

1 Combining a prioritization strategy and functional studies nominates 5'UTR 2 variants underlying inherited retinal disease

3 Alfredo Dueñas Rey^{1,2}, Marta del Pozo Valero^{1,2,3,*}, Manon Bouckaert^{1,2*}, Filip Van Den Broeck^{4,5},
4 Malena Daich Varela^{6,7}, Mattias Van Heetvelde^{1,2}, Marieke De Bruyne^{1,2}, Stijn Van de Sompele^{1,2},
5 Miriam Bauwens^{1,2}, Jamie Ellingford^{8,9,10}, Hanne Lenaerts^{1,2}, Quinten Mahieu^{1,2}, Dragana Josifova¹¹,
6 Genomics England Research Consortium¹⁰, Carlo Rivolta^{12,13,14}, Andrew Webster^{6,7}, Gavin Arno^{6,7},
7 Carmen Ayuso^{3,15}, Julie De Zaeytijd^{4,5}, Bart P. Leroy^{1,4,5,16}, Elfride De Baere^{1,2}, Frauke Coppieters^{1,2,17}

8 1. Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium.

9 2. Department of Biomolecular Medicine, Ghent University, Ghent, Belgium.

10 3. Department of Genetics, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz, University
11 Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM).

12 4. Department of Ophthalmology, Ghent University Hospital, Ghent, Belgium.

13 5. Department of Head & Skin, Ghent University, Ghent, Belgium.

14 6. UCL Institute of Ophthalmology, London, UK.

15 7. Moorfields Eye Hospital, London, UK.

16 8. Division of Evolution, Infection and Genomic Sciences, School of Biological Sciences, Faculty of
17 Biology, Medicines and Health, University of Manchester, Manchester, UK.

18 9. Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS
19 Foundation Trust, Manchester, UK.

20 10. Genomics England, London, UK.

21 11. Clinical Genetics Department, Guy's Hospital, London, UK.

22 **NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**
12. Department of Ophthalmology, University of Basel, Basel, Switzerland.

23 13. Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland.

24 14. Department of Genetics and Genome Biology, University of Leicester, Leicester, UK.

25 15. Center for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III,
26 Madrid, Spain.

27 16. Division of Ophthalmology & Center for Cellular & Molecular Therapeutics, Children's Hospital of
28 Philadelphia, Philadelphia, PA, USA.

29 17. Department of Pharmaceutics, Ghent University, Ghent, Belgium.

30

31 * Shared second co-authors

32 Corresponding author: Frauke Coppieters (frauke.coppieters@ugent.be)

33

34

35

36

37

38

39

40

41

42 **Abstract**

43 **Background** | 5' untranslated regions (5'UTRs) are essential modulators of protein translation.
44 Predicting the impact of 5'UTR variants is challenging and typically not performed in routine
45 diagnostics. Here, we present a combined approach of a comprehensive prioritization strategy
46 and subsequent functional assays to evaluate 5'UTR variation in two large cohorts of patients
47 with inherited retinal diseases (IRDs).

48 **Methods** | We performed an isoform-level re-analysis of retinal RNA-seq data to identify the
49 protein-coding transcripts of 378 IRD genes with highest expression in retina. We evaluated
50 the coverage of these 5'UTRs by different whole exome sequencing (WES) capture kits. The
51 selected 5'UTRs were analyzed in whole genome sequencing (WGS) and WES data from IRD
52 sub-cohorts from the 100,000 Genomes Project (n = 2,417 WGS) and an in-house database (n
53 = 1,682 WES), respectively. Identified variants were annotated for 5'UTR-relevant features and
54 classified into 7 distinct categories based on their predicted functional consequence. We
55 developed a variant prioritization strategy by integrating population frequency, specific
56 criteria for each category, and family and phenotypic data. A selection of candidate variants
57 underwent functional validation using diverse experimental approaches.

58 **Results** | Isoform-level re-quantification of retinal gene expression revealed 76 IRD genes with
59 a non-canonical retina-enriched isoform, of which 20 display a fully distinct 5'UTR compared
60 to that of their canonical isoform. Depending on the probe-design 3-20% of IRD genes have
61 5'UTRs fully captured by WES. After analyzing these regions in both IRD cohorts we prioritized
62 11 (likely) pathogenic variants in 10 genes (*ARL3*, *MERTK*, *NDP*, *NMNAT1*, *NPHP4*, *PAX6*,
63 *PRPF31*, *PRPF4*, *RDH12*, *RD3*), of which 8 were novel. Functional analyses further supported
64 the pathogenicity of 2 variants. The *MERTK*:c.-125G>A variant, overlapping a transcriptional

65 start site, was shown to significantly reduce both luciferase mRNA levels and activity. The
66 *RDH12:c.-123C>T* variant was found *in cis* with the reported hypomorphic *RDH12:c.701G>A*
67 (p.Arg234His) variant in 11 patients. This 5'UTR variant, predicted to introduce an upstream
68 open reading frame, was shown to result in reduced RDH12 protein but unaltered mRNA
69 levels.

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

73 **Keywords**

74 5'untranslated region (5'UTR), upstream open reading frame (uORF), non-coding variation,
75 whole genome sequencing (WGS), whole exome sequencing (WES), *in silico* prioritization,
76 variant interpretation, functional studies, inherited retinal disease (IRD).

77

78

79

80

81

82

83

84

85

86 **Background**

87 Among the numerous non-coding regulatory regions found across the human genome, 5'
88 untranslated regions (5'UTRs) are major determinants of post-transcriptional control and
89 translation efficiency^{1,2}. These regions, with an average length of about 200 nucleotides in
90 humans, are located immediately upstream from protein-coding sequences and include the
91 Kozak consensus sequence around the AUG start codon³. 5'UTRs harbor numerous *cis*-
92 regulatory elements such as internal ribosomal entry sites (IRES) and upstream open reading
93 frames (uORFs), which can recruit scanning ribosomes and initiate translation⁴. In particular,
94 uORFs, defined by an upstream start codon in-frame with a stop codon preceding the end of
95 the primary open reading frame, can decrease downstream protein expression up to 80%⁵.
96 Furthermore, 5'UTRs can serve as platforms for the formation of secondary and tertiary mRNA
97 structures like stem-loops, hairpins, and RNA G-quadruplexes which further influence mRNA
98 translation^{6,7}. Given the critical importance of 5'UTRs as modulators of protein expression,
99 genetic variation within these regions can contribute to disease pathogenesis, as shown by
100 several examples in a wide range of inherited diseases⁸⁻¹¹.

101 With the expanding implementation of massively parallel sequencing technologies in clinical
102 practice, the rare disease field has witnessed a true paradigm shift enabling improved
103 molecular diagnosis for many affected individuals¹²⁻¹⁵. Despite these advancements, and
104 although this varies greatly across individuals with diverse clinical indications, a significant
105 fraction of cases with suspected Mendelian disorders remains unsolved^{12,13,16-18}.

106 To date, 5'UTR variants are often not evaluated because of the multiple challenges they entail.
107 Firstly, detection relies on inclusion of these non-coding regions in targeted resequencing
108 panels or whole genome sequencing, which has been limited to date¹⁹. Secondly,
109 interpretation of the effect is more difficult due to different mechanisms⁸ and our relative lack

110 of understanding of the functional consequences of 5'UTR variants. Finally, functional
111 evidence is necessary to support their pathogenicity and hence confirm the molecular
112 diagnosis. Recently, to close the annotation and interpretation gap, several *in silico* tools²⁰⁻²⁶
113 and guidelines²⁷ have been developed, although their specific applicability to the accurate and
114 comprehensive interpretation of the diverse pathogenic mechanism of 5'UTR variants is yet
115 to be fully established.

116 A clear instance of both the effectiveness and remaining challenges of clinical large-scale
117 genome analyses can be found in inherited retinal diseases (IRDs)²⁸⁻³¹. IRDs comprise a
118 genetically and phenotypically diverse constellation of visually debilitating conditions
119 affecting over 2 million people globally^{32,33}. With over 290 known disease genes³⁴, diverse
120 modes of inheritance²⁹, intersecting phenotypes^{35,36}, and available automated approaches to
121 variant interpretation, establishing a genetic diagnosis in IRD patients can be challenging. Even
122 after the use of unbiased approaches such as whole exome (WES) and whole genome
123 sequencing (WGS), up to 50% of cases remain unsolved^{15,28,37-39}, and hence without the
124 possibility of potential clinical intervention including gene therapy-based treatments^{40,41}.

125 Recent studies have revealed an important contribution of non-coding variation to IRDs,
126 particularly with the identification of deep-intronic mis-splicing variants^{39,42,51,43-50}. Moreover,
127 there is a growing body of evidence supporting the emerging role of genetic variation affecting
128 *cis*-regulatory regions in the molecular pathogenesis of IRDs^{45,48,52-56}. However, thus far only
129 a few 5'UTR variants have been reported to be implicated in ocular diseases, including
130 IRDs^{48,53,54,57-59}, but no large-scale studies have been performed yet.

131 In this work, we conducted a systematic evaluation of 5'UTR variation in IRD genes using WGS
132 and WES data derived from two IRD cohorts comprising 2,417 participants of the 100,000
133 Genomes Project¹² and 1,682 local cases, respectively. Firstly, we obtained a comprehensive

134 selection of 5'UTRs of the most abundant canonical as well as non-canonical protein-coding
135 IRD gene isoforms by performing transcript-level re-quantification of retinal expression data.
136 We then screened these regions for variants in both IRD cohorts and developed a prioritization
137 strategy to identify candidate pathogenic 5'UTR variants. This allowed us to identify 11
138 potentially causative 5'UTR variants, 10 of which were found in unsolved cases. Functional
139 validation of the predicted pathogenetic mechanism was performed for 4 of these, further
140 supporting the potential implication in disease for 2 variants. Overall, we show that 5'UTRs
141 represent understudied targets of non-coding variation that can provide novel molecular
142 diagnoses in IRDs and demonstrate the importance of reassessing these regions in existing
143 exome and genome sequencing data.

144 **Methods**

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

146 We retrieved paired-end FASTQ files (GSE115828) derived from human postmortem retina
147 samples characterized by Ratnapriya *et al.*, (2019)⁶⁰. Only samples derived from donor retinas
148 showing no features of age-related macular degeneration were evaluated (n=102). Transcripts
149 were quantified through pseudoalignment by *Kallisto*⁶¹ (v.0.46.1) using default parameters.
150 Abundance estimates in transcripts per million (TPM) for all annotated transcripts (*Ensembl*
151 human release 107) were retrieved. A custom R (v.4.0.2) script was used for annotating each
152 transcript with its corresponding gene and biotype and flagging them as canonical and/or
153 belonging to the Matched Annotation from NCBI and EMBL-EBI (MANE) and MANE Plus
154 Clinical set (version 1.0)⁶². For each group of isoforms belonging to each gene, we computed
155 the average of abundance estimates across all samples and the isoform exhibiting the highest
156 average among its corresponding group was deemed retina-enriched. This dataset was further

157 filtered by retaining only protein-coding isoforms derived from selected genes, namely IRD
158 genes listed in either the *Retinal disorders panel* (v2.195) from Genomics England PanelApp⁶³
159 or RetNet (<https://sph.uth.edu/retnet/>) (**Table S1**). Additionally, we integrated cap analysis
160 gene expression sequencing (CAGE-seq) data derived from fetal and adult retina (FANTOM5⁶⁴
161 Robust Peak Set using an expression RLE normalized threshold >1) to evaluate the confidence
162 of the annotated transcription start sites (TSS) of the selected transcripts in retina.
163 To aid tiering of variants, the Genomics England PanelApp⁶³ gene color-coded rating system
164 was used: genes with diagnostic-grade rating, borderline evidence and research candidates
165 were flagged with green, amber, and red, respectively. To further assess the retina-
166 enrichment of specific isoforms, their loci were inspected using an integration of multiple
167 publicly available multi-omics datasets derived from human retina (**Table S2**).

168 **Comparison of 5'UTRs of canonical and non-canonical isoforms and selection of 5'UTR** 169 **variant search space**

170 The exact genomic coordinates of the start and end positions of the 5'UTRs of all genes were
171 downloaded from *Ensembl biomart* (Human Ensembl Genes 107, GRCh38.p13) and filtered to
172 only include the transcripts defined above. We first assessed the number of exons composing
173 these 5'UTRs to identify which IRD genes have spliced 5'UTRs. For each IRD gene for which a
174 retina-enriched non-canonical isoform was identified, we compared the canonical and non-
175 canonical 5'UTRs and computed their respective overlap. The calculated overlaps were then
176 used to classify these genes into three different categories, namely genes with transcripts
177 displaying: (i) fully distinct 5'UTRs, (ii) partly overlapping 5'UTRs, and (iii) fully overlapping
178 5'UTRs.

179 For each 5'UTR exon of all selected transcripts, we defined a near-splice region by including
180 25-bp intronic sequence up- and downstream of its respective splice donor and acceptor sites

181 (excluding promoter and coding sequences). The resulting coordinates were annotated with
182 their corresponding transcript identifier and gene name and stored into a sorted BED file
183 (herein after referred to as *5'UTR analysis file*) for downstream variant assessment.

184 **Evaluation of 5'UTR capture by whole-exome sequencing**

185 We assessed the performance of exome captures of the selected 5'UTRs based on the designs
186 provided by the kits which were mostly used for the generation of our in-house WES data,
187 namely the SureSelect Human All Exon V6 and SureSelect Human All Exon V7 (Agilent
188 Technologies), as well as a selection of the most recent versions of commonly used kits from
189 4 different providers: SureSelectXT Human All Exon V8 (Agilent Technologies), KAPA
190 HyperExome V2 (Roche), Twist Exome 2.0 (Twist), Illumina Exome Panel v1.2 (Illumina). BED
191 files containing the genomic coordinates of the capture regions were downloaded from the
192 corresponding design catalogs and intersected with the coordinates of the 5'UTRs of interest
193 using *bedtools intersect* (v2.26.0)⁶⁵ with default parameters. For uniformity, only the capture
194 regions were used to compare between the different designs. Additionally, for the SureSelect
195 Human All Exon V6 and SureSelect Human All Exon V7 kits (Agilent Technologies) both the
196 strict union of all regions covered by baits and a version padded by ± 50 bp extending into
197 intronic regions were considered for the intersections. A custom Python (v.3.6.8) script was
198 then used to compute for each IRD gene the length proportion (%) of its 5'UTR captured by
199 these kits.

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

201 The ClinVar database (ClinVar) was downloaded in a tab-delimited format⁶⁶ directly from the
202 FTP site (<https://ftp.ncbi.nlm.nih.gov/pub/clinvar/>; version from 2023-03-18) and pre-filtered
203 to keep only entries from the GRCh38.p13 build. The resulting file was further filtered to

204 retrieve variants located within the regions defined in the *5'UTR analysis file*. Large copy
205 number gain/loss variants extending into coding regions as well as variants with a protein-
206 altering or synonymous annotation for the canonical transcript were removed from this
207 analysis.

208 **Cohort selection**

209 To assess the contribution of 5'UTR genomic variation to IRDs, individuals from two different
210 cohorts were selected for this study: (i) 2,397 participants (2,100 probands) from a sub-cohort
211 of the Rare Disease arm of the 100,000 Genomes Project (Genomics England –GE– cohort)
212 affected by posterior segment abnormalities (**Table S3**); (ii) an IRD sub-cohort of 1,682 cases
213 (1,030 probands) with a WES analysis performed at the Center for Medical Genetics Ghent
214 (CMGG cohort). In both cases, sequencing data aligned to GRCh38 build were included.

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

220 **Sequencing and variant analysis**

221 We interrogated WGS and WES data from the two cohorts described above to detect germline
222 single-nucleotide variants (SNVs) and small insertions and deletions (indels) overlapping the
223 regions defined in the *5'UTR analysis file*. The sequencing and bioinformatic pipelines used for
224 processing genome data derived from the participants of the GE cohort have been described
225 previously¹².

226 Samples from the CMGG cohort were tested using WES with the SureSelect Human All Exon
227 V6, SureSelect Human All Exon V7 (Agilent Technologies) or HyperExome (Roche) enrichment
228 kits and sequenced on HiSeq 3000 or NovaSeq 6000 instruments (paired-end 150 cycles)
229 (Illumina). Reads were aligned to the human reference genome (GRCh38 build) with *BWA*
230 (v0.7.15)⁶⁷ and *GATK HaplotypeCaller* (v3.82)⁶⁸ was used for calling SNVs and indels. Resulting
231 variant call format (VCF) files from both cohorts were subsequently parsed based on the
232 regions of interest using *BCFtools* (v1.9)⁶⁹. Only variants satisfying the filter criteria for
233 sequencing depth (DP>10) and genotype quality (GQ>15) were retrieved.

234 **Variant annotation and prioritization**

235 From each cohort, a file containing unique variants was created and formatted so that it could
236 be annotated with the *Ensembl Variant Effect Predictor*⁷⁰ (VEP, release 107). Apart from gene
237 and transcript information, each variant was annotated with frequency data retrieved from
238 gnomAD⁷¹ (genomes, v3.1.2) allele frequencies, splicing (dbSNV⁷², SpliceAI²³,
239 MaxEntScan⁷³), pathogenicity predictions (EVE⁷⁴, CADD⁷⁵), and regulatory data⁷⁶. The recently
240 developed *UTRannotator*²⁰ tool was used to annotate variants that create or disrupt uORFs.
241 The classical⁷⁷ and retinal⁷⁸ Kozak consensus sequences were considered altered when
242 variants were identified within the position range -1 to -10 relative to the main AUG (canonical
243 start codon). In addition, nucleotide frequency plots corresponding to the Kozak sequences of
244 the retina-enriched and not retina-enriched transcripts evaluated in this study were generated
245 using were generated with *WebLogo*⁷⁹ (v.3.7.12). Annotation data tables containing predicted
246 changes in translational efficiency⁸⁰ and secondary structure minimum free energy affecting
247 double-stranded RNA or G4 quadruplex structures⁸¹ were downloaded (*5utr* ['*suter*']:
248 <https://github.com/leklab/5utr>) and queried for the variant positions using *tabix*⁸² (v.1.7-2);
249 further evaluation of changes in mRNA secondary structure was performed using the *Ufold*⁸³

250 (v.1.2), *REDfold*⁸⁴ (v1.14.alpha), and *MXfold2*⁸⁵ (v0.1.2) tools. To assess whether variants were
251 located within TSS relevant to retinal gene expression, we made use of the CAGE-seq data
252 described above. Translation initiation-related feature data from the Human Internal
253 Ribosome Entry Sites (IRES) Atlas⁸⁶ was also queried to annotate variants found within these
254 regions involved in cap-independent translation initiation.

255 Annotated 5'UTR variants were classified into the following 7 categories: (i) uAUG gained, (ii)
256 change in existing uORF, (iii) alteration of classical or retinal primary Kozak context, (iv)
257 splicing, (v) change in translational efficiency (TE), (vi) change in secondary structure minimum
258 free energy, and (vii) overlapping a retinal transcription start site and/or an IRES. Variants
259 were first selected when their minor allele frequency was lower than 2% in all populations.
260 The resulting variants were further filtered using specific criteria for each category: (i) uAUG
261 created in a strong or moderate Kozak context, (ii) natural uAUG loss, (iii) variant located in
262 positions -3, -4, -5, -6, or -9, (iv) any SpliceAI Delta Score (DS_AG, DS_AL, DS_DG, DS_DL) higher
263 than 0.2, (v) $|\log_2FC| \geq 0.5$, (vi) $|FC| \geq 1.5$, (vii) variant located within a TSS or IRES. Variants for
264 which the inheritance pattern and phenotype of the patient were compatible with the gene
265 in which the 5'UTR variant was identified were selected as candidates. This prioritization
266 procedure is depicted in **Figure 1**. For comparing the proportion of variants for each category
267 between cohorts, the statistical analysis was performed in R using the χ^2 -test. Finally, for each
268 candidate variant identified in unsolved cases, an additional screening was performed to
269 discard (likely) pathogenic variants, both SNVs and structural variants (SVs), in other IRD genes
270 that could provide an alternative molecular diagnosis. For the cases from the GE cohort in
271 which the identified 5'UTR variant remained as candidate, we requested to have a clinical
272 collaboration with Genomics England.

273 **Cell culture**

274 ARPE-19 (ATCC, CRL-2302™) and HEK-293T cells (ATCC CRL-3216™) cells were grown in either
275 Dulbecco's minimal essential medium (DMEM) with phenol red (Thermo Scientific Life
276 Technologies) or DMEM:F12 (Gibco) medium supplemented with 10% fetal bovine serum
277 (Gibco), 1% penicillin-streptomycin (Gibco), 1% non-essential amino acid solution (Gibco), and
278 0.1% amphotericin B (Gibco), respectively. Cells were cultured at 37°C and 5% CO₂ and tested
279 for mycoplasma contamination prior to use.

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

286 **Cloning and mutagenesis**

287 In order to evaluate the functional effect of 3 selected 5'UTRs variants (*RDH12:c.-123C>T*,
288 *MERTK:c.-125G>A*, *PAX6:c.-44C>T*), we cloned the wild-type 5'UTR of interest (IDT gBlock) into
289 a psiCHECK™-2 dual luciferase vector (Promega) using the Cold Fusion cloning kit (Sanbio BV).
290 The recombinant vectors were then amplified in One Shot TOP10 Chemically Competent *E.*
291 *coli* cells (Invitrogen) and purified using the NucleoBond Xtra Midi kit (Filter Service S.A). For
292 the generation of the overexpression construct to further assess the *RDH12:c.-123C>T* variant,
293 we designed a gBlocks™ fragment comprising the wild-type 5'UTR and coding sequence (CDS)
294 of *RDH12*; both sequences were modified to include downstream Myc and FLAG in-frame tags
295 to evaluate the translation of the uORF introduced by the *RDH12:c.-123C>T* variant and *RDH12*
296 protein levels, respectively. These fragments were then cloned into a pcDNA™3.1(+)
297 (Invitrogen) vector by restriction-ligation cloning and the recombinant vectors were amplified

298 and purified as described above. For all constructs, 5'UTRs variants were introduced using the
299 Q5 Site-Directed Mutagenesis Kit (NEB) using variant-specific primers designed with the
300 *NEBaseChanger* tool. The sequence of each insert was confirmed by Sanger sequencing using
301 the BigDye Terminator v3.1 kit (Life Technologies) and/or long-read whole plasmid sequencing
302 (Plasmidsaurus). A schematic overview of these constructs is shown in **Figure S1**. All primer
303 sequences can be found in **Table S4**.

304 **Dual luciferase assays**

305 ARPE-19 cells were seeded in a 24-well plate (Greiner Bio-One BVBA) at a density of 50,000
306 cells/well in 1mL of medium without antibiotics to reach 70-90% confluence at transfection
307 (18-24h after plating). Cells were transfected at a 3:1 reagent to plasmid DNA ratio using the
308 TransIT-X2[®] Dynamic Delivery System (Mirus Bio) according to the manufacturer's
309 instructions. After 24 h, cells were lysed, and luciferase activity was detected using the Dual-
310 Glo[®] Luciferase Assay System (Promega) in a Glomax 96-Microplate Luminometer (Promega).
311 Each transfection was performed in triplicate and each experiment was repeated at least three
312 times to ensure reproducibility. For each well, the ratio of *Renilla* luciferase activity was
313 normalized to *Firefly* luciferase activity. A custom R script was used to evaluate the effect of
314 each variant on luciferase activity through a linear mixed effects model (implemented in the
315 *lme4* package⁸⁷) having set the luciferase vector as fixed effect and the biological replicate as
316 random effect.

317 **RNA isolation and quantitative polymerase chain reaction (qPCR)**

318 Total RNA was extracted using either the RNeasy Mini kit[®] (Qiagen) (ARPE-19 and HEK-293T
319 cells) or the Maxwell RSC simply RNA kit (Promega) (cultured lymphocytes) according to the
320 manufacturer's instructions. Isolated RNA underwent DNase treatment (ArcticZymes, Tromsø,

321 Norway) prior to cDNA synthesis with the iScript cDNA Synthesis Kit (Bio-Rad Laboratories).
322 For each cDNA sample, qPCR assays were prepared using SsoAdvanced Universal SYBR Green
323 Supermix (Bio-Rad Laboratories) and run on LightCycler 480 System (Roche). Data were
324 analyzed with *qbase+*⁸⁸ (CellCarta, v.3.4) and normalized either to a set of housekeeping
325 genes (*YWHAZ*, *HPRT1*, *HMBS*, *SDHA*) or *Firefly* luciferase for normalization of *Renilla*
326 luciferase mRNA levels. All primer sequences can be found in **Table S4**. Statistical analyses
327 were performed in R using the Wilcoxon rank sum test.

328 **Overexpression and immunoblotting**

329 To perform immunoblotting of RDH12 and the predicted peptide encoded by the uORF
330 introduced by the *RDH12:c.-123C>T* variant, ARPE-19 cells were seeded in 12-well plates
331 (Greiner Bio-One BVBA) at a density of 100,000 cells/well, allowed to settle overnight, and
332 then transfected with the wild-type and mutant *RDH12* overexpression vectors using the
333 TransIT-X2[®] Dynamic Delivery System (Mirus Bio) according to the manufacturer's
334 instructions. The pcDNA[™]3.1⁽⁺⁾ (Invitrogen) backbone vector was transfected as negative
335 control. Four hours prior to RNA and protein isolation, cells were treated with 10 μ M MG-132
336 proteasome inhibitor (Merck Life Science). For total protein extraction, cells were lysed with
337 RIPA Buffer (Sigma-Aldrich) including protease inhibitory cocktail (Roche Diagnostics),
338 phosphatase inhibitory cocktail 2, and phosphatase inhibitory cocktail 3 (Sigma-Aldrich).
339 Protein concentrations were measured using the Pierce[™] BCA Protein Assay kit (Fisher
340 Scientific). After centrifugation and reduction with 1M DTT (Sigma-Aldrich), protein lysates
341 were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis using either
342 NuPAGE[™] 4–12% Bis-Tris (RDH12 blot) or Novex[™] 16% Tricine (uORF blot) Protein Gels (Fisher
343 Scientific) with a ladder (Precision Plus Protein All Blue Standards, Bio-Rad Laboratories).
344 Proteins were then transferred to a nitrocellulose membrane using the iBlot 2 Dry Blotting

345 System (Thermo Fisher Scientific). Membranes were blocked for 2 hours at room temperature
346 in 2% ECL™ Blocking Agent (Cytiva Amersham) and incubated at 4°C overnight with anti-FLAG
347 (1:1000, F1804, Merck Life Science) or anti-Myc (1:1000, ab9106, Abcam) primary antibodies.
348 Membranes were subsequently incubated for 2 hours at room temperature with the
349 appropriate horseradish-peroxidase-conjugated secondary antibody (1:2500, 7076S or 7074S,
350 Cell Signaling Technologies) and revealed with the SuperSignal™ West Dura Extended Duration
351 Substrate (Fisher Scientific). Membranes were scanned with an Amersham Imager 680 system
352 (GE Healthcare Life Sciences). Protein quantification was performed by firstly stripping the
353 membranes with Restore™ PLUS Western Blot Stripping Buffer (Thermo Scientific), incubating
354 them for 1 hour at room temperature with a primary antibody against β -tubulin (1:2500,
355 ab6046, Abcam) and for 2 hours with a horseradish-peroxidase-conjugated secondary
356 antibody (1:2500, 7074S, Cell Signaling Technologies). RDH12 (FLAG) signal intensity
357 quantification was achieved using *ImageJ* (NIH, v.1.50i) and normalized to the amount of β -
358 tubulin.

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

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

366 The *RDH12:c.-123C>T* 5'UTR variant was shared with members of the European Retinal
367 Disease Consortium (ERDC: <https://www.erd.c.info/>) to evaluate its pathogenicity and identify

368 additional carriers of the *RDH12*:c.701G>A (p.Arg234His) missense variant to confirm its *cis*-
369 configuration. Primer sequences for PCR and Sanger sequencing are listed in **Table S4**.

370 **Results**

371 **Isoform-level re-quantification of retinal gene expression reveals differential 5'UTRs of IRD** 372 **genes**

373 To obtain a relevant selection of 5'UTRs for downstream variant screening, we performed a
374 transcript-level re-analysis of RNA-seq data to identify protein-coding non-canonical isoforms
375 with relevant retinal expression i.e., higher than its respective canonical isoform, to be
376 retained in addition to the canonical isoforms. A total of 454 canonical and non-canonical
377 transcripts belonging to 378 IRD genes were thus selected (with maximum of 2 isoforms per
378 gene) (**Table S5**), from which their 5'UTRs were retrieved, resulting in 638 genomic regions
379 (**Table S6**). Considering these 454 transcripts, the 5'UTRs of approximately 62% of IRD genes
380 (233/378) are only part of the first coding exon while 33% of IRD genes (126/378) have
381 exclusively transcripts with spliced 5'UTRs, i.e., their 5'UTR comprise additional full non-coding
382 exons. A minor fraction (19/378) of genes has isoforms with both spliced and non-spliced
383 5'UTRs (**Figure 2A; Table S7**). Of the 76 IRD genes for which a retina-enriched non-canonical
384 isoform was identified, 20 display a fully distinct 5'UTR compared to the one of their
385 corresponding canonical isoforms (**Figure 2B; Table S8**). For six of these, the non-canonical
386 TSS is further supported by CAGE-seq data derived from both adult and fetal retina (**Table S5;**
387 **Figure S2**). Two remarkable examples are *CRB1* and *RIMS2*, for which the non-canonical
388 isoforms (ENST00000681519.1: 48.38±14.97 TPM; ENST00000436393.6: 41.58±12.35 TPM)
389 were found to be more abundant in retina compared to their respective canonical isoforms
390 (ENST00000367400.8: 2.54±3.88 TPM; ENST00000696799.1: 12.19±3.85 TPM). These findings

391 were also supported by the integration of multiple multi-omics datasets derived from human
392 retina (**Figure S3; Table S2**).

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

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

406 **A retrospective analysis of 5'UTR variants reported in IRD genes reveals a majority classified** 407 **as Variants of Uncertain Significance (VUS)**

408 A total of 1,547 5'UTR variants in IRD genes have been submitted to the ClinVar database thus
409 far (**Table S11**). All variants except 4 have been clinically interpreted. Only 2% (31/1,547) of
410 the variants have been classified as pathogenic or likely pathogenic, of which 32% (10/31) and
411 16% (5/31) were found in the *PAX6* and *NMNAT1* genes, respectively. On the other hand, 34%
412 (527/1,547) of the 5'UTR variants have been classified as benign or likely benign, of which no
413 single gene was found to account for more than 5% of these variants. While 5% (81/1,547) of

414 the variants had conflicting interpretations of pathogenicity, the largest fraction comprising
415 58% (904/1,547) of all variants have been classified as VUS.

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

418 To systematically assess the contribution of 5'UTR genomic variation to IRDs, we performed a
419 variant analysis within the selected 643 genomic regions in affected individuals from two
420 different IRD sub-cohorts, namely the GE (n=2,397 WGS) and CMGG (n=1,682 WES) cohorts.
421 We identified a total of 2,898 and 381 distinct 5'UTR variants within the GE and CMGG cohorts,
422 respectively. The majority of these variants (2,637/2,898 and 334/381) had a minor allele
423 frequency (MAF) lower than 2% in all populations and a substantial fraction (506/2,898 and
424 92/381) was found to be absent from all reference population public databases. To aid the
425 interpretation of these 5'UTR variants, we classified them into 7 (non-mutually exclusive)
426 categories according to their *in silico* predicted functional consequences (**Figure 1**). A
427 summarized overview of the number of variants that remained after classification and
428 category-specific filtering can be found in **Table 1**. Out of the 1,450 remaining variants
429 (**Dataset S1**), 370 were present in more than one category. Of the remaining variants assigned
430 to a single category, the bulk corresponded to variants overlapping a retinal TSS or an IRES
431 (761/1,450; 221 -TSS-, 488 -IRES-, 52 -TSS & IRES-), followed by variants with a predicted
432 change in secondary structure minimum free energy (204/1,450). When comparing the
433 number of variants in each category between the two cohorts, we observed statistically
434 significant differences in the number of variants overlapping a retinal TSS and/or IRES, for
435 which the proportion of variants was higher for the WGS-based GE cohort compared to the
436 WES-based CMGG cohort (~55% and ~35% respectively, $p < 0.05$). These differences are most
437 likely due to the expanded capture of regions including TSS allowed by the use of WGS in the

438 GE cohort. Additionally, it is noteworthy that for the IRD genes with 5'UTRs harboring IRES
 439 (92/378), these elements were found to span on average 61% of the 5'UTR and even its entire
 440 length for certain genes (12/92) (**Table S12**).

441 **Table 1. Summarized overview of the number of 5'UTR variants that remained after functional**
 442 **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	6	22	1.52
Splicing	29	10	39	2.69
Change in translational efficiency	1	0	1	0.07
Change in secondary structure minimum free energy	194	10	204	14.07
Overlapping a retinal transcription start site or an IRES	692	69	761	52.48
> 1 category	276	94	370	25.52
Total	1255	195	1450	

443 Abbreviations: CMGG, Center for Medical Genetics Ghent; GE, Genomics England; IRES, internal ribosomal entry
 444 site; uORF, upstream open reading frame.

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

447 To enhance the likelihood of identifying potential pathogenic 5'UTR variants, we prioritized
 448 the ones with higher predicted impact based on specific criteria for each category (**Figure 1**)
 449 and for which the inheritance patterns and reported phenotypes fitted the genes in which the
 450 5'UTR variants had been identified. The application of this prioritization procedure is
 451 illustrated in **Table S13**. We prioritized 11 candidate pathogenic 5'UTR variants. Of these, 10
 452 were found in unsolved cases and 8 have not been reported before. Of note, although initially
 453 identified through WGS, 6 of these variants would also be covered by WES. An overview of

454 the genetic, phenotypic, and *in silico* prediction details of these variants is shown in **Table 2**.
455 The pedigrees of these families including variant segregation are shown in **Figure S5**.
456 Four of these variants were found to alter splicing in genes linked to autosomal dominant
457 (*PRPF31*), autosomal recessive (*NMNAT1*) and X-linked (*NDP*) disease. The *NDP*:c.-70G>A and
458 *PRPF31*:c.-9+1G>A variants have recently been reported by Daich Varela *et al.* (2022)⁴⁸. We
459 identified a *de novo* variant (*PRPF31*:c.-9+1G>T) in the same 5'UTR position of *PRPF31* in a
460 sporadic case with rod-cone dystrophy in which bi-allelic *GRM6* variants had been identified,
461 one of them reported as likely benign (**Table S14**). Variants in *GRM6* cause autosomal
462 recessive congenital stationary night blindness⁸⁹, which is different from the clinical diagnosis
463 in this proband. The *NMNAT1*:c.-57G>A variant was found in an individual with macular
464 dystrophy that carries *in trans* an extremely rare missense variant for which *in silico*
465 predictions support a pathogenic effect.
466 Changes in secondary structure were predicted for 4 variants in genes linked to autosomal
467 dominant (*ARL3*, *PAX6*) and autosomal recessive (*MERTK*, *RD3*) disease (**Figure S6**). Given the
468 inheritance pattern and *in silico* predictions, a gain-of-function effect was hypothesized for
469 the *ARL3*:c.-88G>A and *PAX6*:c.-44T>C variants. Both the *MERTK*:c.-125G>A and *RD3*:c.-
470 394G>A variants were identified in homozygous state and found to overlap their respective
471 TSS.
472 The introduction of an uORF was predicted for 2 variants in the *NPHP4* and *RDH12* genes. In
473 contrast to the uORF created by the *RDH12*:c.-123C>T variant, the one introduced by the
474 *NPHP4*:c.-21C>T variant is out-of-frame and overlaps the CDS. Although no segregation could
475 be established for this case with non-syndromic rod-cone dystrophy, an ultra-rare missense
476 *NPHP4* variant with pathogenic *in silico* predictions was identified.

477 Only one variant in a gene linked to dominant disease (*PRPF4*) was predicted to affect the
478 primary Kozak consensus sequence. The *PRPF4*:c.-6C>T variant results in a transition to an
479 infrequent nucleotide at this position within the Kozak consensus sequences derived from the
480 IRD gene isoforms selected in this study (**Figure S7**). This variant was found in a sporadic case
481 with rod-cone dystrophy in which 2 likely pathogenic variants have been identified in the
482 *ITF140* gene (**Table S14**); segregation analysis was not possible to establish the phase of any
483 of these alleles.

484 Four of the variants (*ARL3*:c.-88G>A, *MERTK*:c.-125G>A, *PAX6*:c.-44T>C, *RDH12*:c.-123C>T)
485 were selected for functional validation using diverse multiple experimental approaches.

486 **Downstream functional analyses support the pathogenicity of the *MERTK*:c.-125G>A and**
487 ***RDH12*:c.-123C>T variants**

488 Depending on the tissue-specific gene expression and availability of patient material, we
489 conducted either *in vitro* evaluation in ARPE-19 cells (dual luciferase assays or
490 overexpression), or mRNA expression analysis in patient-derived lymphocyte cultures (**Figure**
491 **4A**). The *RDH12*:c.-123C>T and *MERTK*:c.-125G>A variants were found to result in a significant
492 decrease in luciferase activity (**Figure 4B**). To further elucidate the underlying mechanism,
493 *Renilla* mRNA expression analysis was performed. Relative *Renilla* luciferase mRNA levels
494 remain unchanged for *RDH12*:c.-123C>T, whereas a significant decrease was observed for
495 *MERTK*:c.-125G>A (**Figure 4C**), hence suggesting a translational and a transcriptional effect for
496 these variants, respectively.

497 Taking advantage of the expression of *ARL3* in accessible tissues, we performed qPCR-based
498 quantification of *ARL3* mRNA abundance in lymphocyte cDNA from the two affected siblings
499 in whom we identified the *ARL3*:c.-88G>A variant and five healthy controls. No significant
500 differences were observed in *ARL3* mRNA abundance between the affected carriers and

501 controls (**Figure 4D**). Interestingly, *ARL3* expression levels were slightly higher in the patient
502 samples derived from the lymphocyte cultures treated with puromycin, a translation inhibitor
503 used to suppress nonsense-mediated mRNA decay, compared to the corresponding untreated
504 samples and controls. Therefore, although the *ARL3*:c.-88G>A variant was not predicted to
505 affect splicing, we also performed Sanger sequencing on the cDNA derived from the
506 puromycin-treated and untreated patient lymphocyte cultures. Neither splicing defects nor
507 allele-specific expression were observed (*data not shown*).

508 In view of the negative results obtained for the *PAX6*:c.-44T>C and *ARL3*:c.-88G>A variants,
509 the phenotypes of patients F9 and F10 were re-evaluated. Neither anterior segment
510 abnormalities nor other clinical presentations compatible with a *PAX6*-related disease were
511 observed in patient F9 or her affected relatives. Regarding F10, re-evaluation of one of the
512 affected siblings revealed an acquired vascular ocular condition instead of an IRD.

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

515 We identified the *RDH12*:c.-123C>T variant in one solved case from the GE cohort
516 characterized by bi-allelic *RDH12* coding variants (c.[701G>A];[c.735_743del], p.[Arg234His];
517 [Cys245_Leu247del]). The 5'UTR variant, reported as VUS in ClinVar, was found in *cis* with the
518 p.Arg234His allele. Further evaluation of this 5'UTR variant in 10 additional *RDH12* bi-allelic
519 patients from 8 families carrying the p.Arg234His variant revealed that the c.-123C>T and
520 p.Arg234His variants always form a complex allele (**Table S15**). This was also shown for 7
521 carriers of the p.Arg234His variant affected by other pathologies (*data not shown*).

522 Here, we report the first patient who is homozygous for the *RDH12*: c.[-123C>T];[701G>A]
523 complex allele (**Table S15**). Compared to other patients carrying the 5'UTR variant *in trans*

524 with a null *RDH12* allele⁹⁰, patient F21 presents with a milder phenotype (mild decreased
525 acuity (20/25 OD and OS) in fourth decade of life, foveal sparing maculopathy with a
526 circumscribed area of atrophy within the vascular arcades, without a nasal component.

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

537 **Discussion**

538 Given the emerging role of non-coding variation underlying IRDs^{45,48,52–55} and the essential
539 regulatory function of 5'UTRs^{2,6}, we set out to evaluate their contribution to this
540 heterogeneous group of disorders by analyzing, systematically annotating, filtering and
541 prioritizing 5'UTR variants in two large IRD cohorts in combination with different experimental
542 approaches for functional validation.

543 Thus far, only a few studies have implicated 5'UTR variation in the molecular pathogenesis of
544 IRD cases^{48,53,54}. Our retrospective analysis of 5'UTR variants in IRD genes listed in the ClinVar
545 database⁶⁶ revealed that as many as 58% of all variants have been classified as VUS. A large
546 fraction of the underrepresentation of (likely) pathogenic 5'UTR variants in clinical databases

547 can be explained by their exclusion from downstream tiering pipelines, which prioritize
548 variants with protein-altering consequences and hence neglect the potential impact of 5'UTR
549 variants⁹¹. Furthermore, the identification of these variants is not always feasible by means of
550 commercial exome enrichment platforms, in particular when variants are found within 5'UTR
551 exons that are not part of or proximal to the first protein-coding exon¹⁹. Here, we show that
552 the enriched 5'UTR fraction is highly dependent on the WES capture kit used, and ranges from
553 7% to 39%. Overall, as previous studies have also argued⁹²⁻⁹⁴, our analysis underscores the
554 diagnostic value of re-analyzing exome data, in which causative variants might already be
555 present but previously disregarded.

556 Furthermore, the identification of the genetic defects underlying IRDs is greatly impacted by
557 the inherent complexity of the retinal transcriptional landscape. Variants in retina-enriched
558 isoforms and tissue-specific mis-splicing have been shown to be important molecular
559 mechanisms underpinning disease pathogenesis and phenotypic heterogeneity⁹⁵⁻⁹⁹. As gene
560 isoforms can display differential 5'UTRs that can result in differential translational
561 efficiencies¹⁰⁰, we performed a transcript-level re-analysis of retinal expression data to obtain
562 a relevant selection of 5'UTRs of IRD genes for downstream variant analysis. We identified 76
563 IRD genes with alternative isoforms exhibiting retinal expression levels higher than their
564 respective canonical isoforms, of which 20 displayed a fully distinct 5'UTR. This analysis also
565 found the recently identified photoreceptor-specific non-canonical *CRB1* isoform, which bears
566 a unique 5'UTR exon¹⁰¹. Similarly, we confirmed the retinal enrichment of an alternative
567 *RIMS2* isoform containing an unconventional 5'UTR exon, which has been shown to be
568 photoreceptor-specific and functionally conserved in mouse (Del Pozo-Valero *et al.*,
569 *unpublished data*). By revealing that a significant fraction of IRD genes express alternative

570 retinal isoforms, our analysis highlights the importance of isoform-aware variant annotation
571 for adequate interpretation.

572 To aid the assessment of 5'UTR variant pathogenicity, *in silico* tools have recently been
573 developed, with an emphasis put on uORF-perturbing variants^{20,26}. A systematic
574 characterization of this class of 5'UTR variants showed that they are subject to strong negative
575 selection, which could even be equivalent to that observed against missense variants¹¹.
576 However, thus far, these tools do not provide comprehensive annotations of all possible
577 effects 5'UTR variants can exert⁸. Here, we designed an integrative annotation of 5'UTR-
578 informative features and built a prioritization strategy combining family and phenotypic data
579 to increase the probability of identifying potential pathogenic 5'UTR variants. Although we
580 performed these analyses in IRD cohorts, the features evaluated, and the methods presented
581 in this study can be extrapolated and applied in any setting involving clinical interpretation of
582 5'UTR variation.

583 We prioritized 11 candidate pathogenic 5'UTR variants. The great majority of these variants
584 (8/11) have not been reported before, most likely due to their exclusion from routine
585 prioritization pipelines and subsequent clinical interpretation⁹¹. These 11 variants were
586 assigned to 5 of the 7 categories we defined to classify their potential functional
587 consequences, thus reflecting the diversity in predicted functional effect. Of note, the majority
588 (4/11) was predicted to have an effect on splicing which, in line with our analysis revealing
589 that almost one third of all IRD genes have their 5'UTRs spliced, supports that these regions
590 are also susceptible to disease-causing splicing defects. Notably, we also provided a set of
591 filtered annotated 5'UTR variants with *in silico* predictions pointing to a pathogenic effect that
592 were not prioritized as candidates due to inconsistent genotypes or phenotypes in the patients
593 evaluated. Therefore, we cannot dismiss the possibility that these variants can have a

594 pathogenic effect when present in the right context, e.g. *in trans* with a second pathogenic
595 allele in a gene linked to recessive disease.

596 Given the particular importance of providing functional evidence to support the pathogenicity
597 of variants located within non-coding regions, we evaluated the predicted functional
598 consequence of 4 candidate variants. Despite the extreme rarity and disease co-segregation
599 of the *PAX6*:c.-44C>T variant, both the lack of functional evidence and the unusual phenotype
600 established upon clinical re-evaluation indicate that this variant is not explaining the
601 phenotype. Likewise, we could not confirm the predicted effect of the *ARL3*:c.-88G>A variant
602 using accessible patient-derived material. The sibling of the proband was eventually
603 considered not affected by an IRD, indicating that the variant is either not causative,
604 incompletely penetrant, or simply a recessive allele. Interestingly, a phenotype similar to that
605 of the proband has been associated with a pathogenic variant in this gene¹⁰², which could
606 suggest that this variant exerts a retina-specific effect that cannot be evaluated with the
607 current experimental setting. Nevertheless, it is important to note that in this case both
608 siblings have been studied by WES, but were not tested for known deep-intronic and structural
609 variants. Before proceeding with further functional characterization of the *ARL3*:c.-88G>A
610 variant, WGS should thus be performed to confirm or rule out such variants. Altogether, these
611 results highlight the need and potential limitations of experimental assays to validate the
612 functionality of non-coding variants.

613 We could provide a novel molecular diagnosis in a patient with severe early-onset retinal
614 dystrophy that remained unsolved after WGS screening. This individual was found to be a
615 homozygous carrier for the extremely rare c.-125G>A variant in *MERTK*, overlapping its
616 transcription start site, for which our *in silico* predictions pointed to an effect at the
617 transcriptional level. A drastic reduction of luciferase activity accompanied by decreased

618 relative *Renilla* luciferase mRNA levels confirmed the hypothesized transcriptional effect of
619 the variant, which is consistent with a loss of function of *MERTK*. To our knowledge, this is the
620 first reported disease-causing 5'UTR variant in *MERTK*.

621 One of the candidate pathogenic 5'UTR variants was found in a case solved with bi-allelic
622 *RDH12* coding variants, namely the p.Arg234His hypomorphic missense^{90,103,104} and the
623 p.Cys245_Leu247del in-frame variants. The identified c.-123C>T 5'UTR variant was found *in*
624 *cis* with p.Arg234His, both in this case as well as in 10 additional *RDH12* bi-allelic patients
625 screened for this 5'UTR variant, with one of them being homozygous for this complex allele.
626 The pathogenicity of p.Arg234His has been questioned in view of functional assays revealing
627 *RDH12* protein levels and catalytic activity comparable to the wild-type or polymorphic alleles,
628 respectively¹⁰⁵. We therefore hypothesized a pathogenic effect exerted by the *RDH12*:c.-
629 123C>T variant as part of this complex allele. This 5'UTR variant was predicted to introduce an
630 upstream start codon (uAUG) into a strong Kozak consensus, thereby creating an uORF. Given
631 that ribosomes would first encounter and start translating from this gained uAUG, we
632 hypothesized that this variant would decrease normal *RDH12* translation and thus result in
633 lower protein levels. Furthermore, this variant has been argued to affect the function of an
634 alternative promoter of *RDH12* based on the observed decreased activity in a luciferase
635 reporter assay⁵². Importantly, this experimental setup did not include comparison between
636 mRNA levels and luciferase protein activity and was therefore unable to pinpoint at which
637 level the effect occurs. Here, using a dual luciferase assay with the 5'UTR instead of the
638 genomic sequence, followed by downstream evaluation at both the mRNA and protein level,
639 we confirmed the hypothesized uORF-mediated effect. We tested this hypothesis further
640 using an overexpression setting and demonstrated unaltered mRNA but significantly reduced
641 *RDH12* protein levels, hence confirming the translational effect of *RDH12*:c.-123C>T variant.

642 Although we could not obtain experimental evidence supporting the translation of the uORF-
643 encoded peptide, which might suggest its instability, we cannot rule out alternative
644 mechanisms such as ribosome stalling⁹. However, further studies are needed to confirm this
645 hypothesis. Overall, our results indicate that not p.Arg234His alone, but rather the
646 combination of p.Arg234His and c.-123C>T, or even c.-123C>T by itself, is disease-causing
647 **(Figure 5)**.

648 **Conclusions**

649 5'UTRs are essential modulators of post-transcriptional and translational control. Even though
650 they can be captured by WES to a large extent, variants within these regions are typically
651 overlooked in both research and routine clinical settings, mainly due to their challenging
652 interpretation. In this study, we developed a systematic strategy to comprehensively annotate
653 and prioritize 5'UTR variants in IRD genes. This strategy combined with functional studies
654 allowed us to pinpoint 5'UTR variants providing either novel molecular diagnoses or the full
655 picture explaining the pathogenicity of previously reported hypomorphic variants. Overall, in
656 this study we highlight the importance of multi-layered annotation and validation of non-
657 coding variation potentially underlying disease and provide a 5'UTR interpretation approach
658 that could be extrapolated to other rare diseases.

659

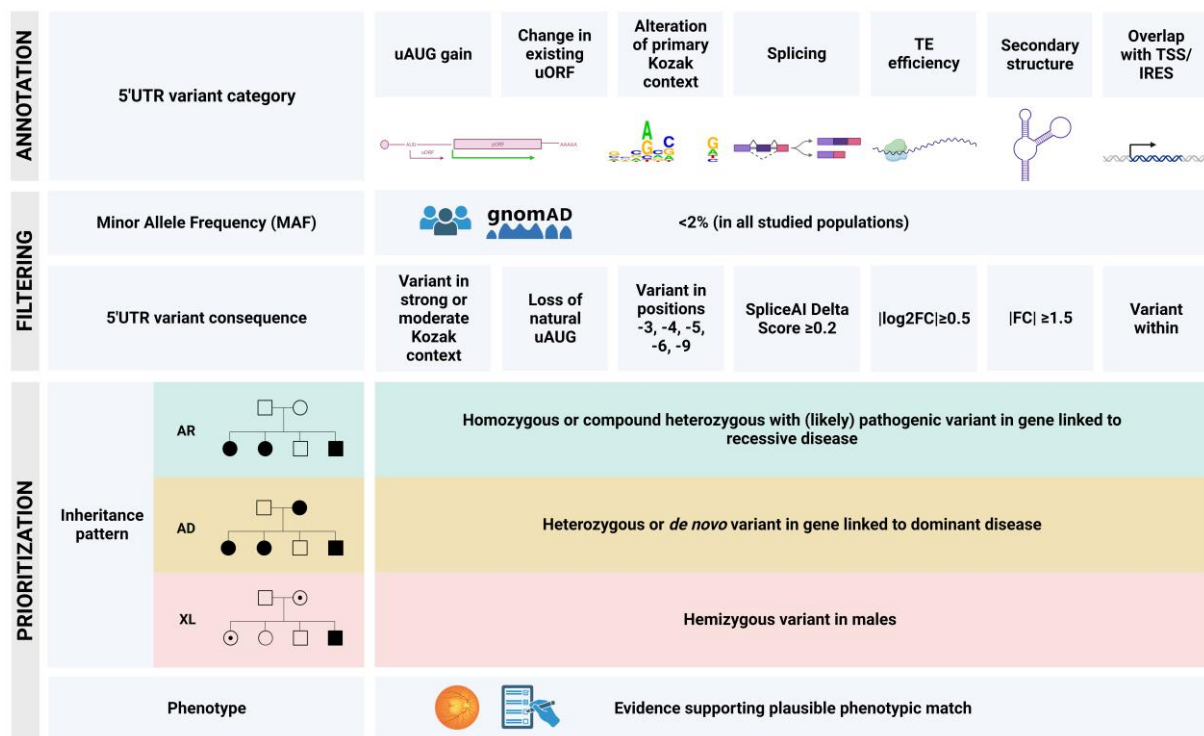
660

661

662

663 **Abbreviations**

- 664 5'UTR: 5' untranslated region
- 665 AD: Autosomal dominant
- 666 AR: Autosomal recessive
- 667 CAGE-seq: Cap analysis gene expression sequencing
- 668 CMGG: Center for Medical Genetics Ghent
- 669 CDS: Coding sequence
- 670 DMEM: Dulbecco's minimal essential medium
- 671 ERDC: European Retinal Disease Consortium
- 672 FC: Fold change
- 673 GE: Genomics England
- 674 IRD: Inherited retinal disease
- 675 IRES : Internal ribosomal entry sites
- 676 MAF: Minor allele frequency
- 677 OD/OS: *Oculus dexter/sinister* (right/left eye)
- 678 qPCR: Quantitative polymerase chain reaction
- 679 SNV: Single-nucleotide variant
- 680 SV: Structural variant
- 681 TSS : Transcription start site
- 682 TPM: Transcripts per million
- 683 TE: Translational efficiency
- 684 uORF: Upstream open reading frame
- 685 VUS: Variant of uncertain significance
- 686 WES/WGS: Whole exome/genome sequencing



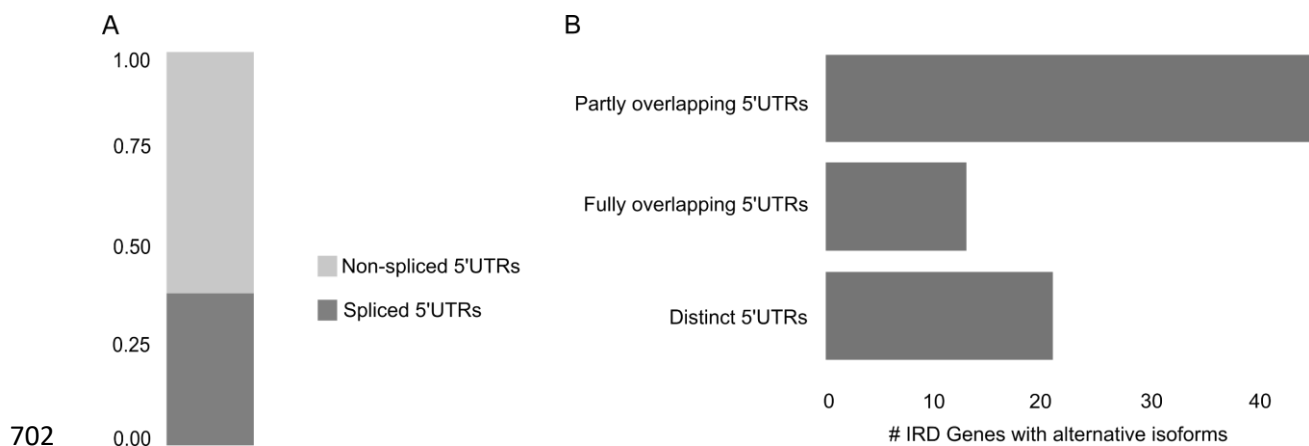
687

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

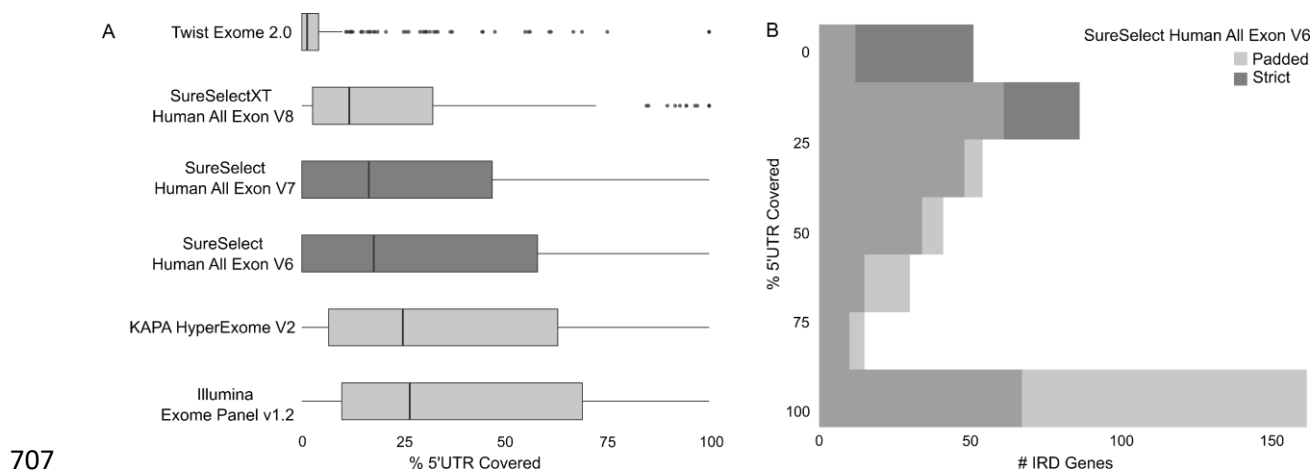
699

700

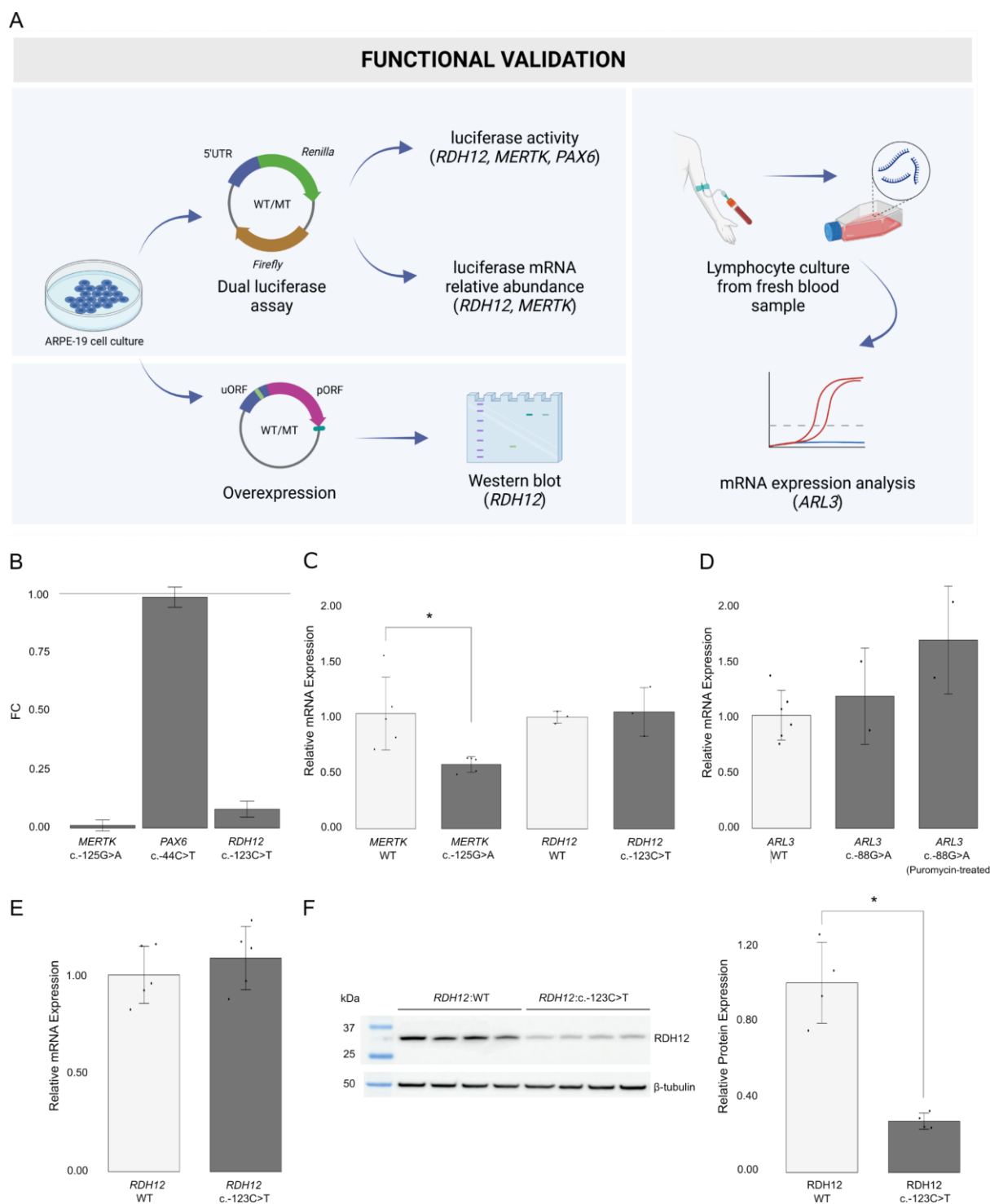
701



702
703 **Figure 2. Characterization of 5'UTRs of IRD genes. (A)** Representation of the proportion of IRD gene
704 isoforms based on the structure of their 5'UTRs. **(B)** Classification of IRD genes with a retina-enriched
705 isoform differing to the canonical one based on the comparison of their respective 5'UTRs.
706



707
708 **Figure 3. Evaluation of 5'UTR capture by whole-exome sequencing (WES). (A)** Boxplots showing the
709 capture performance (*x-axis*) of commonly used commercial exome capture designs on the selected
710 5'UTRs. The kits that were mostly used for the generation of our in-house WES data are highlighted in
711 *darker gray*. **(B)** Histogram representing the portions of 5'UTRs (*y-axis*) of the selected IRD genes which
712 are captured by the SureSelect Human All Exon V6 kit (Agilent Technologies) considering a *strict* or a
713 *padded* design (*see Methods*).



714

715 **Figure 4. Functional evaluation of candidate pathogenic 5'UTR variants in the *ARL3*, *MERTK*, *PAX6*,**

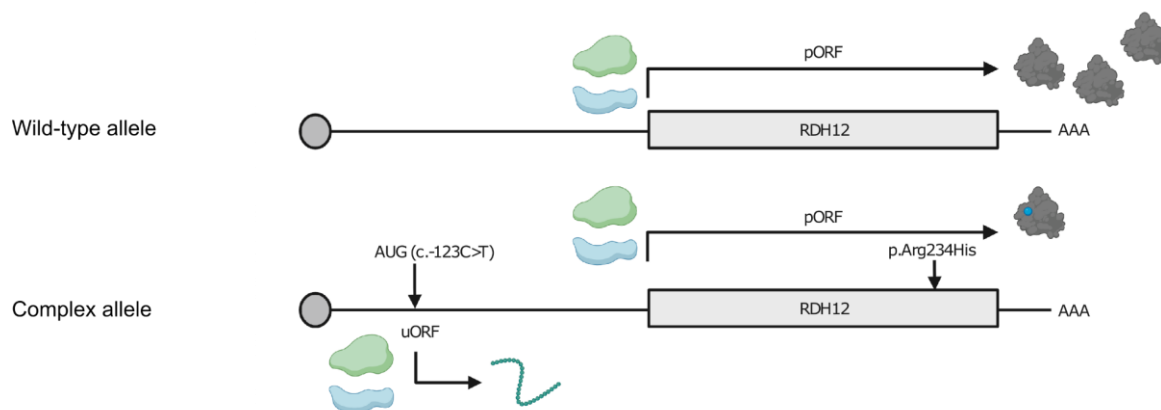
716 **and *RDH12* genes. (A)** Various approaches were used for functionally evaluating candidate variants,

717 including *in vitro* studies (dual luciferase reporter assays and overexpression) and experiments with

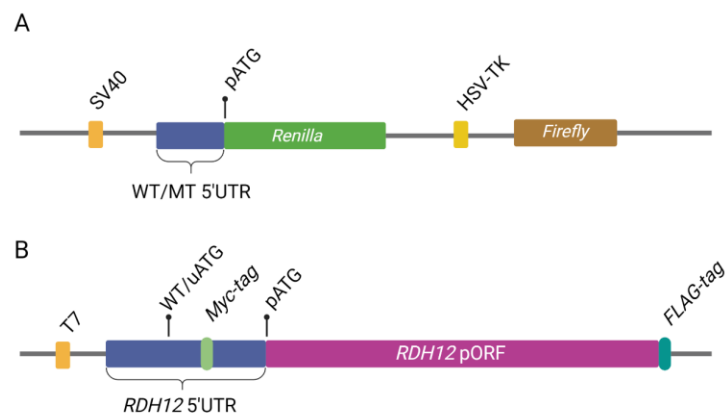
718 clinically-accessible tissues (expression analysis in patient-derived lymphocytes). **(B)** Results from the

719 luciferase assays for the *MERTK*:c.-125G>A, *PAX6*:c.-44C>T, and *RDH12*:c.-123C>T variants. The bar

720 plot shows, for each variant, the fold change (FC) of the luciferase reporter level relative to the level
721 of their corresponding wild-type (WT) construct luciferase vector (FC= 1). The *RDH12*:c.-123C>T and
722 *MERTK*: c.-125G>A variants resulted in significant ($p<0.001$) decrease in luciferase activity (~92% and
723 ~99%, respectively). **(C)** Relative *Renilla* luciferase mRNA levels were significantly decreased (~42%,
724 $p<0.01$) for the *MERTK*:c.-125G>A variant while they remain the same for the *RDH12*:c.-123C>T variant
725 when normalized to mRNA of *Firefly* luciferase and compared to their corresponding wild-type (WT)
726 construct luciferase vectors. **(D)** qPCR quantification of *ARL3* mRNA abundance in lymphocyte cDNA of
727 two affected siblings carrying the *ARL3*:c.-88G>A novel variant and five healthy controls. No significant
728 differences were observed in *ARL3* mRNA abundance between the affected carriers and controls. **(E-**
729 **F)** Using an overexpression setting, the *RDH12*:c.-123C>T variant was shown to result in **(E)** unaltered
730 mRNA levels but **(F)** significantly reduced (~73%, $p<0.01$) RDH12 protein levels.

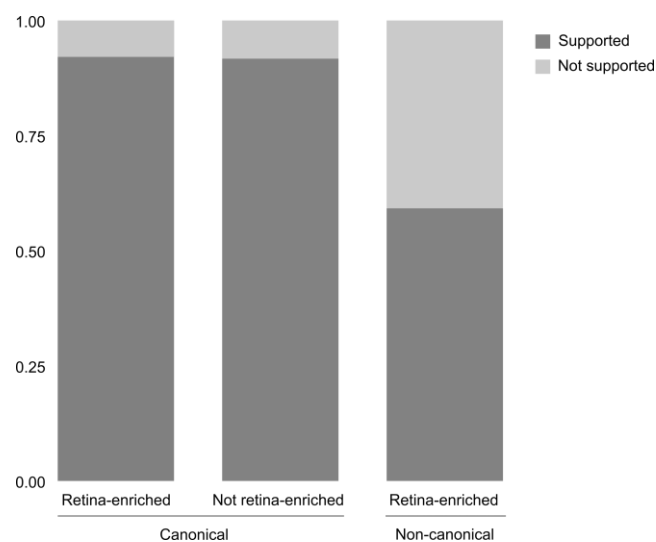


731
732 **Figure 5. Proposed pathogenic mechanism of the *RDH12*: c.-123C>T];[701G>A] complex allele.**
733 Graphical representation of the *RDH12* wild-type allele (*top*) and the *RDH12* complex allele harboring
734 the c.701G>A (p.Arg234His) and the c.-123C>T variants (*bottom*). The 5'UTR variant introduces an
735 upstream start codon into a strong Kozak sequence which can be recognized by ribosomes. Either
736 initiation or active translation of the introduced upstream open reading frame (uORF) results in lower
737 translational efficiency of the primary open reading frame (pORF). As a result, there is a decreased
738 level of RDH12 protein with the arginine-to-histidine amino acid substitution at position 734 (*blue*
739 *point*).



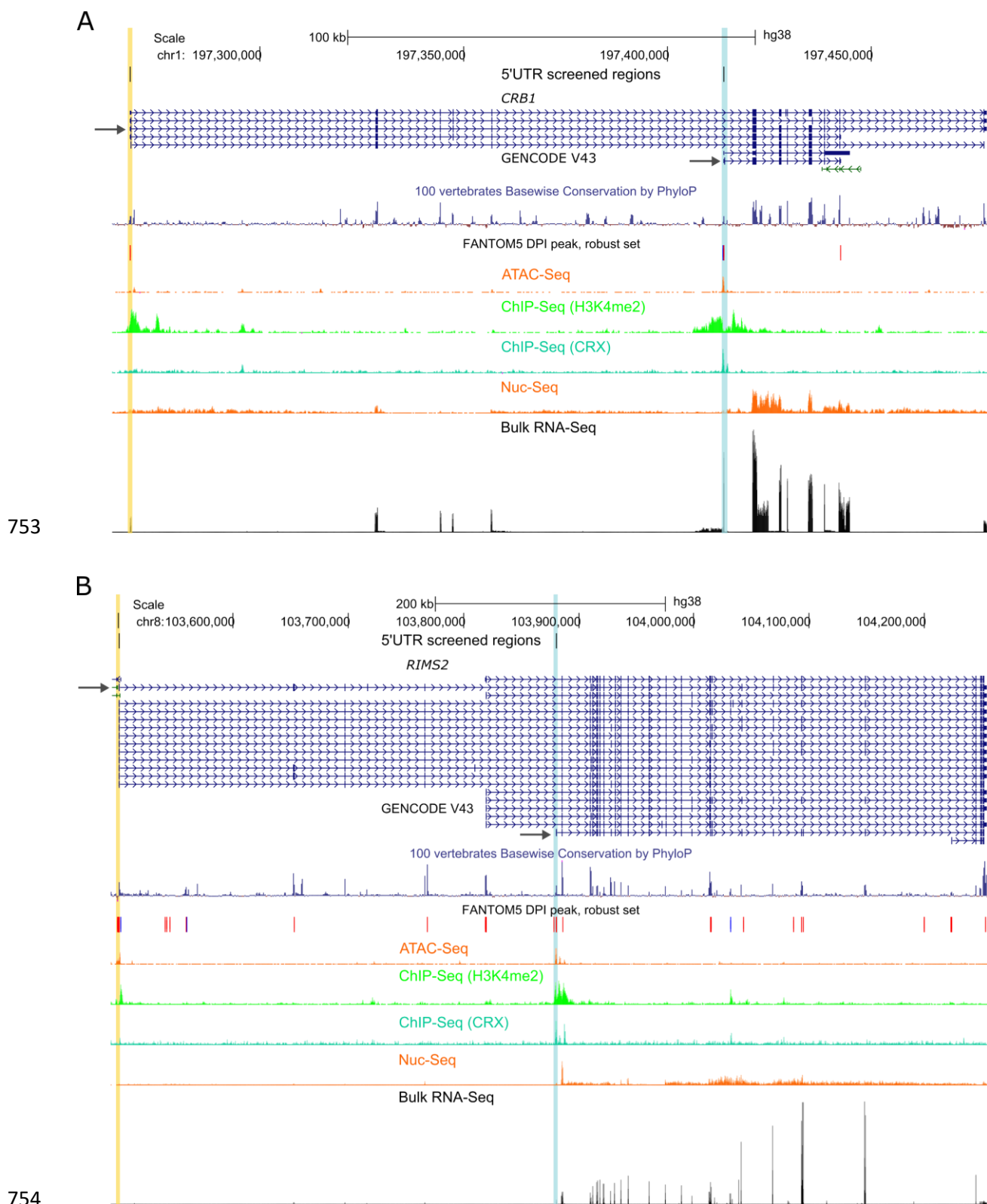
740

741 **Figure S1. Schematic overview of constructs used for functional studies.** (A) The wild-type (WT)
 742 5'UTRs of the *MERTK*, *PAX6*, and *RDH12* genes were cloned in-frame into a psiCHECK™-2 dual luciferase
 743 vector (Promega), containing the *Renilla* and *Firefly* luciferase reporter genes under the regulation of
 744 the SV40 and HSV-TK promoters, respectively. (B). Depiction of the overexpression RDH12 construct
 745 comprising the 5'UTR and primary open reading frame (pORF) of *RDH12*, for which Myc and FLAG in-
 746 frame tags were included downstream, respectively. This fragment was cloned into a pcDNA™3.1⁽⁺⁾
 747 (Invitrogen) vector downstream its T7 promoter. For all constructs, 5'UTRs variants were created by
 748 site-directed mutagenesis to obtain mutant (MT) constructs with the variants of interest.



749

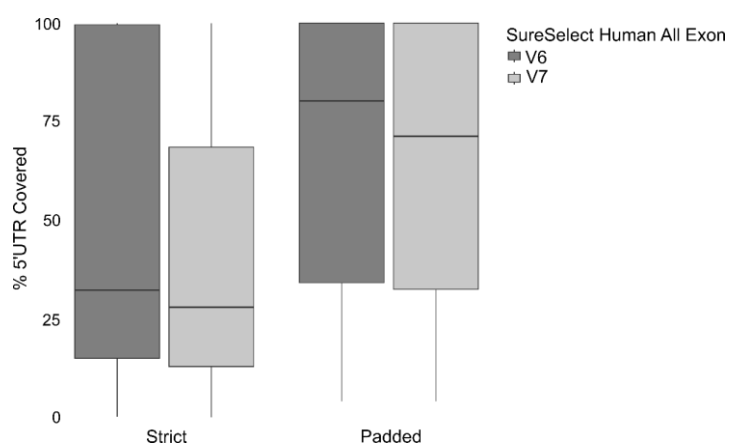
750 **Figure S2. Assessment of transcription start site (TSS) confidence by CAGE-seq in retina.**
 751 Representation of the proportion of IRD gene isoforms based on the support of their annotated TSS
 752 provided by CAGE-seq data derived from adult and fetal retina.



755 **Figure S3. Retina-enriched non-canonical isoforms of the *CRB1* and *RIMS2* genes.** The shorter
756 alternative isoforms of (A) *CRB1* (ENST00000681519) and (B) *RIMS2* (ENST00000436393) contain
757 5'UTRs which are completely distinct to those of their respective canonical isoforms

758 (ENST00000367400.8 and ENST00000696799.1, respectively). The transcription start sites (supported
759 by CAGE-seq) of each canonical and non-canonical isoform (indicated by arrows) are highlighted in
760 *yellow* and *blue*, respectively. Active retinal transcription is supported by bulk RNA-seq and Nuc-seq
761 derived from human retina. The enrichment in retina of these isoforms is further supported by
762 signatures of open chromatin, H3K4me2, and binding of the retina-specific transcription factor CRX
763 (datasets used are listed in **Table S2**).

764



765

766 **Figure S4. Evaluation of 5'UTR capture by the kits mostly used for generating our in-house WES data.**

767 Comparison of the capture performance of the SureSelect Human All Exon V6 and V7 kits (Agilent

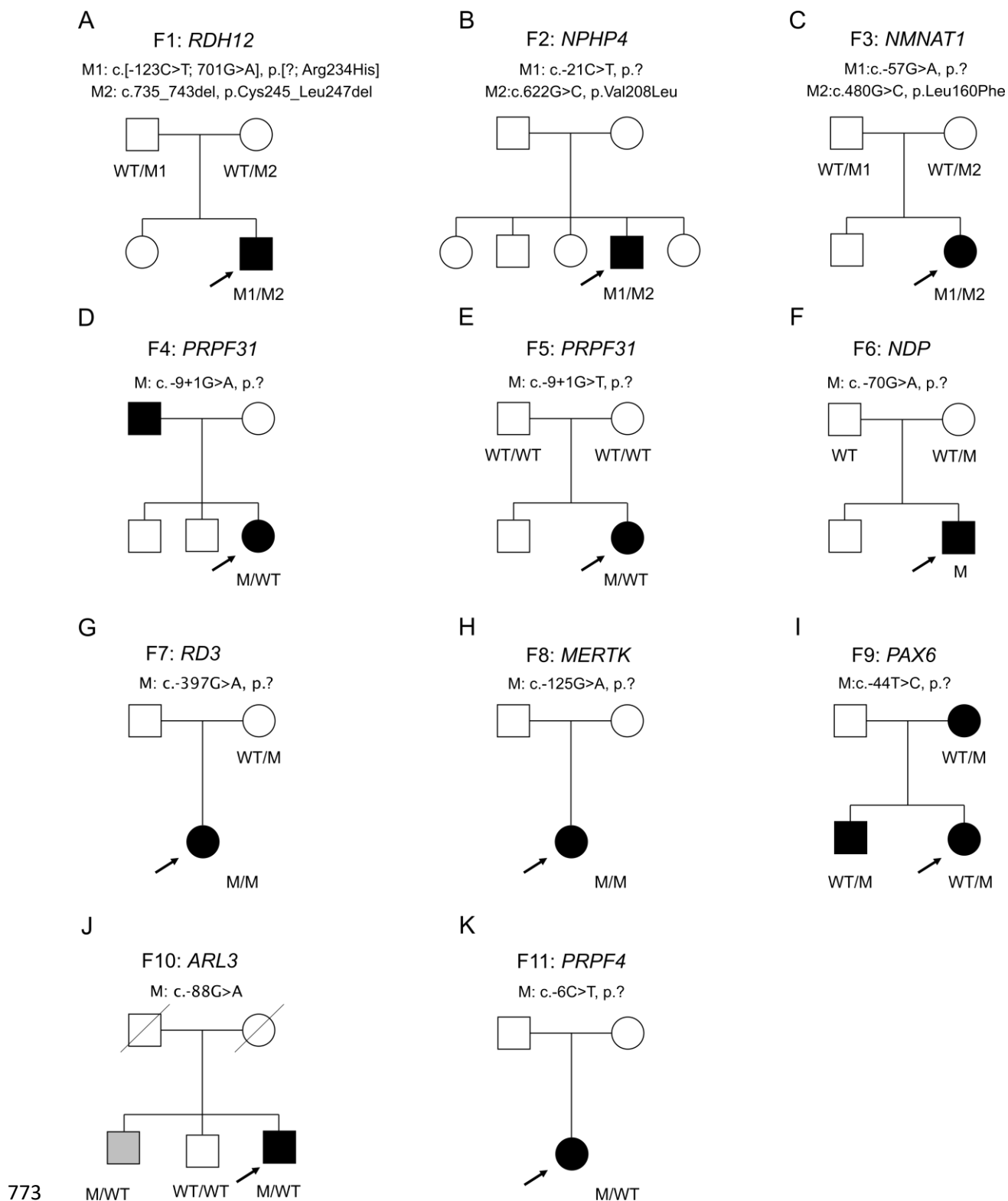
768 Technologies) considering *strict* or *padded* designs (*see Methods*).

769

770

771

772

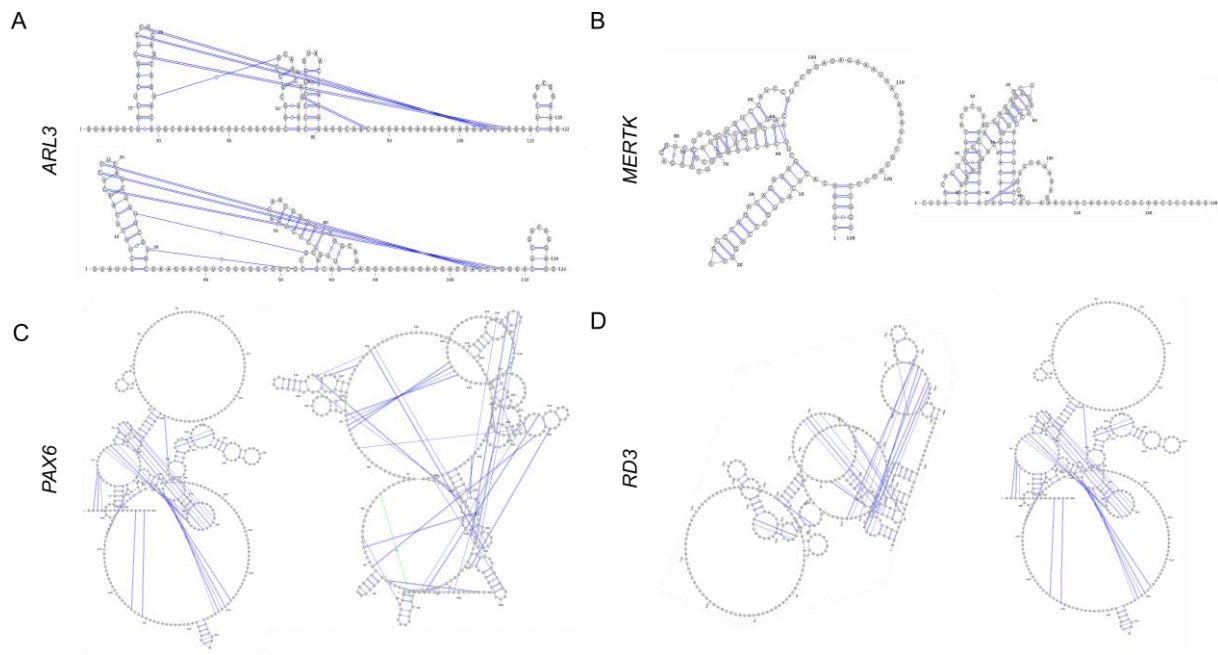


773

774 **Figure S5. Pedigrees of families carrying the 11 candidate 5'UTR variants and segregation analysis.**

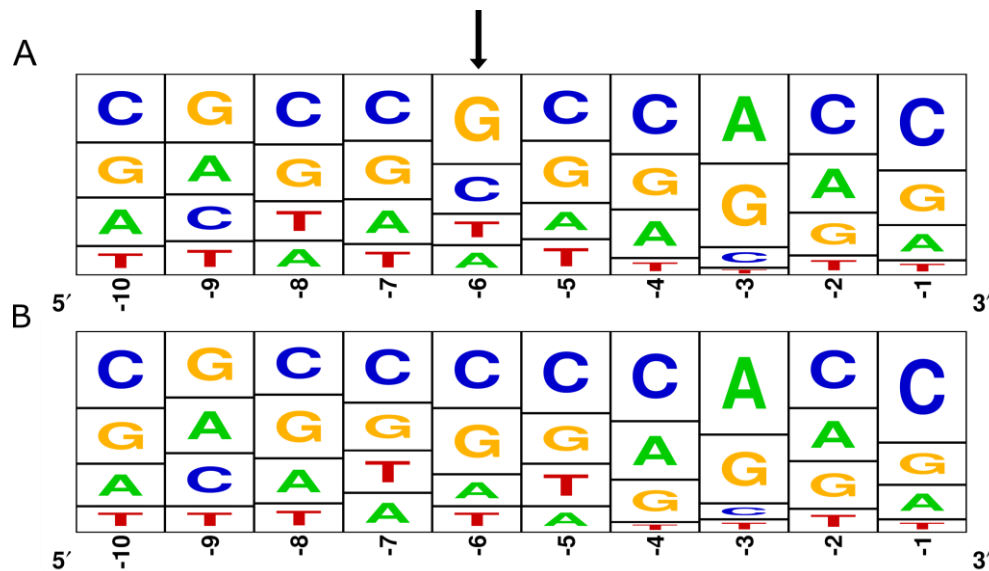
775 (J) Upon clinical re-evaluation, the sibling of F10 (*gray*) was found to be affected by an acquired

776 vascular ocular condition instead of an IRD. Abbreviations: WT, wild-type allele; M, mutated allele.



777

778 **Figure S6. Secondary structure prediction for wild-type and mutated 5'UTR sequences.** Secondary
 779 structure analysis of the 5'UTR sequences (*Ufold* using default parameters) revealed different folding
 780 for the (A) *ARL3:c.-88G>A*, (B) *MERTK:c.-125G>A*, (C) *PAX6:c.-44T>C*, (D) *RD3:c.-394G>A* variants in
 781 comparison to their corresponding wild-type 5'UTRs.



782

783 **Figure S7. Kozak consensus sequence of all IRD gene isoforms evaluated in this study.** The nucleotides
 784 within the -1 to -10 positions relative to the main AUG were retrieved for all selected transcripts.
 785 Frequency plots corresponding to the Kozak sequences of the (A) retina-enriched (373) and (B) not
 786 retina-enriched canonical transcripts (76).

787 **Tables**

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

789 Abbreviations: AD, autosomal dominant; AR, autosomal recessive; CMGG, Center for Medical

790 Genetics Ghent; GE, Genomics England; LCA, Leber Congenital Amaurosis; MAF, minor allele

791 frequency; TSS, transcription start site; uORF, upstream open reading frame.

792

793 **Additional files**

794 **Additional file 1 - Supplementary Tables.**

795 **Table S1.** Custom diagnostic gene panel comprising all IRD genes listed in either the Retinal

796 disorders panel (v2.195) from Genomics England PanelApp or RetNet. Abbreviations: AD,

797 autosomal dominant; AR, autosomal recessive; HGNC, HUGO Gene Nomenclature Committee.

798 **Table S2.** Publicly available multi-omics datasets derived from human retina used in this study.

799 **Table S3.** Overview of the IRD sub-cohort from the Rare Disease arm of the 100,000 Genomes

800 Project (Genomics England).

801 **Table S4.** Primer sequences used in this study.

802 **Table S5.** Overview of selected canonical and non-canonical protein-coding transcripts of IRD

803 genes and their expression in adult human retina. Abbreviations: MANE, Matched Annotation

804 from NCBI and EMBL-EBI; SD, standard deviation; TPM, transcripts per million; TSS,

805 transcription start site.

806 **Table S6.** Genomic coordinates (GRCh38) of the regions in which variants were searched in

807 this study (*5'UTR analysis file*).

808 **Table S7.** Overview of the structure of 5'UTRs of the transcripts included in this study.

809 **Table S8.** Comparison of 5'UTRs of non-canonical transcripts with respect to their
810 corresponding canonical isoforms.

811 **Table S9.** Overview of the performance of commercial exome capture designs on the 5'UTRs
812 screened in this study.

813 **Table S10.** Overview of the performance of commercial exome capture kits mostly used for
814 the generation of our in-house WES data considering strict and padded designs.

815 **Table S11.** Overview of all 5'UTR variants reported in IRD genes submitted to the ClinVar
816 database.

817 **Table S12.** Overview of the span of internal ribosomal entry sites (IRES) contained in the 5'UTR
818 of IRD genes.

819 **Table S13.** Illustration of the designed prioritization strategy applied to a selection of the
820 identified variants. Variants which did not pass all criteria (*strikethrough*) were not prioritized
821 as candidates (*bold*).

822 **Table S14.** Overview of additional variants reported for patients F5, F8, F11. Abbreviations:
823 ACMG, American College of Medical Genetics and Genomics; MAF, minor allele frequency.

824 **Table S15.** Overview of the 11 patients carrying the *RDH12*:c.701G>A (p.Arg234His) variant
825 found *in cis* with *RDH12*:c.-123C>T. Abbreviations: ARG: Argentina; CMGG, Center for Medical
826 Genetics Ghent; GE, Genomics England; HU FJD: Hospital Universitario Fundación Jiménez
827 Díaz; IOB: Institute of Molecular and Clinical Ophthalmology Basel; MEH: Moorfields Eye
828 Hospital.

829 **Additional file 2 - Dataset S1.** Annotation of the 1,450 5'UTRs variants that remained after
830 functional classification and filtering. For each cohort (GE or CMGG) there is one tab per
831 functional category.

832

833 **Declarations**

834 **Ethics approval and consent to participate**

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

840 **Consent for publication**

841 Not applicable. We present only de-identified data.

842 **Availability of data and materials**

843 The data that support the findings of this study are available within the Genomics England
844 (protected) Research Environment but restrictions apply to the availability of these data, as
845 access to the Research Environment is limited to protect the privacy and confidentiality of
846 participants. Likewise, healthcare and genomic data derived from individuals included in the
847 Center for Medical Genetics Ghent (CMGG) cohort are not publicly available to comply with
848 the consent given by those participants. De-identified data as well as analysis scripts are
849 available from the authors upon reasonable request. Extended data generated in this study
850 are available in the supplementary materials.

851 **Competing interests**

852 The authors declare that they have no competing interests.

853 **Funding**

854 This work was supported by the Ghent University Special Research Fund (BOF20/GOA/023;
855 BOF/STA/201909/016) (EDB, BPL, FCP); H2020 Marie Skłodowska-Curie Innovative Training
856 Networks (ITN) StarT (grant No. 813490) (ADR, EDB, FC); Ghent University Hospital under the

857 NucleUZ Grant (EDB, FC); Foundation Fighting Blindness (TA-GT-0621-0810-UGENT) (FCP);
858 EJPRD19-234 Solve-RET (EDB); Fundación Alfonso Martín Escudero (MDPV); Instituto de Salud
859 Carlos III (ISCIII) of the Spanish Ministry of Health (CA; FIS: PI22/00321); University Chair UAM-
860 IIS-FJD of Genomic Medicine (CA). GA is funded by a Fight For Sight UK Early Career
861 Investigator Award (5045/46), National Institute of Health Research Biomedical Research
862 Centre (NIHR-BRC) at Moorfields Eye Hospital and UCL Institute of Ophthalmology and NIHR-
863 BRC at Great Ormond Street Hospital Institute for Child Health.

864 **Authors' contributions**

865 ADR: Conception and project design, acquisition of data, analysis and interpretation of data,
866 drafting and revising the manuscript.

867 MDPV: Acquisition of data, analysis and interpretation of data, drafting and revising the
868 manuscript.

869 MB: Acquisition of data, analysis and interpretation of data, revising the manuscript.

870 FVB: Acquisition of data, revising the manuscript.

871 MDV: Acquisition of data, revising the manuscript.

872 MVH: Acquisition of data, revising the manuscript.

873 MDB: Acquisition of data, revising the manuscript.

874 SVS: Acquisition of data, revising the manuscript.

875 MBW: Acquisition of data, revising the manuscript.

876 JE: Acquisition of data, revising the manuscript.

877 GER: Acquisition of data, revising the manuscript.

878 CR: Acquisition of data, revising the manuscript.

879 AW: Acquisition of data, revising the manuscript.

880 GA: Acquisition of data, revising the manuscript.

881 CA: Acquisition of data, revising the manuscript.

882 JDZ: Acquisition of data, revising the manuscript.

883 BL: Acquisition of data, revising the manuscript.

884 EDB: Project supervision, acquisition of data, revising the manuscript.

885 FC: Conception and project supervision, acquisition of data, analysis and interpretation of
886 data, revising the manuscript.

887 **Acknowledgements**

888 This research was made possible through access to the data and findings generated by the
889 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England
890 Limited (a wholly owned company of the Department of Health and Social Care). The 100,000
891 Genomes Project is funded by the National Institute for Health Research and NHS England.
892 The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded
893 research infrastructure. The 100,000 Genomes Project uses data provided by patients and
894 collected by the National Health Service as part of their care and support. We thank Olga Zurita
895 for their excellent technical assistance. We are grateful to Dr. Jacqueline Cook and Dr. Savita
896 Madhusudhan for their clinical input.

897

898

899

900

901

902

903 References

- 904 1. Jackson RJ, Hellen CUT, Pestova T V. THE MECHANISM OF EUKARYOTIC TRANSLATION
905 INITIATION AND PRINCIPLES OF ITS REGULATION. doi:10.1038/nrm2838
- 906 2. Araujo PR, Yoon K, Ko D, et al. Before It Gets Started: Regulating Translation at the 5' UTR.
907 *Comp Funct Genomics*. 2012;2012. doi:10.1155/2012/475731
- 908 3. Kozak M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc*
909 *Natl Acad Sci U S A*. 1986;83(9):2850-2854. doi:10.1073/PNAS.83.9.2850
- 910 4. Sonenberg N, Hinnebusch AG. Regulation of Translation Initiation in Eukaryotes: Mechanisms
911 and Biological Targets. *Cell*. 2009;136(4):731. doi:10.1016/J.CELL.2009.01.042
- 912 5. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread
913 reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S*
914 *A*. 2009;106(18):7507. doi:10.1073/PNAS.0810916106
- 915 6. Leppek K, Das R, Barna M. Functional 5' UTR mRNA structures in eukaryotic translation
916 regulation and how to find them HHS Public Access. *Nat Rev Mol Cell Biol*. 2018;19(3):158-
917 174. doi:10.1038/nrm.2017.103
- 918 7. Bugaut A, Balasubramanian S. 5'-UTR RNA G-quadruplexes: translation regulation and
919 targeting. doi:10.1093/nar/gks068
- 920 8. Steri M, Idda ML, Whalen MB, Orrù V. Genetic Variants in mRNA Untranslated Regions. *Wiley*
921 *Interdiscip Rev RNA*. 2018;9(4):e1474. doi:10.1002/WRNA.1474
- 922 9. Barbosa C, Peixeiro I, Romão L. Gene Expression Regulation by Upstream Open Reading
923 Frames and Human Disease. *PLOS Genet*. 2013;9(8):e1003529.
924 doi:10.1371/JOURNAL.PGEN.1003529
- 925 10. Soukarieh O, Meguerditchian C, Proust C, et al. Common and Rare 5'UTR Variants Altering
926 Upstream Open Reading Frames in Cardiovascular Genomics. *Front Cardiovasc Med*.
927 2022;9:542. doi:10.3389/FCVM.2022.841032/BIBTEX
- 928 11. Whiffin N, Karczewski KJ, Zhang X, et al. Characterising the loss-of-function impact of 5'
929 untranslated region variants in 15,708 individuals. *Nat Commun* 2020 111. 2020;11(1):1-12.
930 doi:10.1038/s41467-019-10717-9
- 931 12. Smedley D, Smith KR, Martin A, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in
932 Health Care — Preliminary Report. *N Engl J Med*. 2021;385(20):1868-1880.
933 doi:10.1056/nejmoa2035790
- 934 13. Wright CF, FitzPatrick DR, Firth H V. Paediatric genomics: diagnosing rare disease in children.
935 *Nat Rev Genet* 2018 195. 2018;19(5):253-268. doi:10.1038/nrg.2017.116
- 936 14. Fernández-Marmiesse A, Gouveia S, Couce ML. NGS Technologies as a Turning Point in Rare
937 Disease Research, Diagnosis and Treatment. *Curr Med Chem*. 2018;25(3):404.
938 doi:10.2174/0929867324666170718101946
- 939 15. Carss K, Arno G, Erwood M, et al. Comprehensive Rare Variant Analysis via Whole-Genome
940 Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum*
941 *Genet*. 2017;100(1):75-90. doi:10.1016/J.AJHG.2016.12.003
- 942 16. Retterer K, Juusola J, Cho MT, et al. Clinical application of whole-exome sequencing across
943 clinical indications. *Genet Med* 2016 187. 2015;18(7):696-704. doi:10.1038/gim.2015.148

- 944 17. Lionel AC, Costain G, Monfared N, et al. Improved diagnostic yield compared with targeted
945 gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic
946 test. *Genet Med*. 2018;20(4):435-443. doi:10.1038/GIM.2017.119
- 947 18. Dai P, Honda A, Ewans L, et al. Recommendations for next generation sequencing data
948 reanalysis of unsolved cases with suspected Mendelian disorders: A systematic review and
949 meta-analysis. *Genet Med*. 2022;24(8):1618-1629. doi:10.1016/J.GIM.2022.04.021
- 950 19. Wright CF, Quaife NM, Ramos-Hernández L, et al. Non-coding region variants upstream of
951 MEF2C cause severe developmental disorder through three distinct loss-of-function
952 mechanisms. *Am J Hum Genet*. 2021;108(6):1083. doi:10.1016/J.AJHG.2021.04.025
- 953 20. Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5'untranslated region
954 variants with the UTRannotator. *Bioinformatics*. 2021;37(8):1171-1173.
955 doi:10.1093/BIOINFORMATICS/BTAA783
- 956 21. Wang J, Gribskov M. IRESpy: An XGBoost model for prediction of internal ribosome entry
957 sites. *BMC Bioinformatics*. 2019;20(1):1-15. doi:10.1186/S12859-019-2999-7/FIGURES/10
- 958 22. Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol*
959 *Cell*. 2009;101(5):251-262. doi:10.1042/BC20080104
- 960 23. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting Splicing from
961 Primary Sequence with Deep Learning. *Cell*. 2019;176(3):535-548.e24.
962 doi:10.1016/J.CELL.2018.12.015
- 963 24. Zhao J, Wu J, Xu T, Yang Q, He J, Song X. IRESfinder: Identifying RNA internal ribosome entry
964 site in eukaryotic cell using framed k-mer features. *J Genet Genomics*. 2018;45(7):403-406.
965 doi:10.1016/J.JGG.2018.07.006
- 966 25. Lim Y, Arora S, Schuster SL, et al. Multiplexed functional genomic analysis of 5' untranslated
967 region mutations across the spectrum of prostate cancer. *Nat Commun* 2021 121.
968 2021;12(1):1-18. doi:10.1038/s41467-021-24445-6
- 969 26. Filatova A, Reveguk I, Piatkova M, et al. Annotation of uORFs in the OMIM genes allows to
970 reveal pathogenic variants in 5'UTRs. *Nucleic Acids Res*. 2023;51(3):1229-1244.
971 doi:10.1093/nar/gkac1247
- 972 27. Ellingford JM, Ahn JW, Bagnall RD, et al. Recommendations for clinical interpretation of
973 variants found in non-coding regions of the genome. *Genome Med*. 2022;14(1):1-19.
974 doi:10.1186/S13073-022-01073-3/FIGURES/3
- 975 28. Pontikos N, Arno G, Jurkute N, et al. Genetic Basis of Inherited Retinal Disease in a Molecularly
976 Characterized Cohort of More Than 3000 Families from the United Kingdom. Published online
977 2020. doi:10.1016/j.ophtha.2020.04.008
- 978 29. Ratnapriya R, Swaroop A. Genetic architecture of retinal and macular degenerative diseases:
979 the promise and challenges of next-generation sequencing. *Genome Med*. 2013;5(10).
980 doi:10.1186/GM488
- 981 30. O'sullivan J, Mullaney BG, Bhaskar SS, et al. A paradigm shift in the delivery of services for
982 diagnosis of inherited retinal disease. doi:10.1136/jmedgenet-2012-100847
- 983 31. Dockery A, Whelan L, Humphries P, Farrar GJ, Malerba G. Next-Generation Sequencing
984 Applications for Inherited Retinal Diseases. Published online 2021. doi:10.3390/ijms22115684
- 985 32. Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and
986 vitreoretinal diseases. *Prog Retin Eye Res*. 2010;29(5):335-375.

- 987 doi:10.1016/J.PRETEYERES.2010.03.004
- 988 33. Hanany M, Rivolta C, Sharon D. Worldwide carrier frequency and genetic prevalence of
989 autosomal recessive inherited retinal diseases. *Proc Natl Acad Sci U S A*. 2020;117(5):2710-
990 2716. doi:10.1073/PNAS.1913179117
- 991 34. RetNet - Retinal Information Network. Accessed April 6, 2023.
992 <https://web.sph.uth.edu/RetNet/>
- 993 35. Tatour Y, Ben-Yosef T. Syndromic Inherited Retinal Diseases: Genetic, Clinical and Diagnostic
994 Aspects. *Diagnostics (Basel, Switzerland)*. 2020;10(10). doi:10.3390/DIAGNOSTICS10100779
- 995 36. Ellingford JM, Hufnagel RB, Arno G. Phenotype and Genotype Correlations in Inherited Retinal
996 Diseases: Population-Guided Variant Interpretation, Variable Expressivity and Incomplete
997 Penetrance. *Genes (Basel)*. 2020;11(11):1-4. doi:10.3390/GENES11111274
- 998 37. Sharon D, Ben-Yosef T, Goldenberg-Cohen N, et al. A nationwide genetic analysis of inherited
999 retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC).
1000 *Hum Mutat*. 2020;41(1):140-149. doi:10.1002/HUMU.23903
- 1001 38. Perea-Romero I, Gordo G, Iancu IF, et al. Genetic landscape of 6089 inherited retinal
1002 dystrophies affected cases in Spain and their therapeutic and extended epidemiological
1003 implications. *Sci Rep*. 2021;11(1). doi:10.1038/S41598-021-81093-Y
- 1004 39. Fadaie Z, Whelan L, Ben-Yosef T, et al. Whole genome sequencing and in vitro splice assays
1005 reveal genetic causes for inherited retinal diseases. *NPJ genomic Med*. 2021;6(1).
1006 doi:10.1038/S41525-021-00261-1
- 1007 40. Lenassi E, Clayton-Smith J, Douzgou S, et al. Clinical utility of genetic testing in 201 preschool
1008 children with inherited eye disorders. *Genet Med*. 2020;22(4):745-751. doi:10.1038/S41436-
1009 019-0722-8
- 1010 41. Lam BL, Leroy BP, Black G, Ong T, Yoon D, Trzuppek K. Genetic testing and diagnosis of
1011 inherited retinal diseases. *Orphanet J Rare Dis*. 2021;16(1):1-9. doi:10.1186/S13023-021-
1012 02145-0/TABLES/2
- 1013 42. Liquori A, Vaché C, Baux D, et al. Whole USH2A Gene Sequencing Identifies Several New Deep
1014 Intronic Mutations. *Hum Mutat*. 2016;37(2):184-193. doi:10.1002/HUMU.22926
- 1015 43. Holtan JP, Selmer KK, Heimdal KR, Bragadóttir R. Inherited retinal disease in Norway - a
1016 characterization of current clinical and genetic knowledge. *Acta Ophthalmol*. 2020;98(3):286-
1017 295. doi:10.1111/AOS.14218
- 1018 44. Reurink J, Weisschuh N, Garanto A, et al. Whole genome sequencing for USH2A-associated
1019 disease reveals several pathogenic deep-intronic variants that are amenable to splice
1020 correction. *HGG Adv*. 2023;4(2). doi:10.1016/J.XHGG.2023.100181
- 1021 45. Bauwens M, Garanto A, Sangermano R, et al. ABCA4-associated disease as a model for missing
1022 heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory,
1023 structural, and recurrent hypomorphic variants. *Genet Med*. 2019;21(8):1761-1771.
1024 doi:10.1038/S41436-018-0420-Y
- 1025 46. Sangermano R, Garanto A, Khan M, et al. Deep-intronic ABCA4 variants explain missing
1026 heritability in Stargardt disease and allow correction of splice defects by antisense
1027 oligonucleotides. *Genet Med*. 2019;21(8):1751-1760. doi:10.1038/S41436-018-0414-9
- 1028 47. Khan M, Cornelis SS, Pozo-Valero M Del, et al. Resolving the dark matter of ABCA4 for 1054
1029 Stargardt disease probands through integrated genomics and transcriptomics. *Genet Med*.

- 1030 2020;22(7):1235-1246. doi:10.1038/S41436-020-0787-4
- 1031 48. Daich Varela M, Bellingham J, Motta F, et al. Multidisciplinary team directed analysis of whole
1032 genome sequencing reveals pathogenic non-coding variants in molecularly undiagnosed
1033 inherited retinal dystrophies. *Hum Mol Genet.* 2023;32(4):595-607.
1034 doi:10.1093/HMG/DDAC227
- 1035 49. Qian X, Wang J, Wang M, et al. Identification of Deep-Intronic Splice Mutations in a Large
1036 Cohort of Patients With Inherited Retinal Diseases. *Front Genet.* 2021;12:276.
1037 doi:10.3389/FGENE.2021.647400/BIBTEX
- 1038 50. Weisschuh N, Sturm M, Baumann B, et al. Deep-intronic variants in CNGB3 cause
1039 achromatopsia by pseudoexon activation HHS Public Access. *Hum Mutat.* 2020;41(1):255-264.
1040 doi:10.1002/humu.23920
- 1041 51. Jamshidi F, Place EM, Mehrotra S, et al. Contribution of noncoding pathogenic variants to
1042 RPGRIP1-mediated inherited retinal degeneration. *Genet Med.* 2019;21(3):694-704.
1043 doi:10.1038/S41436-018-0104-7
- 1044 52. Cherry TJ, Yang MG, Harmin DA, et al. Mapping the cis-regulatory architecture of the human
1045 retina reveals noncoding genetic variation in disease. *Proc Natl Acad Sci U S A.*
1046 2020;117(16):9001-9012. doi:10.1073/PNAS.1922501117/-/DCSUPPLEMENTAL
- 1047 53. Coppieters F, Todeschini AL, Fujimaki T, et al. Hidden Genetic Variation in LCA9-Associated
1048 Congenital Blindness Explained by 5'UTR Mutations and Copy-Number Variations of NMNAT1.
1049 *Hum Mutat.* 2015;36(12):1188. doi:10.1002/HUMU.22899
- 1050 54. Ruberto FP, Balzano S, Namburi P, et al. Heterozygous deletions of noncoding parts of the
1051 PRPF31 gene cause retinitis pigmentosa via reduced gene expression. *Mol Vis.* 2021;27:107.
1052 Accessed April 11, 2023. /pmc/articles/PMC8056469/
- 1053 55. Van de Sompele S, Small KW, Cicekdağ MB, et al. Multi-omics approach dissects cis-regulatory
1054 mechanisms underlying North Carolina macular dystrophy, a retinal enhanceropathy. *Am J*
1055 *Hum Genet.* 2022;109(11):2029-2048. doi:10.1016/J.AJHG.2022.09.013
- 1056 56. Small KW, DeLuca AP, Whitmore SS, et al. North Carolina Macular Dystrophy is caused by
1057 dysregulation of the retinal transcription factor PRDM13. *Ophthalmology.* 2016;123(1):9.
1058 doi:10.1016/J.OPHTHA.2015.10.006
- 1059 57. Filatova AY, Vasilyeva TA, Marakhonov A V., et al. Upstream ORF frameshift variants in the
1060 PAX6 5'UTR cause congenital aniridia. *Hum Mutat.* 2021;42(8):1053-1065.
1061 doi:10.1002/HUMU.24248
- 1062 58. Zuercher J, Neidhardt J, Magyar I, et al. Alterations of the 5'Untranslated Region of SLC16A12
1063 Lead to Age-Related Cataract. *Invest Ophthalmol Vis Sci.* 2010;51(7):3354.
1064 doi:10.1167/IOVS.10-5193
- 1065 59. Zhang W, Kassels AC, Barrington A, et al. Macular corneal dystrophy with isolated peripheral
1066 Descemet membrane deposits. *Am J Ophthalmol Case Reports.* 2019;16.
1067 doi:10.1016/J.AJOC.2019.100571
- 1068 60. Ratnapriya R, Sosina OA, Starostik MR, et al. Retinal transcriptome and eQTL analyses identify
1069 genes associated with age-related macular degeneration. *Nat Genet.* 2019;51(4):606-610.
1070 doi:10.1038/S41588-019-0351-9
- 1071 61. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification.
1072 *Nat Biotechnol.* 2016;34(5):525-527. doi:10.1038/NBT.3519

- 1073 62. Morales J, Pujar S, Loveland JE, et al. A joint NCBI and EMBL-EBI transcript set for clinical
1074 genomics and research. *Nature*. 2022;604(7905). doi:10.1038/S41586-022-04558-8
- 1075 63. Martin AR, Williams E, Foulger RE, et al. PanelApp crowdsources expert knowledge to
1076 establish consensus diagnostic gene panels. *Nat Genet*. 2019;51(11):1560-1565.
1077 doi:10.1038/S41588-019-0528-2
- 1078 64. Forrest ARR, Kawaji H, Rehli M, et al. A promoter-level mammalian expression atlas. *Nature*.
1079 2014;507(7493):462-470. doi:10.1038/NATURE13182
- 1080 65. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
1081 *Bioinforma Appl NOTE*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033
- 1082 66. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically
1083 relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862-D868. doi:10.1093/NAR/GKV1222
- 1084 67. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
1085 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324
- 1086 68. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework
1087 for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.
1088 doi:10.1101/GR.107524.110
- 1089 69. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience*.
1090 2021;10(2). doi:10.1093/GIGASCIENCE/GIAB008
- 1091 70. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*.
1092 2016;17(1). doi:10.1186/S13059-016-0974-4
- 1093 71. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from
1094 variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/S41586-020-
1095 2308-7
- 1096 72. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in
1097 the human genome. *Nucleic Acids Res*. 2014;42(22):13534-13544. doi:10.1093/NAR/GKU1206
- 1098 73. Yeo G, Burge CB. Maximum Entropy Modeling of Short Sequence Motifs with Applications to
1099 RNA Splicing Signals. <https://home.liebertpub.com/cmb>. 2004;11(2-3):377-394.
1100 doi:10.1089/1066527041410418
- 1101 74. Frazer J, Notin P, Dias M, et al. Disease variant prediction with deep generative models of
1102 evolutionary data. *Nat 2021 5997883*. 2021;599(7883):91-95. doi:10.1038/s41586-021-04043-
1103 8
- 1104 75. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for
1105 estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-
1106 315. doi:10.1038/NG.2892
- 1107 76. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. The ensembl regulatory build.
1108 *Genome Biol*. 2015;16(1):56-56. doi:10.1186/S13059-015-0621-5
- 1109 77. Kozak M. The scanning model for translation: an update. *J Cell Biol*. 1989;108(2):229-241.
1110 doi:10.1083/JCB.108.2.229
- 1111 78. McClements ME, Butt A, Piotter E, Peddle CF, Maclaren RE. An analysis of the Kozak
1112 consensus in retinal genes and its relevance to gene therapy. *Mol Vis*. 2021;27:233. Accessed
1113 March 15, 2023. /pmc/articles/PMC8116250/
- 1114 79. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome*

- 1115 Res. 2004;14(6):1188-1190. doi:10.1101/GR.849004
- 1116 80. Sample PJ, Wang B, Reid DW, et al. Human 5' UTR design and variant effect prediction from a
1117 massively parallel translation assay. *Nat Biotechnol* 2019 377. 2019;37(7):803-809.
1118 doi:10.1038/s41587-019-0164-5
- 1119 81. Murat P, Marsico G, Herdy B, Ghanbarian A, Portella G, Balasubramanian S. RNA G-
1120 quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent
1121 translation of human mRNAs. *Genome Biol.* 2018;19(1):1-24. doi:10.1186/S13059-018-1602-
1122 2/FIGURES/6
- 1123 82. Li H. Tabix: fast retrieval of sequence features from generic TAB-delimited files.
1124 *Bioinformatics.* 2011;27(5):718. doi:10.1093/BIOINFORMATICS/BTQ671
- 1125 83. Fu L, Cao Y, Wu J, Peng Q, Nie Q, Xie X. Ufold: fast and accurate RNA secondary structure
1126 prediction with deep learning. *Nucleic Acids Res.* 2022;50(3):e14-e14.
1127 doi:10.1093/NAR/GKAB1074
- 1128 84. Chen CC, Chan YM. REDfold: accurate RNA secondary structure prediction using residual
1129 encoder-decoder network. *BMC Bioinformatics.* 2023;24(1):122. doi:10.1186/S12859-023-
1130 05238-8/FIGURES/3
- 1131 85. Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning
1132 with thermodynamic integration. *Nat Commun* 2021 121. 2021;12(1):1-9.
1133 doi:10.1038/s41467-021-21194-4
- 1134 86. Yang TH, Wang CY, Tsai HC, Liu CT. Human IRES Atlas: an integrative platform for studying
1135 IRES-driven translational regulation in humans. *Database.* 2021;2021:1-16.
1136 doi:10.1093/DATABASE/BAAB025
- 1137 87. Bates D, Mächler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models Using lme4. *J*
1138 *Stat Softw.* 2015;67(1):1-48. doi:10.18637/JSS.V067.I01
- 1139 88. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative
1140 quantification framework and software for management and automated analysis of real-time
1141 quantitative PCR data. Published online 2007. doi:10.1186/gb-2007-8-2-r19
- 1142 89. Dryja TP, McGee TL, Berson EL, et al. Night blindness and abnormal cone electroretinogram
1143 ON responses in patients with mutations in the GRM6 gene encoding mGluR6. *Proc Natl Acad*
1144 *Sci U S A.* 2005;102(13). doi:10.1073/pnas.0501233102
- 1145 90. De Zaeytijd J, Van Cauwenbergh C, De Bruyne M, et al. ISOLATED MACULOPATHY AND
1146 MODERATE ROD-CONE DYSTROPHY REPRESENT THE Milder END OF THE RDH12-RELATED
1147 RETINAL DYSTROPHY SPECTRUM. *Retina.* 2021;41(6):1346-1355.
1148 doi:10.1097/IAE.0000000000003028
- 1149 91. Tiering (Rare Disease) - Genomics England Research Environment - Genomics England
1150 Confluence. Accessed April 14, 2023.
1151 <https://cnfl.extge.co.uk/pages/viewpage.action?pageld=113194832>
- 1152 92. Schmitz-Abe K, Li Q, Rosen SM, et al. Unique bioinformatic approach and comprehensive
1153 reanalysis improve diagnostic yield of clinical exomes. *Eur J Hum Genet* 2019 279.
1154 2019;27(9):1398-1405. doi:10.1038/s41431-019-0401-x
- 1155 93. Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data
1156 yields additional diagnoses: implications for providers. *Genet Med.* 2017;19(2):209-214.
1157 doi:10.1038/GIM.2016.88

- 1158 94. de Bruijn SE, Rodenburg K, Corominas J, et al. Optical genome mapping and revisiting short-
1159 read genome sequencing data reveal previously overlooked structural variants disrupting
1160 retinal disease-associated genes. *Genet Med*. 2023;25(3). doi:10.1016/J.GIM.2022.11.013
- 1161 95. Del Pozo-Valero M, Martin-Merida I, Jimenez-Rolando B, et al. Expanded Phenotypic
1162 Spectrum of Retinopathies Associated with Autosomal Recessive and Dominant Mutations in
1163 PROM1. *Am J Ophthalmol*. 2019;207:204-214. doi:10.1016/J.AJO.2019.05.014
- 1164 96. Moreno-Leon L, West EL, O'Hara-Wright M, et al. RPGR isoform imbalance causes ciliary
1165 defects due to exon ORF15 mutations in X-linked retinitis pigmentosa (XLRP). *Hum Mol Genet*.
1166 2021;29(22):3706-3716. doi:10.1093/HMG/DDAA269
- 1167 97. Vig A, Poulter JA, Ottaviani D, et al. DYNC2H1 hypomorphic or retina-predominant variants
1168 cause nonsyndromic retinal degeneration. *Genet Med*. 2020;22(12):2041-2051.
1169 doi:10.1038/S41436-020-0915-1
- 1170 98. Riazuddin SA, Iqbal M, Wang Y, et al. A splice-site mutation in a retina-specific exon of BBS8
1171 causes nonsyndromic retinitis pigmentosa. *Am J Hum Genet*. 2010;86(5):805-812.
1172 doi:10.1016/J.AJHG.2010.04.001
- 1173 99. Mairot K, Smirnov V, Bocquet B, et al. CRB1-Related Retinal Dystrophies in a Cohort of 50
1174 Patients: A Reappraisal in the Light of Specific Müller Cell and Photoreceptor CRB1 Isoforms.
1175 *Int J Mol Sci*. 2021;22(23):12642. doi:10.3390/IJMS222312642
- 1176 100. Weber R, Ghoshdastider U, Spies D, et al. Monitoring the 5'UTR landscape reveals isoform
1177 switches to drive translational efficiencies in cancer. *Oncogene*. 2023;42(9):638-650.
1178 doi:10.1038/S41388-022-02578-2
- 1179 101. Ray TA, Cochran K, Kozlowski C, et al. Comprehensive identification of mRNA isoforms reveals
1180 the diversity of neural cell-surface molecules with roles in retinal development and disease.
1181 *Nat Commun*. 2020;11(1). doi:10.1038/S41467-020-17009-7
- 1182 102. Yu M, Baehr W, Lei B, et al. A Novel ARL3 Gene Mutation Associated With Autosomal
1183 Dominant Retinal Degeneration. Published online 2021. doi:10.3389/fcell.2021.720782
- 1184 103. Wang J, Wang Y, Li S, et al. Clinical and Genetic Analysis of RDH12-Associated Retinopathy in
1185 27 Chinese Families: A Hypomorphic Allele Leads to Cone-Rod Dystrophy. Published online
1186 2022. doi:10.1167/iovs.63.9.24
- 1187 104. Ba-Abbad R, Arno G, Robson AG, et al. Macula-predominant retinopathy associated with
1188 biallelic variants in RDH12. *Ophthalmic Genet*. 2020;41(6):612-615.
1189 doi:10.1080/13816810.2020.1802763
- 1190 105. Thompson DA, Janecke AR, Lange J, et al. Retinal degeneration associated with RDH12
1191 mutations results from decreased 11- cis retinal synthesis due to disruption of the visual cycle.
1192 *Hum Mol Genet*. 2005;14(24):3865-3875. doi:10.1093/HMG/DDI411
- 1193