

RESEARCH

Open Access



# Long-read sequencing improves the genetic diagnosis of retinitis pigmentosa by identifying an Alu retrotransposon insertion in the *EYS* gene

Elena Fernández-Suárez<sup>1,2†</sup>, María González-del Pozo<sup>1,2†</sup>, Cristina Méndez-Vidal<sup>1,2</sup>, Marta Martín-Sánchez<sup>1,2</sup>, Marcela Mena<sup>1,2</sup>, Belén de la Morena-Barrio<sup>3</sup>, Javier Corral<sup>3</sup>, Salud Borrego<sup>1,2\*</sup> and Guillermo Antiñolo<sup>1,2\*</sup>

## Abstract

**Background** Biallelic variants in *EYS* are the major cause of autosomal recessive retinitis pigmentosa (arRP) in certain populations, a clinically and genetically heterogeneous disease that may lead to legal blindness. *EYS* is one of the largest genes (~2 Mb) expressed in the retina, in which structural variants (SVs) represent a common cause of disease. However, their identification using short-read sequencing (SRS) is not always feasible. Here, we conducted targeted long-read sequencing (T-LRS) using adaptive sampling of *EYS* on the MinION sequencing platform (Oxford Nanopore Technologies) to definitively diagnose an arRP family, whose affected individuals ( $n = 3$ ) carried the heterozygous pathogenic deletion of exons 32–33 in the *EYS* gene. As this was a recurrent variant identified in three additional families in our cohort, we also aimed to characterize the known deletion at the nucleotide level to assess a possible founder effect.

**Results** T-LRS in family A unveiled a heterozygous AluYa5 insertion in the coding exon 43 of *EYS* (chr6(GRCh37):g.64430524\_64430525ins352), which segregated with the disease in compound heterozygosity with the previously identified deletion. Visual inspection of previous SRS alignments using IGV revealed several reads containing soft-clipped bases, accompanied by a slight drop in coverage at the Alu insertion site. This prompted us to develop a simplified program using *grep* command to investigate the recurrence of this variant in our cohort from SRS data. Moreover, LRS also allowed the characterization of the CNV as a ~56.4kb deletion spanning exons 32–33 of *EYS* (chr6(GRCh37):g.64764235\_64820592del). The results of further characterization by Sanger sequencing and linkage analysis in the four families were consistent with a founder variant.

**Conclusions** To our knowledge, this is the first report of a mobile element insertion into the coding sequence of *EYS*, as a likely cause of arRP in a family. Our study highlights the value of LRS technology in characterizing and identifying hidden pathogenic SVs, such as retrotransposon insertions, whose contribution to the etiopathogenesis of rare diseases may be underestimated.

<sup>†</sup>Elena Fernández-Suárez and María González-del Pozo contributed equally to this work and share first authorship.

\*Correspondence:

Salud Borrego

salud.borrego.sspa@juntadeandalucia.es

Guillermo Antiñolo

gantino@us.es

Full list of author information is available at the end of the article



**Keywords** Alu insertion, CNV, *EYS*, Inherited retinal diseases, Long-read sequencing, Nanopore sequencing, Retinitis pigmentosa, Retrotransposon

## Background

Inherited retinal dystrophies (IRD) are a group of clinically and genetically heterogeneous pathologies characterized by photoreceptors or retinal pigment epithelial cell dysfunction leading to irreversible and progressive visual impairment [1]. To date, variants in more than 300 genes and *loci* have been associated with autosomal-recessive, autosomal-dominant, X-linked, and mitochondrial inheritance (<https://web.sph.uth.edu/RetNet/> accessed on October 2023), showing the wide heterogeneity of these disorders [2]. The most common form of IRD is retinitis pigmentosa (RP, ORPHA:791), which affects more than 1.5 million patients worldwide (1:4000) [3]. RP typically manifests with night blindness as the first symptom, reflecting the principal dysfunction of rod photoreceptors, followed by concentric visual field loss, and a decrease in visual acuity due to secondary cone dysfunction [3]. Some of the most commonly mutated genes in non-syndromic RP include *RHO* [4], *USH2A* [5], *EYS* [6–8], or *RPGR* [9].

Currently, short-read sequencing (SRS) is the most commonly used approach to genetically diagnose RP patients in clinical routine, enabling the identification of disease-causing variants in an estimated 60% of cases [10, 11]. Although the application of recent advances, such as whole-genome sequencing (WGS), has the potential to increase the diagnostic yield of IRD patients by facilitating the identification of copy number variations (CNVs), and allowing the detection of variants in GC-rich regions, non-coding regions [12–14], or in novel candidate disease genes [15, 16], a substantial number of patients remain without genetic diagnosis [13, 14]. This may be due to the shortcomings of SRS in detecting complex structural variants (SVs) [17], such as, mobile element insertions (MEIs), inversions, or translocations, which have previously been associated with the etiopathogenesis of the IRD [18–20].

In this sense, the arrival of long-read sequencing (LRS) technologies has raised great expectations about their potential to discover unknown etiological variants. Despite the relatively high sequencing error rate, LRS allows more accurate detection and characterization of SVs, overcoming some of the limitations of SRS [21]. Long-read genome sequencing also offers clear advantages in the detection of the physical phasing of genomes and methylation differences, which are simultaneously detected without needing additional

experiments [22]. However, this approach still has certain disadvantages that hamper its use by human genetics researchers and clinicians, including high costs, low throughput, computational overhead, and the lack of large databases for LRS data interpretation [22, 23]. To reduce costs and simplify the analysis, targeted LRS (T-LRS) has been shown to be effective in identifying missing variants in specific genes of interest [24]. Nanopore sequencing, thanks to its flexibility of data acquisition with real-time analysis, allows target enrichment by directly rejecting or accepting DNA molecules during sequencing without specific sample preparation [25]. This *in silico* enrichment, termed adaptive sampling, avoids the sequencing of uninformative or off-target reads, thereby increasing the depth of coverage of the target region [24, 26]. While targeted LRS is useful for the study of genetic diseases caused by mutations in a single gene [27, 28], for more heterogeneous diseases, like IRD, the selection of partially solved patients carrying a monoallelic likely causative variant in an autosomal recessive gene is key to increase the success rate by focusing the analysis on a single genomic region. In this sense, a good candidate to be explored by LRS would be the *EYS* gene, in which has been described that CNVs are a relatively common type of genomic rearrangement [29–31]. *EYS* is one of the largest genes expressed in the retina, spanning over 2Mb of genomic DNA [6, 7, 32], and is one of the most prevalent genes in autosomal recessive RP (arRP) in diverse populations [7, 8, 33]. Remarkably, in the HGMD-pro database (accessed on October 2023), a large number of pathogenic/likely pathogenic variants ( $n=749$ ) are reported, of which 87 correspond to gross deletions and insertions.

Here, the application of T-LRS using adaptive sampling in a patient with arRP carrying a pathogenic deletion of exons 32–33 in the *EYS* gene allowed us to fulfill the complete molecular diagnosis 4 years after the firsts analysis, because we identified an insertion of 352bp Alu repeat sequence in the coding sequence of exon 43 as a potential second causative variant of arRP in this family. Moreover, T-LRS enabled us to determine that both defects were in different alleles and to define the breakpoints of the aforementioned *EYS* deletion. To our knowledge, this is the first time that a mobile element insertion in the *EYS* gene has been reported as a disease-causing variant, enlarging the number of genes affected by this pathogenic mechanism.

## Materials and methods

### Subjects, clinical evaluation, and previous studies

One Spanish family consisting of 8 unaffected and 3 affected individuals with a presumed arRP, was recruited for genetic diagnosis (Family A). Moreover, two additional genetically solved arRP families (Families B and C) and one unclassified IRD family carrying the heterozygous deletion of exons 32–33 of *EYS* were included in this study. Peripheral blood was collected from the subjects to extract genomic DNA (gDNA) using standard procedures. An informed consent form was signed by all participants or their legal guardians for clinical and genetic studies. Experiments were conducted according to the principles of the Declaration of Helsinki (Edinburgh, 2000) [34], and approved by the Institutional Review Boards of the University Hospital Virgen del Rocio and the University Hospital Virgen Macarena (Seville, Spain).

As part of our diagnostic routine, the proband of family A underwent targeted sequencing of a custom panel that included all coding exons and the splice junctions of 1,166 genes associated with different rare diseases [35, 36]. The sequencing was performed on the NextSeq500 instrument (Illumina, San Diego, CA, USA). Single nucleotide variants (SNVs) and indels were analyzed using a corporate prioritization tool, whereas an in-house independent script based on coverage and statistical studies was used for the analysis of CNVs [35]. Briefly, the prioritization of SNVs and indels was done as following: i) application of a virtual panel of 146 IRD genes; ii) frequency filtering, *minor allele frequency* (MAF) below 0.01 in 1000GP, the Exome Aggregation Consortium (ExAC), the Genome Aggregation Database (GnomAD), Exome Variant Server (EVS); and iii) consequence filtering: coding nonsynonymous variants and splice variants (8 bp intronic and 2bp exonic). As for CNVs, the prioritization of the variant was done according to the highest absolute values of the z-score. In addition, we used the Mobile Element Locator Tool (MELT v2.2.2) [37] to discover mobile element insertions (Alu, L1, and SVA elements).

The CNV affecting *EYS* were previously analyzed and validated by Multiplex Dependent Probe Amplification (MLPA) with SALSA MLPA Probe mix P328-A3 (MRC Holland). The MLPA reactions were run on ABI 3730 DNA Analyzer (Applied Biosystems) and the data was evaluated using GeneMarker v.1.75 (SoftGenetics) as previously described [30].

### Long-read sequencing and data processing

For nanopore sequencing, we used the sequencing service provided by LongSeq Applications (Murcia, Spain) using the MinION device (Oxford Nanopore Technologies). Briefly, approximately 2,305 ng of gDNA was used to prepare the sequencing library using the ONT Ligation

Sequencing Kit (SQK-LSK109) following the manufacturer's protocol with slight modifications. Bead-based washes were performed using Low Fragment Buffer and the final library was eluted in 15 µl of Elution Buffer, following a 10 min incubation at 37°C. Approximately, 200 ng of DNA library were loaded onto a MinION Flow Cell (R9.4.1). *EYS* enrichment was performed using adaptive sampling tool, implemented in the MinKNOW software (ONT) [38], whose input was the FASTA file obtained from this genomic coordinate: chr6:63783736–66808386 (GRCh37/hg19). Sequencing experiment were run for up to ~40h.

Bioinformatic analysis of LRS data was performed as previously described [27] and consisted of: i) base calling using Guppy, which is integrated within the MinKNOW software [38]; ii) alignment to the human reference assembly (GRCh37/hg19) using Minimap2 [39, 40]; iii) variants calling with Sniffles software for SVs [41] and Clair3 for SNVs [42]. The SVs file was annotated using AnnotSV v3.3.6 [43], whereas the VCF file containing SNVs was annotated with Alamut Batch v1.11 and SnpEff v5.1 [44] to add the SpliceAI and CADD v.1.6 scores. The prioritization of SNVs detected by ONT was conducted using our pipeline as previously described [15, 36]. For SVs analysis, variants involving coding exons were prioritized.

### Validations and breakpoints sequence analysis

In the index patients from the four families, the breakpoints of the *EYS* exons 32–33 deletion were assessed by PCR using mutation-specific primers: 5'-CCTTTACAA GACATGAGCATGCTGGGA-3' (intron 33, forward) and 5'-ATTCCTTACTCCCTAGCCCTGCTGTAA-3' (intron 31, reverse). The amplification reaction was performed using Multiplex PCR Master Mix (Qiagen) followed by 35 cycles of 94 °C for 30 s, 60 °C for 90 s, and 72 °C for 90 s. Under these conditions, only the mutant allele can be amplified, as the wild-type allele is larger (~56kb).

The validation of the *EYS*-Alu insertion in family A was performed by PCR using mutation-specific primers designed using the information obtained by nanopore sequencing: 5'-TTTTAGCCGGGATGGTCTCGATCT CC-3' (AluYa5, forward) and 5'-GAGAAACCTCCAGTT CACTACTATATCC-3' (exon 43, reverse) for the 5'-junction and 5'-TGTAGGAAAACAATCAGAACCTT CAGTG-3' (exon 43, forward) and 5'-GGAGATCGA GACCATCCCGGCTAAAA-3' (AluYa5, reverse) for the 3'-junction. The PCR reaction was performed using NXT Taq PCR Master Mix 2X (EURx Ltd.) according to the manufacturer's protocol. The amplification conditions were the following: 95°C for 5 min; 35 cycles of 95°C for 5 s, 60°C for 5 s and 68°C for 40 s; final extension at 68°C for 30 s.

PCR products were analyzed using the QIAxcel capillary electrophoresis system and QIAxcel ScreenGel software (Qiagen) [45]. Sanger sequencing was performed after PCR cleanup (ExoSap-IT, Affymetrix, Santa Clara, CA, USA) and sequenced (BigDye® Terminator v3.1 Cycle Sequencing Kit, 3730 DNA Analyzer, Applied Biosystems, USA)) using the primers described above.

**Linkage analysis by short tandem repeat markers**

To assess a putative founder effect of the CNV-deletion of exons 32–33 of *EYS*, linkage analysis using short tandem repeat (STR) markers was carried out in 19 individuals from the four unrelated families, 14 of which carried the CNV. For this purpose, a total of eight STR markers flanking *EYS* were selected from the literature [32] and UCSC Genome Browser (D6S1573, D6S402, D6S1658, D6S1026, D6S1670, D6S430, D6S1557 and D6S1681). Multiplex PCR Master Mix (Qiagen) was employed to amplify and label the interest regions. PCR products were genotyped using 3730 DNA Analyzer (Applied Biosystems). The results were analyzed by GeneMapper v.4.0 software (Applied Biosystems).

**Screening for EYS-Alu insertion in SRS data**

In order to evaluate the recurrence of the inserted Alu in our population, the Linux grep command (Table 1) was used to search in compressed FASTQ for the junctions between *EYS* exon 43 reference sequences and the beginning/end of the Alu insertions in the previously generated short-read NGS data of 327 additional individuals, including 149 unaffected and 178 IRD patients, as described elsewhere [46, 47]. For this purpose, we used four chimeric sequences of 23 nucleotides in length, containing the two boundaries of the insert (*EYS*-Alu and Alu-*EYS*) at both forward and reverse (Table 1). These sequences were observed by both LRS and Sanger in the index of family A, and were found to be specific for the mutant allele, as they did not match any other region of the genome. The wild-type sequence was also searched to determine the genotype of the Alu insertion. The grep commands returned the number of reads containing the matching sequences in each file, whose value was dependent on the coverage depth in that area. After the

screening, the variant allele frequency (VAF) was calculated as mutant alleles/total alleles.

**Results**

**LRS data quality**

Nanopore sequencing of the proband of family A using a MinION device yielded 8.25Gb with 9,832,520 total reads. After quality analysis by *qualimap*, 936,188 reads remained with a mean read length of 2,876kb and a maximum read length of ~110kb. The percentage of mapped reads across the reference genome was 98.66% with a mean mapping quality of 59.68. The mean coverage in the region of interest (chr6:63,783,736–66,808,386) was 7.0x ± 2.7 (mean ± SD).

**Identification of a novel mobile element insertion**

Index patient from family A received a clinical diagnosis of arRP (Table 2). In previous studies, targeted SRS in the proband of family A only allowed the identification of a heterozygous deletion of exons 32–33 in the *EYS* gene, which was also detected in both affected siblings (Fig. 1). Therefore, it was necessary to identify a second disease-causing variant to complete the genetic diagnosis in this family.

Nanopore sequencing revealed 12 additional SVs and 325 rare (MAF < 0.01) SNVs/indels in *EYS*. Among the SVs, a heterozygous insertion of ~352bp in coding exon 43 was first prioritized in this family because it was the only one with an exonic breakpoint, and it was in a different phase than the previously detected deletion (Fig. 1). LRS provided the complete nucleotide sequence of the insert, which was analyzed using BLAST. The insert sequence included an AluYa5 element belonging to the SINE1/7SL non-LTR retrotransposon class [48], a poly(T) tail of 56 bp, and the characteristic target site duplication (TSD, sequence: AACATTGTATCCTTC) (Fig. 1). The Alu insertion disrupting the coding exon 43 would, if translated, result in an insertion of 39 aberrant amino acids followed by a premature termination codon.

The Alu repeat mobile element insertion was validated by mutant allele-specific PCR and Sanger sequencing (Fig. 1). Family segregation studies in additional family members confirmed that the *Alu* insertion segregated

**Table 1** Command line to search for the AluYa5 insertion in coding exon 43 of the *EYS* gene using the grep search program

Allele	FASTQ files	BAM files
WT	zgrep -c "ATTAGAAGGATACAATGT TTATG CATAAACATTGTATCCTTCTAAT"*fastq.gz > output.txt	samtools view sample1.bam   grep -c -e ATTAGAAGGATACAATGTTTATG -e CATAAACATTGTATCCTTCTAAT > out- put.txt
MUT	zgrep c-"GAAGGATACAATGTTGGCCGGGC\  GCCCGGCCAACATTGTATCCTTC\ ACATTGTATCCTTCTTTTTTT AAAAAAGAAGGATACA ATGT"*fastq.gz > output.txt	samtools view sample1.bam   grep -c -e GAAGGATACAATGTTGGCCGGGC -e GCC CGGCCAACATTGTATCCTTC -e ACATTGTATCCTTCTTTTTTT -e AAAAAAGAAG GATACAATGT > output.txt

**Table 2** Clinical characteristics of the index patients of the families carrying the *EYS* exons 32–33 deletion

Pedigree subject	Clinical diagnosis	Onset age	First symptom	Symptoms at time of the genetic assessment	Fundus examination
Fam. A - II:8 (female)	RP	15 years	Night blindness and reduction of the visual field	ERG consistent with RP; decreased of visual acuity	Narrowed vessels; bone spicule pigmentation
Fam B - II:1 (male)	RP	17 years	Night blindness and reduction of the visual field	Concentric reduction of the visual field (10° central);	Bone spicule pigmentation; Pallor of the optic disc
Fam C - III: 4 (female)	RP	35 years	Concentric reduction of the visual field	Night blindness; Concentric reduction of the visual field (10° central); Decreased visual acuity; Cataracts	Narrowed vessels; Foveal preservation; ERG consistent with RP
Fam D - II:1 (female)	Unclassified IRD	17 years	Night blindness	Photophobia. Decreased visual acuity	No alterations

Abbreviations: ERG Electroretinography, Fam Family, IRD Inherited retinal dystrophy, RP Retinitis pigmentosa

with the disease in the family in combination with the deletion of exons 32–33 and was transmitted to the third generation (Fig. 1).

In previous targeted SRS studies of this family, the Alu insertion could not be identified, even after applying the MELT software. However, subsequent visual IGV inspection of this data revealed a number of reads with soft-clipped bases in the breakpoint region (76 reads out of 449 total reads; ~17%) (Fig. 1). The analysis of the bases was coincident with the AluYa5 sequence. In addition, a drop in coverage at the insertion point was observed, indicating that the standard BWA-based alignment of Illumina reads may fail to map chimeric reads (Alu-*EYS*) (Fig. 1).

#### Screening for the *EYS*-Alu insertion in additional patients and controls using SRS data

Since the exon 43 of *EYS* was already included in our diagnostic panel, a command line based on grep was first validated using the SRS data of the index patient of family A (Table 1). In order to determine the recurrence and the prevalence of this Alu insertion in our cohort, the optimized command was applied for the screening of FASTQ files from 327 individuals, including 178 IRD patients and 149 unaffected individuals. However, none harbored the Alu insertion, supporting its low frequency, and thus, its pathogenicity. Taken together, these data led us to consider the AluYa5 insertion in exon 43 of *EYS* (chr6(GRC h37):g.64430524\_64430525ins352, NM\_001142800.2:c.9402\_9403ins[[JQ403527.1:g.57\_367];9388\_9402]) as the most likely disease-causing second hit in the family A.

#### Characterization of structural variants by LRS

In addition, nanopore sequencing using adaptive sampling in the index patient of family A enabled the

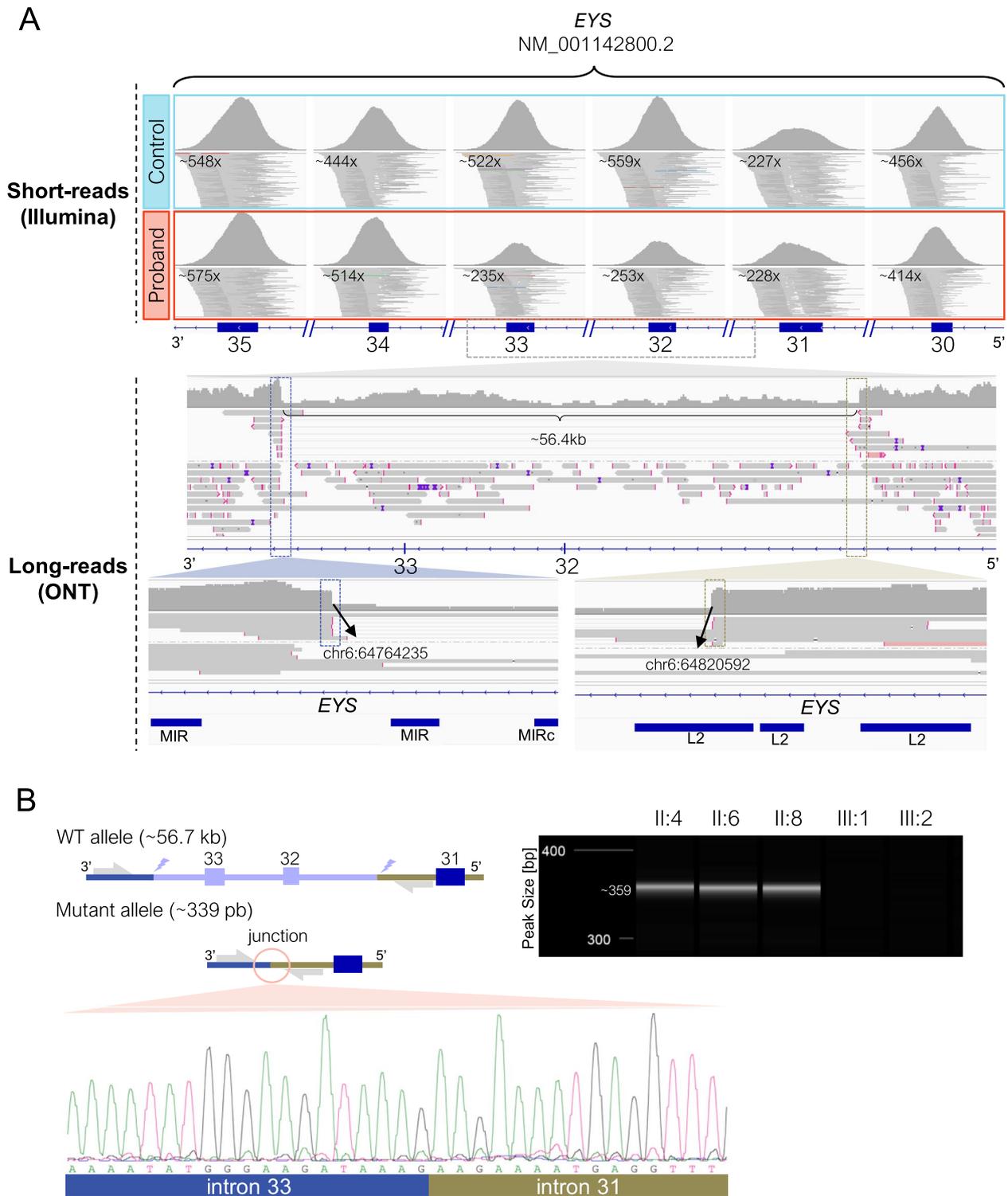
characterization of the previously identified *EYS* CNV at nucleotide-level resolution with a 10× coverage (Fig. 2), defining a ~56.4kb deletion that included exons 32–33 of *EYS* (chr6(GRC h37):g.64764235\_64820592 del, NM\_001142800.2:c.6425-28697\_6725+11996del). Visual inspection in IGV software using the RepeatMasker database, allowed us to locate the 5' breakpoint of the CNV in a long interspersed element (LINE) belonging to the L2 family of intron 31, whereas the 3' breakpoint was flanked (~200bp upstream and 600bp downstream) by a short interspersed element (SINE) belonging to the mammalian interspersed repeats (MIR) family (Fig. 2). To verify the breakpoint junctions, mutation-specific primers were designed using the information obtained by LRS in the flanking regions of the CNV, allowing amplification only in the three affected siblings carrying the deletion (Fig. 2). Sanger sequencing confirmed the deletion breakpoints previously detected by nanopore sequencing.

Since a similar deletion involving exons 32–33 of *EYS* had previously been detected in three additional IRD families from our cohort (Table 2), we performed Sanger sequencing and confirmed that they share the same breakpoints (Fig. 3). Then, microsatellite analysis revealed that the four families shared a region of ~1,9Mb, which extended up to ~12,5 Mb in families A, B and D (Fig. 3).

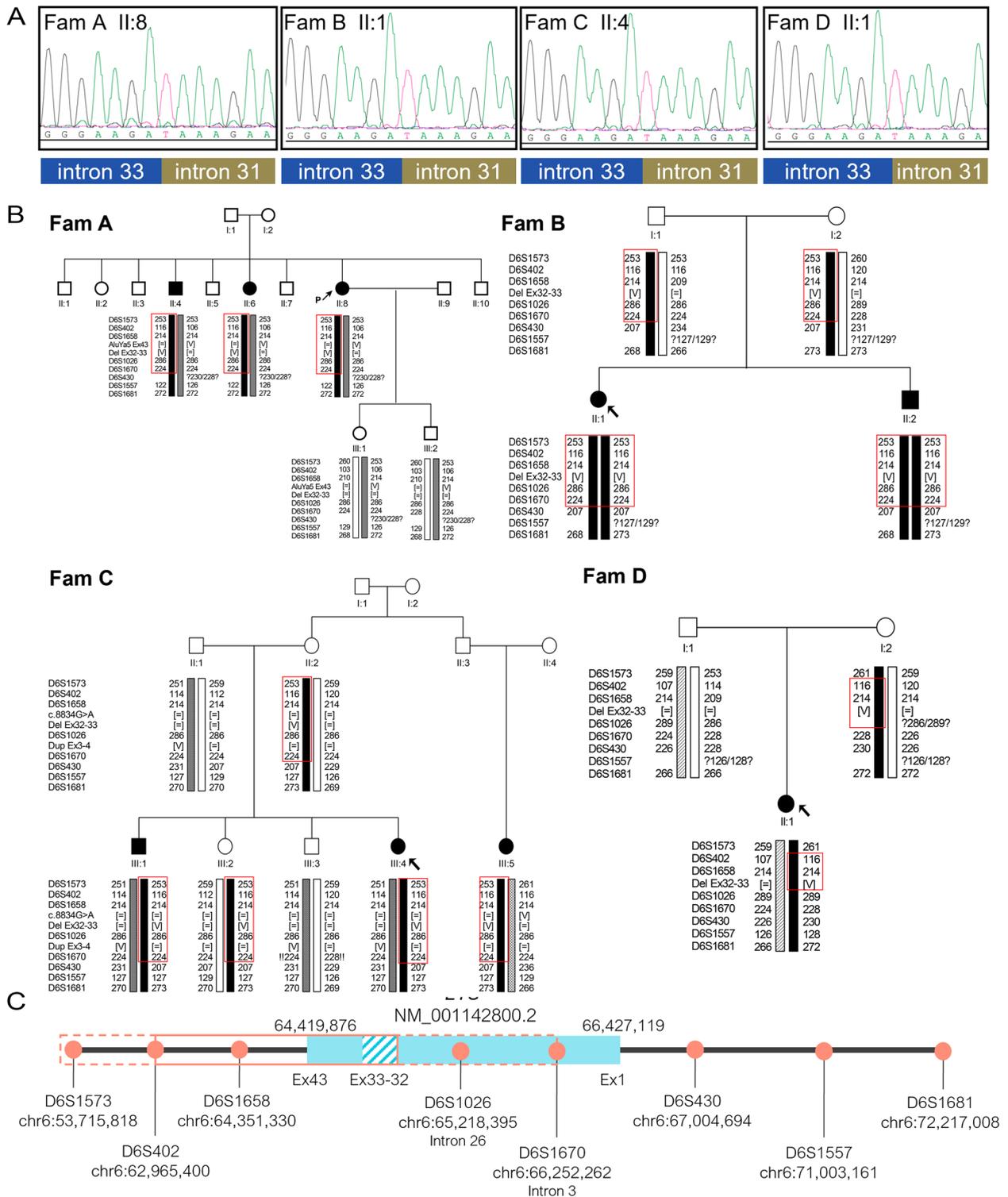
#### Discussion

Sequencing technologies have evolved rapidly since the discovery of Sanger sequencing over 50 years ago. However, despite these advances, the diagnostic yield for IRD remains in the range of 52 to 74%, depending on the phenotype [10, 49], indicating that a substantial proportion of causative variants remain unidentified or misinterpreted. Therefore, a closer look at the detection





**Fig. 2** Characterization of the deletion of exons 32–33 in the *EYS* gene. **A** Integrative Genomics Viewer (IGV) screenshots showing the deletion of exons 32–33 in the *EYS* gene detected by short-read sequencing and its characterization by long-read sequencing. **B** Validation by PCR and Sanger sequencing of the deletion junction in all individuals of family A



**Fig. 3** Linkage analysis of the four IRD families carrying the deletion of exons 32–33 of *EYS*. **A** Sanger sequencing of the deletion junction in the four index patients. **B** STR markers analysis in the 19 individuals from the four families. The allele carrying the deletion is shown in black. The region linked in each family is represented by a red square. **C** Representation of the selected STR markers, their chromosomal positions (GRCh37) and the putatively linked interval

of complex or non-coding variants that currently elude diagnostic pipelines would be of great benefit in increasing the diagnostic yield in IRD patients.

Here, we used LRS to identify the insertion of an AluYa5 element in the *EYS* coding exon 43 as one of the disease-causing variants in one arRP family, which was missed by a previous SRS approach. Alu elements are ~300 bp sequences belonging to a class of mobile elements or retrotransposons called SINEs that comprise 11% of the human genome, with nearly a million copies located primarily in introns and intergenic regions [50]. Among the Alu subfamilies, AluYa5 and AluYb8 are currently the most active subfamilies in the human lineage [51]. Although retrotransposons have contributed in many ways to genetic and functional diversity during evolution, their insertion can also be deleterious, disrupting coding exons or key regulatory elements, and serving as substrates for non-allelic recombination leading to CNVs [52]. In fact, the role of MEIs in the etiopathogenesis of a significant number of Mendelian diseases, including IRD, have already been described [19, 46, 53]. Examples of this are the recurrent retrotransposon insertions in *MAK* [19], *RPI* [54] or *BBS1* [55], which were first serendipitously discovered in linked families from populations with strong founder effects. Interestingly, the poly(T) tail of the identified AluYa5 in *EYS* is longer than 50 bp, indicating that the insert is quite young. In fact, the A-tails of very recent Alu insertions have been described to be between 40 and 97 bp in length [56]. Long A-tails tend to shorten relatively quickly towards 30 bases in terms of generations [56], which could be consistent with the variant being private to the family in study. The identified variant is inserted within the coding sequence of the *EYS* gene [6, 32], which is often the target of SVs, mainly CNVs [30, 31]. Examination of the recently updated gnomAD SVs v4 database revealed the presence of two Alu insertions with different breakpoints disrupting the same coding exon in two heterozygous carriers (INS\_CHR6\_9AB69B96 and INS\_CHR6\_DD0F655F). The identification of three independent Alu insertions within the same exon may indicate a site of high susceptibility to these events. This would make the *EYS* gene a good candidate to explore for the identification of Alu insertions as disease-causing variants.

In addition, thanks to the capability of LRS to cover entire SVs, we characterized at the nucleotide level a recurrent CNV-deletion comprising exons 32–33 of *EYS*. This variant has been reported in the literature in Portuguese, Spanish and French arRP families [29, 30, 33, 57], as well as in three heterozygous carriers (MAF=2.38e-05) from the recently updated gnomAD SVs v4 database from diverse genetic ancestry groups (Middle Eastern, Admixed American and African) (Variant ID:

DEL\_CHR6\_4F5408B3). Moreover, we identified the same deletion in three unrelated Spanish families from our IRD cohort. In this study, the linkage analysis results were consistent with a founder effect variant that may have originated in the Iberian Peninsula and spread to other regions, although studies in other populations are needed. The importance of identifying prevalent founder variants is increasing, as they are potential candidates for variant-specific therapies that may benefit a larger number of patients. These therapies may include antisense oligonucleotides (AONs), which allow targeting of large genes that, as *EYS*, cannot be treated with conventional gene supplementation therapies [58]. Regarding large deletions, AONs can modulate pre-mRNA splicing, and restore the disrupted reading frame. This is the case of Eteplirsén, the drug approved for Duchenne muscular dystrophy [59, 60].

Our results are in line with previous studies [61, 62] that suggest that the impact of complex SVs may be underrepresented in a variety of Mendelian disorders, reinforcing the need of applying systematic detection methods of MEIs, especially in partially or completely unsolved cases. While some SVs, such as deletions or duplications of one or more coding exons, are easy to identify computationally using targeted SRS data, other SVs, such as MEIs, are much more technically challenging due to their genome-wide distribution and related alignment difficulties in SRS data on repetitive regions, requiring the use of specific calling algorithms preferably applicable to WGS [63]. In fact, in our hands, MELT failed to identify the insertion of the Alu element using previous targeted SRS data. However, most clinical genetics centers still prefer to use targeted SRS in their diagnostic routine due to its robustness and cost-effectiveness, which means that the full landscape of SVs may not be fully uncovered. Therefore, there is an urgent need to develop feasible strategies to detect these SVs in the clinical setting without investing in additional costly WGS protocols [61]. In this work, we proposed a re-analysis strategy to uncover specific *EYS*-AluYa5 insertion on available SRS data (FASTQ and/or BAM files) using the Linux grep command as previously described [46, 47].

As we have shown, multiple types of SVs, including a MEI and a CNV, can be detected simultaneously by enriching a particular genomic region by LRS using the ONT adaptive sampling method. This method is powerful and can be performed in a cost-efficient manner using a Nanopore MinION flowcell, which ensures sufficient depth of coverage while reducing problems associated with the introduction of PCR artifacts and PCR length restrictions [64]. It also allows to gather additional clinically relevant information such as the precise SVs breakpoints, the full sequence of the inserted Alu, and

the phasing of the two compound heterozygous variants, which together may result in the resolution of heterogeneous genetic traits.

## Conclusions

In conclusion, our research has explored the potential of LRS to increase the diagnostic yield of an unsolved arRP family. As a result, adaptive sampling of the *EYS* gene revealed the precise location of a previously known recurrent CNV, as well as a novel Alu insertion both of which segregated with the disease in the family. In addition, haplotype analysis of unrelated families harboring the recurrent CNV was consistent with a founder variant, which may open the door to genomic medicine approaches in these patients. This is the first report of a pathogenic Alu insertion in the *EYS* gene, expanding its genotypic spectrum and strengthening the role of MEIs in the etiopathogenesis of IRDs.

## Abbreviations

AONs	Antisense oligonucleotides
arRP	Autosomal recessive retinitis pigmentosa
CNV	Copy number variations
EVS	Exome Variant Server
gDNA	Genomic DNA
GnomAD	Genome Aggregation Database
IGV	Integrative Genomics Viewer
IRD	Inherited retinal dystrophies
LINE	Long interspersed element
LRS	Long-read sequencing
MAF	Minor allele frequency
MEIs	Mobile element insertions
MELT	Mobile Element Locator Tool
MIR	Mammalian-wide interspersed repeats
MLPA	Multiplex Dependent Probe Amplification
ONT	Oxford Nanopore Technologies
RP	Retinitis pigmentosa
SINE	Short interspersed element
SNVs	Single nucleotide variants
SRS	Short-read sequencing
SVs	Structural variants
TSD	Target site duplication
VAF	Variant allele frequency
WGS	Whole-genome sequencing

## Acknowledgements

The authors thank the families who participated in this study, and the Andalusian Association of Retinitis Pigmentosa.

## Authors' contributions

E.F.S. and M.G.P. are co-first authors. G.A. and S.B. conceived and designed the study. E.F.S., and B.M.B. conducted the experiments. E.F.S., M.G.P., C.M.V., M.M.-S., B.M.B., J.C., analyzed and interpreted the generated data. E.F.S. and M.G.P. wrote the manuscript with the collaboration of all co-authors. G.A., S.B., C.M.V., M.M.S., M.M., B.M.B., and J.C., revised the paper critically for important intellectual content. All authors approved the final version to be published.

## Funding

This work was supported by the Instituto de Salud Carlos III (ISCIII), Spanish Ministry of Science and Innovation, Spain and co-funded by ERDF ("A way to make Europe") [PI21-00244]; The strategic plan for the Precision Medicine Infrastructure associated with Science and Technology—IMPACT [IMP-0009], Regional Ministry of Health and Families of the Autonomous Government of Andalusia [PEER-0501–2019] and the Foundation Isabel Gemio/Foundation Cajasol [FGEMIO-2019–01]. EFS is supported by fellowship F119/00091 from

ISCIII (ESF, "Investing in your future") and a formative mobility action from CIBERER (ERF02MOV/2023). MMS [RH-0049–2021] is supported by a fellowship funded by the Regional Ministry of Health and Families of the Autonomous Government of Andalusia.

## Availability of data and materials

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding authors. The novel MEI was submitted to ClinVar database under the accession ID: SCV004697438.

## Declarations

### Ethics approval and consent to participate

This study was conducted according to the ethical principles for medical research involving human subjects according to the Declaration of Helsinki (Edinburgh, 2000). Prior to the study, informed consents were collected from all participants or their legal guardians for clinical genetic testing and research studies, which was approved by the Ethical Committees of the University Hospital Virgen del Rocío (Seville) and the University Hospital Virgen Macarena (Seville).

### Consent for publication

Consent for publication was obtained from all study subjects.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Department of Maternofetal Medicine, Genetics and Reproduction, Institute of Biomedicine of Seville (IBIS), University Hospital Virgen del Rocío/CSIC, University of Seville, Seville, Spain. <sup>2</sup>Center for Biomedical Network Research On Rare Diseases (CIBERER), Seville, Spain. <sup>3</sup>Servicio de Hematología y Oncología Médica, Hospital Universitario Morales Meseguer, Centro Regional de Hemodonación, Universidad de Murcia, IMIB-Pascual Parrilla, CIBERER-ISCIII, Murcia, Spain.

Received: 21 February 2024 Accepted: 10 April 2024

Published online: 04 May 2024

## References

1. Cremers FPM, Boon CJF, Bujakowska K, Zeitz C. Special Issue Introduction: Inherited Retinal Disease: Novel Candidate Genes, Genotype-Phenotype Correlations, and Inheritance Models. *Genes (Basel)*. 2018;9(4):215.
2. Hamblion EL, Moore AT, Rahi JS. British Childhood Onset Hereditary Retinal Disorders N: Incidence and patterns of detection and management of childhood-onset hereditary retinal disorders in the UK. *Br J Ophthalmol*. 2012;96:360–5.
3. Verbakel SK, van Huet RAC, Boon CJF, den Hollander AI, Collin RWJ, Klaver CCW, Hoyng CB, Roepman R, Klevering BJ. Non-syndromic retinitis pigmentosa. *Prog Retin Eye Res*. 2018;66:157–86.
4. Fernandez-San Jose P, Blanco-Kelly F, Corton M, Trujillo-Tiebas MJ, Gimenez A, Avila-Fernandez A, Garcia-Sandoval B, Lopez-Molina MI, Hernan I, Carballo M, et al. Prevalence of Rhodopsin mutations in autosomal dominant Retinitis Pigmentosa in Spain: clinical and analytical review in 200 families. *Acta Ophthalmol*. 2015;93:e38–44.
5. Falsini B, Placidi G, De Siena E, Savastano MC, Minnella AM, Maceroni M, Midena G, Ziccardi L, Parisi V, Bertelli M, et al. USH2A-Related Retinitis Pigmentosa: Staging of Disease Severity and Morpho-Functional Studies. *Diagnostics (Basel)*. 2021;11(2):213.
6. Abd El-Aziz MM, Barragan I, O'Driscoll CA, Goodstadt L, Prigmore E, Borrego S, Mena M, Pieras JI, El-Ashry MF, Safieh LA, et al. *EYS*, encoding an ortholog of *Drosophila* spacemaker, is mutated in autosomal recessive retinitis pigmentosa. *Nat Genet*. 2008;40:1285–7.
7. Barragan I, Borrego S, Pieras JI, Gonzalez-del Pozo M, Santoyo J, Ayuso C, Baiget M, Millan JM, Mena M, Abd El-Aziz MM, et al. Mutation spectrum

- of EYS in Spanish patients with autosomal recessive retinitis pigmentosa. *Hum Mutat.* 2010;31:E1772–1800.
8. Arai Y, Maeda A, Hirami Y, Ishigami C, Kosugi S, Mandai M, Kurimoto Y, Takahashi M. Retinitis Pigmentosa with EYS Mutations Is the Most Prevalent Inherited Retinal Dystrophy in Japanese Populations. *J Ophthalmol.* 2015;2015:819760.
  9. Chivers M, Li N, Pan F, Wiewfer H, Slowik R, Leartsakulpanitch J. The Burden of X-Linked Retinitis Pigmentosa on Patients and Society: A Narrative Literature Review. *Clinicoecon Outcomes Res.* 2021;13:565–72.
  10. Britten-Jones AC, Gocuk SA, Goh KL, Huq A, Edwards TL, Ayton LN. The Diagnostic Yield of Next Generation Sequencing in Inherited Retinal Diseases: A Systematic Review and Meta-analysis. *Am J Ophthalmol.* 2023;249:57–73.
  11. Martin-Sanchez M, Bravo-Gil N, Gonzalez-Del Pozo M, Mendez-Vidal C, Fernandez-Suarez E, Rodriguez-de la Rúa E, Borrego S, Antinolo G: A Multi-Strategy Sequencing Workflow in Inherited Retinal Dystrophies: Routine Diagnosis, Addressing Unsolved Cases and Candidate Genes Identification. *Int J Mol Sci.* 2020;21(24):9355.
  12. Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S, et al. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum Genet.* 2017;100:75–90.
  13. Weissschuh N, Mazzola P, Zuleger T, Schaeferhoff K, Kuhlewein L, Kortum F, Witt D, Liebmann A, Falb R, Pohl L, et al. Diagnostic genome sequencing improves diagnostic yield: a prospective single-centre study in 1000 patients with inherited eye diseases. *J Med Genet.* 2023;61(2):186–95.
  14. Biswas P, Villanueva AL, Soto-Hermida A, Duncan JL, Matsui H, Borooah S, Kurmanov B, Richard G, Khan SY, Branham K, et al. Deciphering the genetic architecture and ethnographic distribution of IRD in three ethnic populations by whole genome sequence analysis. *PLoS Genet.* 2021;17:e1009848.
  15. Gonzalez-Del Pozo M, Fernandez-Suarez E, Bravo-Gil N, Mendez-Vidal C, Martin-Sanchez M, Rodriguez-de la Rúa E, Ramos-Jimenez M, Morillo-Sanchez MJ, Borrego S, Antinolo G: A comprehensive WGS-based pipeline for the identification of new candidate genes in inherited retinal dystrophies. *NPJ Genom Med.* 2022;7:17.
  16. Chrystal PW, Lambacher NJ, Doucette LP, Bellingham J, Schiff ER, Noel NCL, Li C, Tsiropoulou S, Casey GA, Zhai Y, et al. The inner junction protein CFP20 functions in motile and non-motile cilia and is critical for vision. *Nat Commun.* 2022;13:6595.
  17. Begum G, Albanna A, Bankapur A, Nassir N, Tambi R, Berdiev BK, Akter H, Karuvantevida N, Kellam B, Alhashmi D, et al. Long-Read Sequencing Improves the Detection of Structural Variations Impacting Complex Non-Coding Elements of the Genome. *Int J Mol Sci.* 2021;22(4):2060.
  18. de Bruijn SE, Rodenburg K, Corominas J, Ben-Yosef T, Reurink J, Kremer H, Whelan L, Plomp AS, Berger W, Farrar GJ, et al. Optical genome mapping and revisiting short-read genome sequencing data reveal previously overlooked structural variants disrupting retinal disease-associated genes. *Genet Med.* 2023;25:100345.
  19. Tucker BA, Scheetz TE, Mullins RF, DeLuca AP, Hoffmann JM, Johnston RM, Jacobson SG, Sheffield VC, Stone EM. Exome sequencing and analysis of induced pluripotent stem cells identify the cilia-related gene male germ cell-associated kinase (MAK) as a cause of retinitis pigmentosa. *Proc Natl Acad Sci U S A.* 2011;108:E569–576.
  20. Nikopoulos K, Cisarova K, Quinodoz M, Koskiniemi-Kuendig H, Miyake N, Farinelli P, Rehman AU, Khan MI, Prunotto A, Akiyama M, et al. A frequent variant in the Japanese population determines quasi-Mendelian inheritance of rare retinal ciliopathy. *Nat Commun.* 2019;10:2884.
  21. Beyter D, Ingimundardottir H, Oddsson A, Eggertsson HP, Bjornsson E, Jonsson H, Atlason BA, Kristmundsdottir S, Mehlinger S, Hardarson MT, et al. Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. *Nat Genet.* 2021;53:779–86.
  22. Mastrorosa FK, Miller DE, Eichler EE. Applications of long-read sequencing to Mendelian genetics. *Genome Med.* 2023;15:42.
  23. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, Khara AV, Lowther C, Gauthier LD, Wang H, et al. A structural variation reference for medical and population genetics. *Nature.* 2020;581:444–51.
  24. Wojcik MH, Reuter CM, Marwaha S, Mahmoud M, Duyzend MH, Barseghyan H, Yuan B, Boone PM, Groopman EE, Delot EC, et al. Beyond the exome: what's next in diagnostic testing for Mendelian conditions. *ArXiv* 2023.
  25. Loose M, Malla S, Stout M. Real-time selective sequencing using nanopore technology. *Nat Methods.* 2016;13:751–4.
  26. Nakamichi K, Van Gelder RN, Chao JR, Mustafi D. Targeted adaptive long-read sequencing for discovery of complex phased variants in inherited retinal disease patients. *Sci Rep.* 2023;13:8535.
  27. de la Morena-Barrio B, Palomo A, Padilla J, Martin-Fernandez L, Rojo-Carrillo JJ, Cifuentes R, Bravo-Perez C, Garrido-Rodriguez P, Minano A, Rubio AM, et al. Impact of genetic structural variants in factor XI deficiency: identification, accurate characterization, and inferred mechanism by long-read sequencing. *J Thromb Haemost.* 2023;21:1779–88.
  28. Okubo M, Noguchi S, Awaya T, Hosokawa M, Tsukui N, Ogawa M, Hayashi S, Komaki H, Mori-Yoshimura M, Oya Y, et al. RNA-seq analysis, targeted long-read sequencing and in silico prediction to unravel pathogenic intronic events and complicated splicing abnormalities in dystrophinopathy. *Hum Genet.* 2023;142:59–71.
  29. Soares RM, Carvalho AL, Simao S, Soares CA, Raimundo M, Alves CH, Ambrosio AF, Murta J, Saraiva J, Silva R, Marques JP. Eyes Shut Homolog-Associated Retinal Degeneration: Natural History, Genetic Landscape, and Phenotypic Spectrum. *Ophthalmol Retina.* 2023;7:628–38.
  30. Pieras JJ, Barragan I, Borrego S, Audo I, Gonzalez-Del Pozo M, Bernal S, Baiget M, Zeitl C, Bhattacharya SS, Antinolo G. Copy-number variations in EYS: a significant event in the appearance of arRP. *Invest Ophthalmol Vis Sci.* 2011;52:5625–31.
  31. Liu X, Hu F, Zhang D, Li Z, He J, Zhang S, Wang Z, Zhao Y, Wu J, Liu C, et al. Whole genome sequencing enables new genetic diagnosis for inherited retinal diseases by identifying pathogenic variants. *npj Genomic Medicine.* 2024;9(1):6.
  32. Barragan I, Abd El-Aziz MM, Borrego S, El-Ashry MF, O'Driscoll C, Bhattacharya SS, Antinolo G. Linkage validation of RP25 Using the 10K gene-chip array and further refinement of the locus by new linked families. *Ann Hum Genet.* 2008;72:454–62.
  33. Audo I, Sahel JA, Mohand-Said S, Lancelot ME, Antonio A, Moskova-Doumanova V, Nandrot EF, Doumanov J, Barragan I, Antinolo G, et al. EYS is a major gene for rod-cone dystrophies in France. *Hum Mutat.* 2010;31:E1406–1435.
  34. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310:2191–4.
  35. Puppo Moreno AM, Bravo-Gil N, Mendez-Vidal C, Adsuar Gomez A, Ruiz Gomez FT, Jimenez De Juan C, Fernandez Garcia RM, Martin Bermudez R, Lopez Sanchez JM, Martin Sastre S, et al. Genetic profile in patients with complicated acute aortic syndrome: the GEN-AOR study. *Rev Esp Cardiol (Engl Ed).* 2023;76:434–43.
  36. Fernández-Suárez E, González-del Pozo M, García-Núñez A, Méndez-Vidal C, Martín-Sánchez M, Mejías-Carrasco JM, Ramos-Jiménez M, Morillo-Sánchez MJ, Rodríguez-de la Rúa E, Borrego S, Antinolo G: Expanding the phenotype of THRB: a range of macular dystrophies as the major clinical manifestations in patients with a dominant splicing variant. 2023;11:1197744.
  37. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, Mills RE, Genomes Project C, Devine SE. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res.* 2017;27:1916–29.
  38. Ueno Y, Arita M, Kumagai T, Asai K. Processing sequence annotation data using the Lua programming language. *Genome Inform.* 2003;14:154–63.
  39. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34:3094–100.
  40. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics.* 2021;37:4572–4.
  41. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods.* 2018;15:461–8.
  42. Zheng Z, Li S, Su J, Leung AW-S, Lam T-W, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nature Computational Science.* 2022;2:797–803.
  43. Geoffroy V, Guignard T, Kress A, Gaillard JB, Solli-Nowlan T, Schalk A, Gatinois V, Dollfus H, Scheidecker S, Muller J. AnnotSV and knotAnnotSV: a web server for human structural variations annotations, ranking and analysis. *Nucleic Acids Res.* 2021;49:W21–8.

44. Cingolani P, Platts A, Le Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w<sup>1118</sup>; iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92.
45. McMurray CL, Hardy KJ, Hawkey PM. Rapid, automated epidemiological typing of methicillin-resistant *Staphylococcus aureus*. *J Microbiol Methods*. 2010;80:109–11.
46. Bujakowska KM, White J, Place E, Consugar M, Comander J. Efficient In Silico Identification of a Common Insertion in the MAK Gene which Causes Retinitis Pigmentosa. *PLoS ONE*. 2015;10:e0142614.
47. Won D, Hwang JY, Shim Y, Byeon SH, Lee J, Lee CS, Kim M, Lim HT, Choi JR, Lee ST, Han J. In Silico identification of a common mobile element insertion in exon 4 of RP1. *Sci Rep*. 2021;11:13381.
48. Kojima KK. Human transposable elements in Repbase: genomic footprints from fish to humans. *Mob DNA*. 2018;9:2.
49. Perea-Romero I, Gordo G, Iancu IF, Del Pozo-Valero M, Almoguera B, Blanco-Kelly F, Carreno E, Jimenez-Rolando B, Lopez-Rodriguez R, Lorda-Sanchez I, et al. Genetic landscape of 6089 inherited retinal dystrophies affected cases in Spain and their therapeutic and extended epidemiological implications. *Sci Rep*. 2021;11:1526.
50. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet*. 2002;3:370–9.
51. Konkel MK, Walker JA, Hotard AB, Ranck MC, Fontenot CC, Storer J, Stewart C, Marth GT, Batzer MA. Sequence Analysis and Characterization of Active Human Alu Subfamilies Based on the 1000 Genomes Pilot Project. *Genome Biol Evol*. 2015;7:2608–22.
52. Payer LM, Burns KH. Transposable elements in human genetic disease. *Nat Rev Genet*. 2019;20:760–72.
53. Taskesen M, Collin GB, Evsikov AV, Guzel A, Ozgul RK, Marshall JD, Naggert JK. Novel Alu retrotransposon insertion leading to Alstrom syndrome. *Hum Genet*. 2012;131:407–13.
54. Nishiguchi KM, Fujita K, Ikeda Y, Kunikata H, Koyanagi Y, Akiyama M, Abe T, Wada Y, Sonoda KH, Nakazawa T. A founder Alu insertion in RP1 gene in Japanese patients with retinitis pigmentosa. *Jpn J Ophthalmol*. 2020;64:346–50.
55. Delvallée C, Nicaise S, Antin M, Leuvre AS, Nourisson E, Leitch CC, Kellaris G, Stoetzel C, Geoffroy V, Scheidecker S, et al. A BBS1 SVA F retrotransposon insertion is a frequent cause of Bardet-Biedl syndrome. *Clin Genet*. 2021;99:318–24.
56. Roy-Engel AM, Salem AH, Oyeniran OO, Deininger L, Hedges DJ, Kilroy GE, Batzer MA, Deininger PL. Active Alu element "A-tails": size does matter. *Genome Res*. 2002;12:1333–44.
57. Martin-Merida I, Avila-Fernandez A, Del Pozo-Valero M, Blanco-Kelly F, Zurita O, Perez-Carro R, Aguilera-Garcia D, Riveiro-Alvarez R, Arteché A, Trujillo-Tiebas MJ, et al. Genomic Landscape of Sporadic Retinitis Pigmentosa: Findings from 877 Spanish Cases. *Ophthalmology*. 2019;126:1181–8.
58. Kuijper EC, Bergsma AJ, Pijnappel W, Aartsma-Rus A. Opportunities and challenges for antisense oligonucleotide therapies. *J Inher Metab Dis*. 2021;44:72–87.
59. Niks EH, Aartsma-Rus A. Exon skipping: a first in class strategy for Duchenne muscular dystrophy. *Expert Opin Biol Ther*. 2017;17:225–36.
60. Aartsma-Rus A, Krieg AM. FDA Approves Eteplirsén for Duchenne Muscular Dystrophy: The Next Chapter in the Eteplirsén Saga. *Nucleic Acid Ther*. 2017;27:1–3.
61. Demidov G, Park J, Armeanu-Ebinger S, Roggia C, Faust U, Cordts I, Blandfort M, Haack TB, Schroeder C, Ossowski S. Detection of mobile elements insertions for routine clinical diagnostics in targeted sequencing data. *Mol Genet Genomic Med*. 2021;9:e1807.
62. Gardner EJ, Prigmore E, Gallone G, Danecek P, Samocha KE, Handsaker J, Gerety SS, Ironfield H, Short PJ, Sifrim A, et al. Contribution of retrotransposition to developmental disorders. *Nat Commun*. 2019;10:4630.
63. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, Mills RE, Devine SE. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res*. 2017;27:1916–29.
64. Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, Leggett RM. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol*. 2022;23:11.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.