



HAL
open science

Validation of Nanopore long-read sequencing to resolve RPGR ORF15 genotypes in individuals with X-linked retinitis pigmentosa

Christel Vaché, Valérie Faugère, David Baux, Luke Mansard, Charles van Goethem, Claire-Marie Dhaenens, Olivier Grunewald, Isabelle Audo, Christina Zeitz, Isabelle Meunier, et al.

► To cite this version:

Christel Vaché, Valérie Faugère, David Baux, Luke Mansard, Charles van Goethem, et al.. Validation of Nanopore long-read sequencing to resolve RPGR ORF15 genotypes in individuals with X-linked retinitis pigmentosa. *European Journal of Human Genetics*, In press, 10.1038/s41431-024-01649-0 . hal-04638052

HAL Id: hal-04638052

<https://hal.science/hal-04638052>

Submitted on 8 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Validation of Nanopore long-read sequencing to resolve RPGR ORF15 genotypes in individuals with X-linked retinitis pigmentosa

Christel Vaché^{1 2}, Valérie Faugère³, David Baux^{3 4 5}, Luke Mansard^{3 4}, Charles Van Goethem^{3 5}, Claire-Marie Dhaenens⁶, Olivier Grunewald⁶, Isabelle Audo^{7 8}, Christina Zeitz⁷, Isabelle Meunier^{4 9}, Béatrice Bocquet^{4 9}, Mireille Cossée^{3 10}, Anne Bergougnoux^{3 10}, Vasiliki Kalatzis^{4 9}, Anne-Françoise Roux^{3 4 11}

¹ Molecular Genetics Laboratory, Univ Montpellier, CHU Montpellier, Montpellier, France. christel.vache@inserm.fr.

² Institute for Neurosciences of Montpellier (INM), Univ Montpellier, Inserm, Montpellier, France. christel.vache@inserm.fr.

³ Molecular Genetics Laboratory, Univ Montpellier, CHU Montpellier, Montpellier, France.

⁴ Institute for Neurosciences of Montpellier (INM), Univ Montpellier, Inserm, Montpellier, France.

⁵ Montpellier BioInformatique pour le Diagnostic Clinique (MOBIDIC), CHU Montpellier, Montpellier, France.

⁶ Univ. Lille, Inserm, CHU Lille, U1172-LilNCog-Lille Neuroscience and Cognition, Lille, France.

⁷ Sorbonne Université, Inserm, CNRS, Institut de la Vision, Paris, France.

⁸ CHNO des Quinze-Vingts, Centre de Référence Maladies Rares REFERET, Inserm-DGOS CIC1423, Paris, France.

⁹ National Reference Center for Inherited Sensory Diseases, Univ Montpellier, CHU, Montpellier, France.

¹⁰ PhyMedExp, Univ Montpellier, Inserm, CNRS, Montpellier, France.

¹¹ GCS AURAGEN, Lyon, France.

Abstract

X-linked retinitis pigmentosa (XLRP) is characterized by progressive vision loss leading to legal blindness in males and a broad severity spectrum in carrier females. Pathogenic alterations of the retinitis pigmentosa GTPase regulator gene (RPGR) are responsible for over 70% of XLRP cases. In the retina, the RPGR^{ORF15} transcript includes a terminal exon, called ORF15, that is altered in the large majority of RPGR-XLRP cases. Unfortunately, due to its highly repetitive sequence, ORF15 represents a considerable challenge in terms of sequencing for molecular diagnostic laboratories. However, in a recent preliminary work Yahya et al. reported a long-read sequencing approach seeming promising. Here, the aim of the study was to validate and integrate this new sequencing strategy in a routine screening workflow. For that purpose, we performed a masked test on 52 genomic DNA samples from male and female individuals carrying 32 different pathogenic ORF15 variations including 20 located in the highly repetitive region of the exon. For the latter, we have obtained a detection rate of 80-85% in males and 60-80% in females after bioinformatic analyses. These numbers raised to 100% for both status after adding a complementary visual inspection of ORF15 long-reads. In accordance with these results, and considering the frequency of ORF15 pathogenic variations in XLRP, we suggest that a long-read screening of ORF15 should be systematically considered before any other sequencing approach in subjects with a diagnosis compatible with XLRP.

INTRODUCTION

Retinitis pigmentosa (RP), also known as rod-cone dystrophy, is a rare inherited retinal disorder with a high clinical and genetic heterogeneity (OMIM #268000). Among the different inheritance patterns of RP, the X-linked mode is responsible for one of the most severe phenotypes (X-linked RP, XLRP) affecting predominantly males. A recent study [1] estimated the prevalence of XLRP to between 4.0–5.2 per 100,000 males, and, in particular, the prevalence of XLRP due to a pathogenic variant in the retinitis pigmentosa GTPase regulator gene (RPGR) to between 3.4–4.4 per 100,000 males, highlighting the major contribution of this gene to the disease. The majority of males suffering from RPGR-XLRP present night blindness in the first decades, followed by rapid progression of peripheral and central vision loss resulting in legal blindness in the third or fourth decade [2]. Compared to males, females carrying heterozygous RPGR mutations usually have a more favorable visual prognosis, even though a broad spectrum of severity has been reported influenced by the random inactivation of X chromosomes ([3, 4]).

The RPGR gene encodes several isoforms of the RP GTPase regulator protein as a result of alternative splicing (<https://www.gtexportal.org/home/transcriptPage>). Among them, the RPGRORF15 isoform (1,152 amino acids, NP_001030025.1), which is predominantly expressed in retina and brain ([5, 6]) plays a key role in the function of the photoreceptor cells [7]. The RPGRORF15 transcript (NM_001034853.2) consists of the first fourteen exons of RPGR and a large terminal exon (exon 15 plus part of intron 15) called open reading frame 15 (ORF15), which encompasses a highly repetitive, purine-rich domain [5]. Because of this low complexity, ORF15 is prone to mutational events and is considered as a hot spot responsible for approximately 80% of RPGR-XLRP cases ([5]; Leiden Open Variation Database (LOVD) Global Variome Shared Instance (LOVD GVShared, <https://databases.lovd.nl/shared/genes/RPGR>)), making its screening essential for an comprehensive molecular diagnosis. Moreover, this screening is also particularly important because ORF15 variants are usually associated with a more deleterious phenotype compared to variants located in other exons [8].

Currently, most molecular diagnostic laboratories use second generation sequencing technology to screen pathogenic variants in gene panels, exomes or genomes. Unfortunately, this sequencing method is well known for its inability to correctly cover low-complexity sequences like ORF15, leading to an incomplete RPGR screening and consequently to misdiagnosis for a number of affected individuals [9]. To overcome this technological limitation, and because conventional ORF15 Sanger sequencing is challenging, development of new approaches leading to accurate, rapid, and cost-effective screening is of great importance. In this context, a third-generation sequencing strategy, using the Nanopore technology (Oxford Nanopore Technologies), was recently proposed by Yahya et al. [10] to sequence a PCR-amplified fragment encompassing ORF15. Although the reported data seem promising, the limited number and nature of the different identified pathogenic variants (2 substitutions and four 2-bp deletions), as well as the low number of heterozygous carriers, known to be more challenging to sequence [11], tested (3 females), rendered an extensive study mandatory to definitively validate this sequencing strategy.

Thus, in order to implement this promising ORF15 sequencing approach in a routine molecular laboratory, we performed a masked test from a set of controls and affected individuals carrying in the hemizygous and/or heterozygous state 32 different ORF15

pathogenic variations including 20 located in the highly repetitive region of the exon. The workflow used in this study allowed us to quickly identify each alteration in each PCR-amplified DNA sample, validating the Nanopore technology for ORF15 screening.

MATERIALS AND METHODS

Samples

A total of 52 genomic DNA samples from affected individuals (29 males and 23 females) carrying hemizygous or heterozygous pathogenic RPGRORF15 variants and 4 controls (two males and two females), previously analyzed in three French laboratories, were included in the study. In males 19 out of 29 pathogenic variants were located in the highly repetitive region of ORF15 (g.38,285,847 - 38,286,647 (hg38)) and in females they were 15 out of 23. Samples were used in a masked test to validate the long-read sequencing approach, i.e. persons who analyzed the MinION sequencing data did not know the number of positive cases and did not have the list of variants previously identified; results of the long-read analyses were compared a posteriori with data from the three laboratories. Informed consent for genetic analyses was obtained from all participants or their legal representatives. This study was conducted according to the guidelines of the Declaration of Helsinki and in accordance with the French law on bioethics: revised 7 July 2011, number 2011-814. This study was approved by the Montpellier University Hospital (CHU Montpellier) as part of the molecular diagnostic activity. The authorization number given by the Agence Régionale de la Santé (ARS) is LR/2013-N°190. For affected individuals seen at the CHU Lille, U1172-LilNCog-Lille Neuroscience and Cognition, the Lille Database “BASE-OPH” CNIL authorization number is DR- 2023-061. For affected individuals seen at the CHNO des Quinze-Vingts, Centre de Référence Maladies Rares REFERET, the studies were approved by a national ethics committee (CPP Ile de France V, Project number 06693, N°EUDRACT 2006-A00347-44, 11 December 2006).

ORF15 amplification

Each DNA sample was used as template for PCR amplification of a 2062 bp sequence encompassing the entire ORF15 reading frame using the forward primer 5'-AGCAGCCTGAGGCAATAGAA-3' paired with the reverse primer 5'-CAAATTTACCAGTGCCTCCT-3'. PCR reactions (25 µl) contained 50 ng of DNA, 12.5 µl of PCR Master Mix (Promega, Charbonnières, France), 0.4 µM of forward and reverse primers and thermocycling conditions were 96 °C for 6 min, followed by 40 cycles of 96 °C for 30 s, 55 °C for 30 s, 72 °C for 2.5 min, and a final cycle of 72 °C for 8 min. PCR products were purified using Qiagen columns (Promega) and quantified using a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

Long-read sequencing with the MinION device

Libraries were prepared with the ligation sequencing kit SQK-LSK109 (Oxford Nanopore Technologies, Oxford, UK) according to the supplier's protocol, and contained 12 or 24 ORF15 amplicons barcoded with the Nanopore rapid barcoding kits EXP-NBD104 and EXP-NBD114. In this protocol, samples were pooled in equimolar ratios. Runs with 12 amplicons were carried out by injection of the library (approximately 50 fmol) into a MinION flow cell R9.4.1 for a sequencing time of 8 h. One run including 24 amplicons was performed with a first injection of 25 fmol of the library and a sequencing time of 2.5 h, followed by a flow cell

wash of 1 h (Flow cell wash kit EXP-WSH004) and a second injection of the residual library volume for an additional sequencing time of 2.5 h.

All experiments were performed with MinKNOW v22.08.9 to control the MinION device [12]. Basecalling was performed by Guppy v6.5.7 (High accuracy mode), using the graphics processing mode on a Nvidia® Quadro RTX4000 graphics card. The resulting fastq were concatenated and aligned with a home-made pipeline (<https://github.com/MobiDL/ONTVariantCalling>) based on minimap2 (v2.22, [13]). Variant calling was performed with Clair3 (v0.1-r12, <https://github.com/HKU-BAL/Clair3> [14]) with default settings. Resulting VCF files were imported and analyzed with the SEAL NGS analysis software (<https://github.com/mobidic/SEAL>), which annotates the VCF files using Variant Effect Predictor (v104.3, [15]) and displays variants in a user-friendly interface. This bioinformatic analysis was completed by a direct visual inspection of the sequenced reads along the entire target region using the Integrative Genomics Viewer (IGV) software (v2.7.2, [16]).

Variant validation

Sanger sequencing, using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Courtaboeuf, France) on an Applied Biosystems® 3500Dx Genetic Analyzer (Applied Biosystems), was used in order to validate all candidate variants detected by long-read sequencing. The 2062 bp ORF15 amplicon was used as the template and Sanger sequencing using internal primers (Supplementary Table 1) was restricted to the region surrounding the variant.

Nomenclature of the variants follows the Human Genome Variation Society [17] recommendations v20.05 (<http://varnomen.hgvs.org/>) with nucleotide +1 corresponding to the A of the ATG initiation codon in the RPGRORF15 reference sequence NM_001034853.2.

The genome reference consortium human build 38 patch release 14 (GRCh38.p14; hg38) was used to define genomic positions.

The RPGR pathogenic variants studied here are reported in the dedicated LOVD database or have been submitted to MobiDetails [18] (<https://tinyurl.com/4v7nydmy>), and classifications have been submitted to the Global Variome Shared LOVD instance (variants #chrX_018940 to #chrX_018949).

RESULTS

Data obtained with 12 barcoded amplicons per run

Metrics.

All 56 genomic DNA samples from subjects involved in our masked test study were correctly amplified and used as template for long-read sequencing. Bases called per run were between 119 and 274 Mb (See Supplementary Table 1 for more details), leading to a full coverage of the whole ORF15 amplicon in comparison to results obtained with second-generation sequencing technology (Supplementary Fig. 1a). For each run, the pore activity graph, provided in the final MinION report, highlighted a rapid and massive pore availability decrease within the 2 first hours of sequencing (Supplementary Fig. 1b). As a result, the

cumulative output graph showed a rapid and large increase of reads produced in the first 2 h of the experiment followed by a slight gain or a plateau over the 6 last hours (Supplementary Fig. 1b).

Detection of pathogenic variants in males.

In males, 26 out of 29 pathogenic variants were correctly reported by the bioinformatic tools (Table 1), confirmed by visual inspection of the reads using the Integrative Genomics Viewer (IGV) tool (Supplementary Fig. 2) and validated by Sanger sequencing. When annotated with other potential deleterious variations, a manual inspection of the patient's reads in comparison with controls always allowed us to immediately discriminate them from false positives (Supplementary Fig. 3 and Supplementary Table 2). The 3 false negatives consisted of two deletions, c.2522del and c.2792del, and a 29 bp insertion, c.2931_2932ins[AAAGG;2908_2931]. The presence of these sequence defects was quickly pinpointed by the IGV tool, and their precise type, location and size were obtained by Sanger sequencing (Fig. 1).

Data obtained for variants located in the g.38,285,847 - 38,286,647 (hg38) region were detailed in Supplementary Table 2.

In male controls, no pathogenic variant but one duplication was revealed by IGV screening. After Sanger sequencing, this alteration was clearly identified as a 21-bp benign polymorphic duplication (c.2820_2840dup; ClinVar variation ID: 1267917).

Detection of pathogenic variants in females.

Twenty-three females carrying a different pathogenic variant in the heterozygous state were used in the masked test. Twenty of these variants were also present in the male cohort, including 2 out the 3 Clair3 undetected variants i.e. c.2792del and c.2931_2932ins[AAAGG;2908_2931]. As expected, the bioinformatic analysis was also unable to report these 2 pathogenic variants when present in the heterozygous state. Again, IGV inspection followed by Sanger sequencing solved the cases. In addition, 5 other variants (c.2426_2427del, c.2506del, c.2964_2965del, c.2997_2998del and c.3286_3287delinsC) were not reported by the bioinformatic analysis (although they were correctly reported when in a hemizygous state) and were only detected by a visual inspection of the reads (Fig. 2). The 16 remaining pathogenic variants were accurately detected by our bioinformatic analysis and identified by IGV screening (Supplementary Fig. 2). All of them were confirmed by Sanger sequencing.

Data obtained for variants located in the g.38,285,847 -38,286,647 (hg38) region were detailed in Supplementary Table 2.

In one female control, we also found a suspicious alteration by visual inspection of the reads that we finally identified as the well-known benign polymorphic c.2919_2939dup variation after Sanger sequencing (see the LOVD-RPGR database).

Table 1. Results of the masked test from bioinformatic data obtained from 52 affected individuals carrying 32 different ORF15 pathogenic variants.

HGVS DNA	Genomic nomenclature		HGVS protein	Status	Bioinformatic detection		
	(hg38)				Run_12 8 h	Run_24 2.5 h	Run_24 2.5 x2 h
(NM_001034853.2)							
c.1861dup	g.38287139dup		p.(Glu621 Glyfs Ter9)	M	✓		
c.2153del	g.38286848del		p.(Gly718Alafs Ter51)	M	✓		
c.2218G>T	g.38286781C>A		p.(Glu740 Ter)	M	✓		
				F	✓		
c.2234_2237del	g.38286763_38286766del		p.(Arg745LysfsTer69)	M	✓		
				F	✓		
c.2236_2237del	g.38286765_38286766del		p.(Glu746 ArgfsTer23)	M	✓		
				F	✓		
c.2257_2260del	g.38286743_38286746del		p.(Gly753Lysfs Ter61)	M	✓		
				F	✓		
c.2270_2271del	g.38286730_38286731del		p.(Glu757 Glyfs Ter12)	M	✓		
				F	✓		
c.2284G>T	g.38286715C>A		p.(Glu762 Ter)	F	✓		
c.2287G>T	g.38286712C>A		p.(Glu763 Ter)	F	✓		
c.2405_2406del	g.38286595_38286596del		p.(Glu802 Glyfs Ter32)	M	✓		
				F	✓		
c.2416del	g.38286586del		p.(Glu806 ArgfsTer9)	M	✓	✓	✓
c.2426_2427del	g.38286574_38286575del		p.(Glu809 Glyfs Ter25)	M	✓	✓	✓
				F	x	x	✓
c.2442_2445del	g.38286554_38286557del		p.(Gly817Lysfs Ter2)	M	✓	✓	✓
				F	✓	✓	✓
c.2455_2468del	g.38286540_38286553del		p.(Val819A Argfs Ter11)	M	✓	✓	✓
				F	✓	✓	✓
c.2468_2472del	g.38286530_38286534del		p.(Lys823 ArgfsTer10)	M	✓	✓	✓
c.2506del	g.38286497del		p.(Glu836 LysfsTer253)	M	✓	✓	✓
				F	x	x	x
c.2522del	g.38286477del		p.(Glu841 Glyfs Ter248)	M	x	x	x
c.2548G>T	g.38286451C>A		p.(Glu850 Ter)	F	✓		
c.2601_2602del	g.38286399_38286400del		p.(Glu868 Glyfs Ter210)	M	✓		
c.2628_2629del	g.38286372_38286373del		p.(Glu877 Glyfs Ter201)	M	✓	✓	✓
				F	✓	✓	✓
c.2719G>T	g.38286280C>A		p.(Glu907 Ter)	M	✓		
				F	✓		
c.2792del	g.38286207del		p.(Glu931 Glyfs Ter158)	M	x	x	x
				F	x	x	x
c.2864G>A	g.38286135C>T		p.(Trp955Ter)	M	✓		

Table 1. continued

HGVS DNA	Genomic nomenclature	HGVS protein	Status	Bioinformatic detection		
				Run 12 8 h	Run 24 2.5 h	Run 24 2.5 x2 h
(NM_001034853.2)	(hg38)					
c.293 T>G [AAAGGc.2908_293 T]	g.38286091_38286092insCCTTTCCCTTCCCTTCCCTTCTCC	p.(Glu978LysfsTer121)	F	✓	✓	
c.293 T>G	g.38286063_38286064del	p.(Glu980GlyfsTer98)	M	✓	✓	
c.296 A>G	g.38286036_38286037del	p.(Glu989GlyfsTer89)	M	✓	✓	
c.299 T>G	g.38286003_38286004del	p.(Glu100GlyfsTer78)	M	✓	✓	
c.302 T>G	g.38285973_38285974del	p.(Glu101GlyfsTer68)	M	✓	✓	
c.310 A>G	g.38285897_38285898del	p.(Glu103GlyfsTer63)	M	✓	✓	
c.328 G>A [c.328GdelinsC]	g.38285712_38285713delinsG	p.(Lys109HisfsTer35)	M	✓	✓	
c.330 G>A	g.38285691dup	p.(Tyr110Ser)	F	✓	✓	
c.330 G>A	g.106927220_106927221del	p.(Tyr110SerfsTer7)	M	✓	✓	

HGVS DNA (NM_001034853.2), genomic nomenclature (hg38) and HGVS protein nomenclatures are indicated for each pathogenic variant. Variants localized in the refractory region g.38,285,847 - 38,286,647 are indicated in bold. A correct bioinformatic report is indicated by the ✓ symbol and an incorrect by the × symbol.

M male, F female.

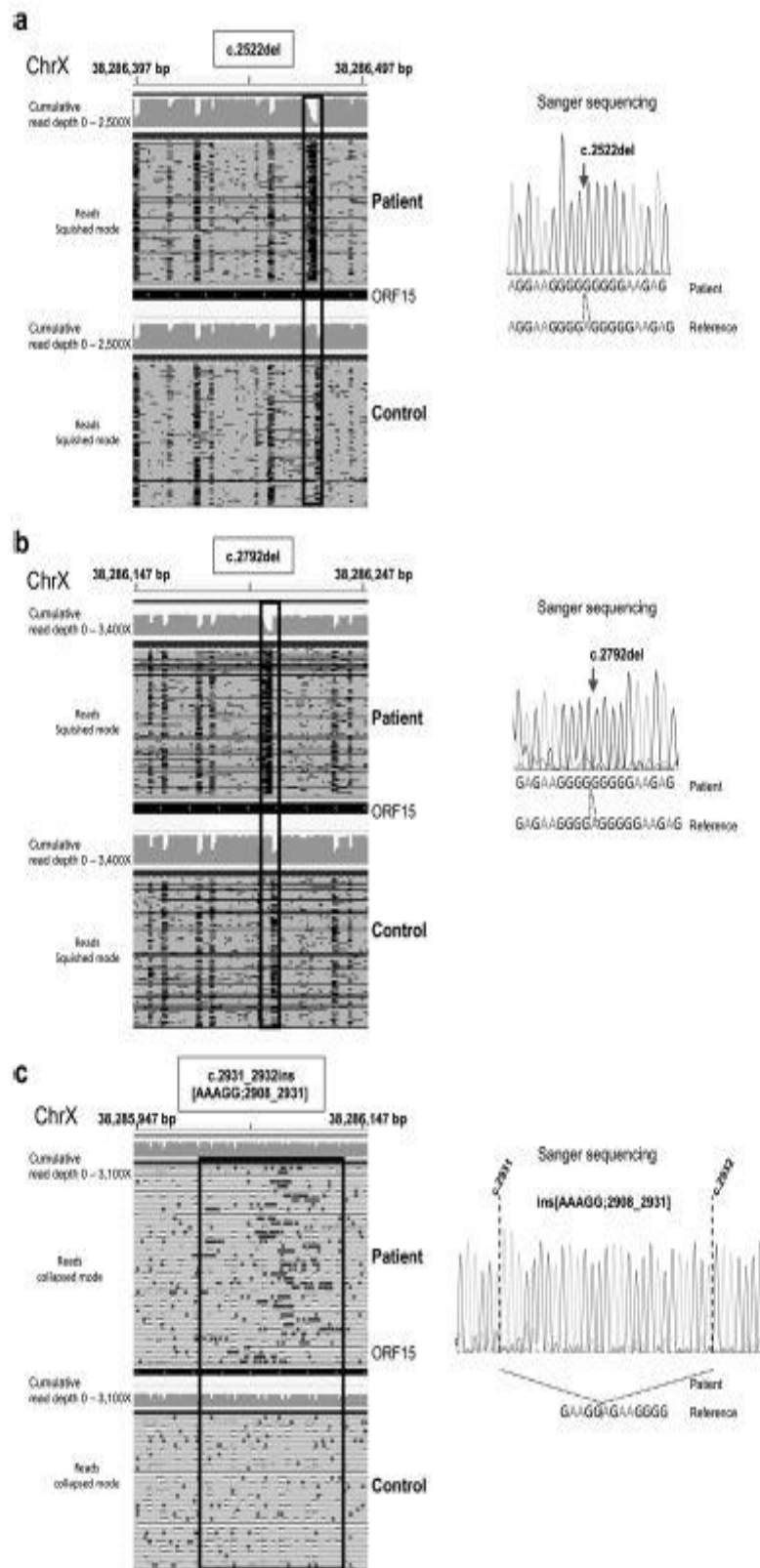


Fig. 1 Inspection of the sequenced reads using the IGV software and Sanger sequencing validation of the 3 hemizygous Clair3 undetected variants. **a** On the left: visual comparison of squished sequenced reads obtained from a male index case carrying the c.2522del and a male control. On the right: validation of the variant by Sanger sequencing. **b** On the left: visual comparison of squished sequence reads obtained from a male index case carrying the c.2792del and a male control. On the right: Validation of the variant by Sanger sequencing. **c** On the left: visual comparison of collapsed sequenced reads obtained from a male index case carrying the c.2931_2932ins[AAAGG;2908_2931] and a male control. The inserted sequences are indicated by dark gray symbols (III). On the right: Validation of the variant by Sanger sequencing. The reads show a representative subset in each IGV capture. The visual difference observed between a patient and its control is highlighted by a black frame. An arrow in Sanger electropherograms indicates the genomic position of each pathogenic variant.

Data obtained with 24 barcoded amplicons per run

In order to test and validate the possibility of increasing the number of samples in a reduced time, we performed one run with 24 selected DNA samples and used a sequencing protocol consisting of two running periods (r1 and r2) of 2.5 h, each separated by a wash step. With this approach, a total of 14 different hemizygous and/or heterozygous variants were tested. Thirteen of them were localized between positions g.38,285,847 - 38,286,647 which delimit the most difficult ORF15 region to sequence. The 14th was located at position g.38,285,712 (c.3286_3287delinsC).

Metrics.

A total of 220 Mb were called during the r1 period, which was in the range of the data obtained using the first protocol (12 amplicons/8 hour running time). As expected, the wash and reloading steps improved yield and total read count with additional 96 Mb obtained during the r2 period for a total of 316 Mb called (See Supplementary Table 1 for details). As a consequence, either r1 or r1 + r2 were efficient to cover the entire ORF15 sequence (Supplementary Fig. 4) with a mean depth of coverage of 1331 ± 422 SD and 1930 ± 606 SD, respectively.

Detection of pathogenic variants in males and females.

Concerning the male cohort, bioinformatic results acquired from the r1 sequencing phase were identical to those obtained using the first protocol, with the exception of c.2931_2932ins[AAAGG;2908_2931], which was unexpectedly reported (Table 1). An IGV-visualization of the reads was still mandatory to obtain a 100% detection rate. Additional data acquired during the r2 period did not improve the variant calling (Table 1).

Concerning the female cohort (Table 1), the r1 period allowed calling of two additional variants in comparison to the first protocol (c.2997_2998del and c.3286_3287delinsC), and two more variants were called after the additional r2 period (c.2426_2427del and c.2964_2965del). Because the heterozygous c.2931_2932ins [AAAGG;2908_2931] was reported at a very low rate (1% of the reads) in the r1 period and was not found when the total run was studied, we decided to consider it as an unidentified variant. Here again, an IGV-inspection of the reads allowed us to obtain a 100% detection rate with both r1 and r1+r2.

Data obtained for variants located in the g.38,285,847 - 38,286,647 (hg38) region were detailed in Supplementary Table 2.

DISCUSSION

Sequencing challenges inherent to the ORF15 region of RPGR are well known to laboratories offering genetic testing for RP. Because this region is a mutational hotspot, these laboratories have always sought to overcome these difficulties by developing new sequencing approaches ([9, 19–21]). The last one, described by Yahya et al. [10], using the 3rd generation sequencing technology, although promising, required additional validation before being integrated into routine molecular diagnostic workflow.

Consequently, in this study, our objective was to evaluate and validate the efficiency of this approach for a diagnostic application using a masked test including controls and ORF15 mutated DNA samples from males and females. For this purpose, a total of 32 different pathogenic variants were tested; 9 of them in hemizygous state, 3 in heterozygous state and the 20 others in both hemizygous and heterozygous states. Various pathogenic variant types were tested including substitutions, small duplications, deletions of 2–14 bp, one insertion of 29 bp, and one delins. More importantly, 20 of these variants were localized between g.38,285,847 - 38,286,647 the well-known ORF15 region difficult to sequence.

It should be noted that unlike Yahya et al., who used a two-step PCR approach to produce their library, we chose a one-step PCR approach. Indeed, considering the nature of the sequence, we thought it was relevant to reduce the number of PCR cycles as much as possible in order to limit amplification bias. Analyses with 12 samples and an 8 h running time were first performed. In accordance with the previous observations of Yahya et al. [10], metrics recorded in the final Nanopore report highlighted a spectacular decrease of active pores, which could be due to the presence of blocking secondary structures in the ORF15 repetitive sequence. The Oxford Nanopore technical services teams have not yet been able to explain this feature. In light of this problem, the possibility to use a Flongle flow cell that has only 126 pores (R9 flow cell : 512 pores) in order to screen a smaller batch of patients must be tested to define the maximum number of samples that could be pooled per run.

Under our experimental conditions, all hemizygous variants localized outside the g.38,285,847 - 38,286,647 sequence (n = 10) were correctly detected by our bioinformatic analysis. Out of the 19 variants localized in the refractory region, only 3 could not be detected. Two variants (c.2506del and c.2792del) created homo-polymers of 9 guanines, which are already known as patterns leading to high error rates when using the Nanopore sequencer (Oxford Nanopore Technologies) [22], and the third one is a 29 bp insertion that includes 21 bp of purine-rich ORF15 sequence (c.2931_2932ins[AAAGG;2908_2931]). Even though this hemizygous insertion was correctly reported when applying our r1 + r2 protocol, the two deletions remained unidentified. This technical limitation will possibly be resolved by using R10.4 flow cells and the latest chemistry (kit SQKLSK114) that seem to improve homo-polymer calling accuracy for lengths up to 10 [23].

A re-analysis of the data obtained from the 5 undetected ORF15-individuals (3 females and 2 males) with the r1 + r2 protocol was performed using the Dorado open source basecaller and the super accuracy mode (<https://github.com/nanoporetech/dorado/>). Unfortunately it did not show any additional detection.

In our study, the limits of the bioinformatics detection for certain variants have been more noticeable in women because of their heterozygous status. Lowering the allele frequency (AF) thresholds in Clair3 does not appear to be a solution, as visual explorations of the reads by the IGV tool clearly show that AF thresholds, defined in the default setting (0.08 for SNV selection as a candidate position and 0.15 for indels), were reached for all heterozygous variants.

From our point of view, this detection problem is linked to misalignment. As an example, the c.2931_2932ins [AAAGG;2908_2931] variant is always undoubtedly misaligned (Fig. 1c) and consequently, when present in heterozygous state, is under-detected or remains undetected depending on the bioinformatics treatment applied (1% in the r1 protocol and 0% in the re-analysis of the same run using Dorado and the super accuracy mode).

In light of these observations, we propose to screen the ORF15 region using 2.5-h sequencing runs with 24 samples per run, followed by a careful IGV-visualization of the reads for negative affected individuals. Presently, in order to deliver a molecular diagnostic result according to good laboratory practice, it is still recommended to validate the presence of these variants by Sanger sequencing. Furthermore, because non-pathogenic insertions or duplications located in the g.38,285,847 - 38,286,647 interval are not always correctly called by the bioinformatic pipeline and sometimes lead to suspect IGV-visualization, we recommend a Sanger sequencing in case of doubt. Although this Sanger sequencing step is delicate, it is limited to a small region encompassing the variant of interest and remains acceptable in routine practice.

Various sequencers based on different long-read sequencing technologies are available on the market, each with its own strengths and limitations. In this study, we used the pocket-sized MinION device of Oxford Nanopore technology and demonstrated its capacity to quickly screen the refractory ORF15 region of multiplexed samples with a 100% detection rate compatible with a routine molecular diagnosis for both males and female carriers. Analyzing reads with the IGV tool, which enabled us to achieve this 100% detection rate, may appear to be a major limitation of this long-reads approach. However, our experience shows that after learning how to read IGV data, using a patient/control comparison, it is quite possible to rapidly detect pathogenic variants.

In the era of gene therapy, it is crucial to provide a proper molecular diagnosis and genetic counseling to the index cases and their families, considering genotype-phenotype correlation [18], especially for RPGR, one of the major RP-associated genes [24]. Consequently, the cost-effective approach validated in this study should be considered for all individuals presenting with a presumed diagnosis of XLRP before conducting massive parallel sequencing studies.

DATA AVAILABILITY

The data that support this study are available from the corresponding author upon reasonable request.

References

1. Vinikoor-Imler LC, Simpson C, Narayanan D, Abbasi S, Lally C. Prevalence of RPGR-mutated X-linked retinitis pigmentosa among males. *Ophthalmic Genet.* 2022;43:581–8. - [DOI](#) - [PubMed](#)
2. Tee JJL, Yang Y, Kalitzeos A, Webster A, Bainbridge J, Michaelides M. Natural History Study of Retinal Structure, Progression, and Symmetry Using Ellipsoid Zone Metrics in RPGR-Associated Retinopathy. *Am J Ophthalmol.* 2019;198:111–23. - [DOI](#) - [PubMed](#) - [PMC](#)
3. Kurata K, Hosono K, Hayashi T, Mizobuchi K, Katagiri S, Miyamichi D, et al. X-linked Retinitis Pigmentosa in Japan: Clinical and Genetic Findings in Male Patients and Female Carriers. *Int J Mol Sci.* 2019;20:1518. - [DOI](#) - [PubMed](#) - [PMC](#)
4. Talib M, van Schooneveld MJ, Van Cauwenbergh C, Wijnholds J, Ten Brink JB, Florijn RJ, et al. The Spectrum of Structural and Functional Abnormalities in Female Carriers of Pathogenic Variants in the RPGR Gene. *Invest Ophthalmol Vis Sci.* 2018;59:4123–33. - [DOI](#) - [PubMed](#)
5. Vervoort R, Lennon A, Bird AC, Tulloch B, Axton R, Miano MG, et al. Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nat Genet.* 2000;25:462–6. - [DOI](#) - [PubMed](#)
6. Kirschner R, Erturk D, Zeitz C, Sahin S, Ramser J, Cremers FP, et al. DNA sequence comparison of human and mouse retinitis pigmentosa GTPase regulator (RPGR) identifies tissue-specific exons and putative regulatory elements. *Hum Genet.* 2001;109:271–8. - [DOI](#) - [PubMed](#)
7. Hong DH, Pawlyk BS, Adamian M, Sandberg MA, Li T. A single, abbreviated RPGR-ORF15 variant reconstitutes RPGR function in vivo. *Invest Ophthalmol Vis Sci.* 2005;46:435–41. - [DOI](#) - [PubMed](#)
8. Di Iorio V, Karali M, Melillo P, Testa F, Brunetti-Pierri R, Musacchia F, et al. Spectrum of Disease Severity in Patients With X-Linked Retinitis Pigmentosa Due to RPGR Mutations. *Invest Ophthalmol Vis Sci.* 2020;61:36. - [DOI](#) - [PubMed](#) - [PMC](#)
9. Huang XF, Wu J, Lv JN, Zhang X, Jin ZB. Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: a complementary approach to clinical genetic diagnostic testing. *Genet Med.* 2015;17:307–11. - [DOI](#) - [PubMed](#)
10. Yahya S, Watson CM, Carr I, McKibbin M, Crinnion LA, Taylor M, et al. Long-Read Nanopore Sequencing of RPGR ORF15 is Enhanced Following DNase I Treatment of MinION Flow Cells. *Mol Diagn Ther.* 2023;27:525–35. - [DOI](#) - [PubMed](#) - [PMC](#)
11. Maggi J, Roberts L, Koller S, Rebello G, Berger W, Ramesar R. De Novo Assembly-Based Analysis of RPGR Exon ORF15 in an Indigenous African Cohort Overcomes Limitations of a Standard Next-Generation Sequencing (NGS) Data Analysis Pipeline. *Genes.* 2020;11:800. - [DOI](#) - [PubMed](#) - [PMC](#)
12. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol.* 2021;39:1348–65. - [DOI](#) - [PubMed](#) - [PMC](#)
13. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics.* 2021;37:4572–4. - [DOI](#) - [PubMed](#) - [PMC](#)
14. Zheng Z, Li S, Su J, Leung AWS, Lam TW, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci.* 2022;2:797–803. - [DOI](#) - [PubMed](#)
15. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17:122. - [DOI](#) - [PubMed](#) - [PMC](#)

16. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14:178–92. - [DOI](#) - [PubMed](#)
17. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat.* 2016;37:564–9. - [DOI](#)
18. Nassisi M, De Bartolo G, Mohand-Said S, Condroyer C, Antonio A, Lancelot ME, et al. Retrospective Natural History Study of RPGR-Related Cone- and Cone-Rod Dystrophies While Expanding the Mutation Spectrum of the Disease. *Int J Mol Sci.* 2022;23:7189. - [DOI](#) - [PubMed](#) - [PMC](#)
19. Li J, Tang J, Feng Y, Xu M, Chen R, Zou X, et al. Improved Diagnosis of Inherited Retinal Dystrophies by High-Fidelity PCR of ORF15 followed by Next-Generation Sequencing. *J Mol Diagn.* 2016;18:817–24. - [DOI](#) - [PubMed](#)
20. Chiang JPW, Lamey TM, Wang NK, Duan J, Zhou W, McLaren TL, et al. Development of High-Throughput Clinical Testing of RPGR ORF15 Using a Large Inherited Retinal Dystrophy Cohort. *Invest Ophthalmol Vis Sci.* 2018;59:4434–40. - [DOI](#) - [PubMed](#)
21. Nash BM, Ma A, Ho G, Farnsworth E, Minoche AE, Cowley MJ, et al. Whole Genome Sequencing, Focused Assays and Functional Studies Increasing Understanding in Cryptic Inherited Retinal Dystrophies. *Int J Mol Sci.* 2022;23:3905. - [DOI](#) - [PubMed](#) - [PMC](#)
22. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. *PLoS One.* 2021;16:e0257521. - [DOI](#) - [PubMed](#) - [PMC](#)
23. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods.* 2022;19:823–6. - [DOI](#) - [PubMed](#) - [PMC](#)
24. Martinez-Fernandez de la Camara C, Cehajic-Kapetanovic J, MacLaren RE. Emerging gene therapy products for RPGR-associated X-linked retinitis pigmentosa. *Expert Opin Emerg Drugs.* 2022;27:431–43. - [DOI](#) - [PubMed](#)