Article

Towards a Long-Read Sequencing Approach for the Molecular Diagnosis of RPGR^ORF15 Genetic Variants

Gabriele Bonetti [1,2,*,†], William Cozza [3,†], Andrea Bernini [4], Jurgen Kaftalli [3], Chiara Mareso [3], Francesca Cristofoli [3], Maria Chiara Medori [1], Leonardo Colombo [5], Salvatore Martella [5], Giovanni Staurenghi [6], Anna Paola Salvetti [6], Benedetto Falsini [7,8], Giorgio Placidi [7], Marcella Attanasio [9], Grazia Pertile [9], Mario Bengala [10], Francesca Bosello [11], Antonio Petracca [12], Fabiana D'Esposito [3,13,14], Benedetta Toschi [15], Paolo Lanzetta [16,17], Federico Ricci [18], Francesco Viola [19], Giuseppe Marceddu [3] and Matteo Bertelli [1,3,20]

1   MAGI's LAB, 38068 Rovereto, Italy; chiara.medori@assomagi.org (M.C.M.);
    matteo.bertelli@assomagi.org (M.B.)
2   Department of Pharmaceutical Sciences, University of Perugia, 06123 Perugia, Italy
3   MAGI Euregio, 39100 Bolzano, Italy; amministrazione@assomagi.org (W.C.);
    jurgen.kaftalli@assomagi.org (J.K.); chiara.mareso@assomagi.org (C.M.); f.desposito@imperial.ac.uk (F.D.);
    giuseppe.marceddu@assomagi.org (G.M.)
4   Department of Biotechnology, Chemistry and Pharmacy, University of Siena, 53100 Siena, Italy;
    andrea.bernini@unisi.it
5   Department of Ophthalmology, ASST Santi Paolo e Carlo Hospital, University of Milan, 20142 Milan, Italy;
    leonardo.colombo.82@gmail.com (L.C.); martellasalvatore7@gmail.com (S.M.)
6   Eye Clinic, Department of Biomedical and Clinical Science, Luigi Sacco Hospital, University of Milan,
    20157 Milan, Italy; giovanni.staurenghi@unimi.it (G.S.); paola.anna.salvetti@gmail.com (A.P.S.)
7   UOC Oculistica, Fondazione Policlinico Universitario "A. Gemelli" IRCCS, Largo Gemelli 8,
    00168 Rome, Italy; giorgio.placidi@guest.policlinicogemelli.it (G.P.)
8   Istituto di Oftalmologia, Università Cattolica del Sacro Cuore, Largo Francesco Vito 1, 00168 Rome, Italy
9   Ospedale Sacrocuore Don Calabria, Viale Luigi Rizzardi, 4, 37024 Negrar di Valpolicella, Italy;
    marcella.attanasio@sacrocuore.it (M.A.); grazia.pertile@sacrocuore.it (G.P.)
10 Medical Genetics Unit, Department of Oncohematology, Policlinico Tor Vergata, 00133 Rome, Italy;
    mario.bengala@ptvonline.it
11 Department of Surgical Sciences, Dentistry, Paediatrics and Gynaecology, Section of Ophthalmology,
    University of Verona, 37134 Verona, Italy; francesca.bosello87@gmail.com
12 Division of Medical Genetics, Fondazione IRCCS-Casa Sollievo della Sofferenza,
    71013 San Giovanni Rotondo, Italy; a.petracca@operapadrepio.it
13 Imperial College Ophthalmic Research Group (ICORG) Unit, Imperial College, London NW1 5QH, UK
14 Eye Clinic, Department of Neurosciences, Reproductive Sciences and Dentistry, University of Naples
    Federico II, 80138 Naples, Italy
15 Section of Medical Genetics, Department of Medical and Oncological Area, University Hospital of Pisa,
    56126 Pisa, Italy; b.toschi@ao-pisa.toscana.it
16 Department of Medicine-Ophthalmology, University of Udine, 33100 Udine, Italy; paolo.lanzetta@uniud.it
17 Istituto Europeo di Microchirurgia Oculare (IEMO), 33100 Udine, Italy
18 Department of Experimental Medicine, Tor Vergata University of Rome, Viale Oxford, 00133 Rome, Italy;
    rccfrc00@gmail.com
19 Department of Ophthalmology, Fondazione IRCCS Cà Granda, Clinica Regina Elena, 20122 Milan, Italy;
    francesco.viola@unimi.it
20 MAGISNAT, Atlanta Tech Park, 107 Technology Parkway, Peachtree Corners, GA 30092, USA
*   Correspondence: gabriele.bonetti@assomagi.org; Tel.: +39-0365-62-061
†   These authors contributed equally to this work.

Abstract: Sequencing of the low-complexity ORF15 exon of RPGR, a gene correlated with retinitis pigmentosa and cone dystrophy, is difficult to achieve with NGS and Sanger sequencing. False results could lead to the inaccurate annotation of genetic variants in dbSNP and ClinVar databases, tools on which HGMD and Ensembl rely, finally resulting in incorrect genetic variants interpretation. This paper aims to propose PacBio sequencing as a feasible method to correctly detect genetic variants in low-complexity regions, such as the ORF15 exon of RPGR, and interpret their pathogenicity by structural studies. Biological samples from 75 patients affected by retinitis pigmentosa or cone dystrophy were analyzed with NGS and repeated with PacBio. The results showed that NGS has a

low coverage of the ORF15 region, while PacBio was able to sequence the region of interest and detect eight genetic variants, of which four are likely pathogenic. Furthermore, molecular modeling and dynamics of the RPGR Glu-Gly repeats binding to TTLL5 allowed for the structural evaluation of the variants, providing a way to predict their pathogenicity. Therefore, we propose PacBio sequencing as a standard procedure in diagnostic research for sequencing low-complexity regions such as RPGR^{ORF15}, aiding in the correct annotation of genetic variants in online databases.

1. Introduction

Next-generation sequencing (NGS) was a breakthrough in molecular biology, providing a cheaper and faster method for sequencing DNA compared to the previous techniques, such as Sanger sequencing [1]. NGS techniques, among which Illumina is the most represented, rely on specific laboratory procedures, specifically DNA fragmentation, DNA end-repair, adapter ligation, surface attachment, and in situ amplification. A huge amount of data is produced, and thanks to bioinformatics analysis, it is possible to detect genetic mutations like single-nucleotide polymorphisms (SNPs) and indels [1,2]. NGS is considered a "short-reading" approach because it permits the sequencing of short fragments of DNA of 300 bp length (on average), this being its main limitation because it impedes the sequencing of low-complexity and/or GC-rich regions [3]. NGS strongly supports molecular genetics, allowing us to identify many genetic variants and to define the genetic onset of many genetic diseases. However, its limitations impede the study of many genes, which present low-complexity regions, and of specific chromosomal regions, in which massive arrays of tandem repeats predominate [3]. These limitations require developing a new sequencing technique: Third-generation sequencing (TGS). TGS, unlike NGS, comprises long-read sequencing approaches and is represented mainly by Pacific Biotechnology (PacBio) or Oxford Nanopore [2,3].

PacBio works on circular DNA fragments, while Oxford Nanopore works on linear DNA fragments. Moreover, PacBio relies on fluorescently labeled nucleotides for nucleotide detection, while Oxford Nanopore relies on an electronic signal disruption caused by the passage of DNA in a nanopore. The PacBio technique can analyze sequences up to 300 Kb, while Oxford Nanopore can analyze sequences up to 4 Mb. Finally, they can both detect epigenetic modifications. These techniques also have some disadvantages. PacBio is expensive, both in the instrumentation and in the sequencing cost. Moreover, the accuracy of these techniques was originally low, and it did not permit their scalability in a diagnostic setting. PacBio accuracy has recently increased dramatically, reaching 99.99%, while Oxford Nanopore recently reached 99.9% [2–4].

This work aims to propose PacBio sequencing for diagnostic applications, focusing on *RPGR^{ORF15}* sequencing. *RPGR* is a gene related to retinitis pigmentosa (RP) and cone dystrophy. Mutations to *RPGR* are responsible for over 70% of the X-linked Retinitis pigmentosa cases and for over 73% of all cone and rode cone dystrophy cases [5,6]. *RPGR* has two isoforms, namely *RPGR^{default}* and *RPGR^{ORF15}*. *RPGR^{default}* comprises 19 exons, while *RPGR^{ORF15}* shares the first 14 exons with *RPGR^{default}*, along with a distinct ORF15 exon [7], [8]. The first ten exons code for an RCC1 (regulator of chromatic condensation 1-like) domain, which is involved in the regulation of small GTPases, while the ORF15 exon is a 1 kb-long, highly repetitive, low complexity, purine-rich region, terminating with a C-terminal tail region with unknown function (basic domain) [7,9]. The low complexity of ORF15 makes NGS or Sanger sequencing ineffective for routine sequencing, thus increasing the possibility of false negative or false positive results [8,10]. Considering that ORF15 is a mutation hotspot of *RPGR* (Figure S1), a great number of variants associated with this region are uploaded in the database of single nucleotide variants (dbSNP), commonly used worldwide to detect variants and diagnose hereditary genetic diseases and on which online

databases and tools such as HGMD and Ensembl rely [11,12]. Indeed, most of the genetic variants correlated to Retinitis Pigmentosa have been found in *RPGR*[ORF15] [13]; thus, a robust, accurate, and scalable test to sequence ORF15 is necessary for a precise genetic diagnosis. Finally, precise genetic variants identification will permit the implementation of personalized medicine strategies, like gene therapy, thus providing a new treatment option to RP patients [8].

Regarding its molecular structure, the photoreceptor-specific ORF15 variant of *RPGR* RPGR[ORF15] contains multiple Glu-Gly tandem repeats and a C-terminal basic domain unknown in function and is localized to the connective cilium where it is thought to regulate cargo transport. The tubulin tyrosine ligase like-5 (TTLL5) glutamylates RPGR[ORF15] in its Glu-Gly–rich repetitive region, which contains motifs homologous to the α-tubulin C-terminal tail; loss of glutamylation has pathological consequences in developing retinal dystrophy [7]. The C-terminal basic domain of RPGR[ORF15] interacts with the TTLL5 noncatalytic cofactor interaction domain, which is unique among the glutamylases of the TTLL family and targets TTLL5 to glutamate RPGR as a result. TTLL5 is the only glutamylase in the TTLL family that interacts with RPGR[ORF15] when expressed transiently in cells [7]. While the association of TTLL5 variants with loss of glutamylation due to enzymatic deficiency has already been characterized [14–16], less is known for the loss of glutamylation due to variants affecting the RPGR[ORF15]. The interaction of the TTLL enzyme family with regions rich in Glu-Gly repeats has been studied in detail for some cases as glutamylation of tubulin by TTLL4 and TTLL6 [17], and while the general mechanism of initiation and elongation can be extended to the whole family, specific characterization of the TTLL5/RPGR[ORF15] interaction to support the effect of variants on the Glu-Gly-rich region is lacking.

2. Results

2.1. Study Cohort Characteristics

Table 1 reports the clinical characteristics of the probands analyzed for this study.

Table 1. Clinical data of the patients. The pathology was considered familiar if other cases were present in their clinical familiar history. All the analyzed patients were unrelated.

Characteristics		Case Subjects (n = 75)
Age (years)	Mean	51 ± 17
	Median	52 ± 17
Females/Males		35/40 (47%/53%)
Diagnosis	Hereditary non-syndromic retinal dystrophies	5 (7%)
	Retinitis pigmentosa	56 (75%)
	Cone dystrophy	8 (10%)
	Macular distrophy	6 (8%)
Age of onset (years)	Mean	25 ± 16
	Median	24 ± 16
	Unknown	n = 8
Familiarity	Familiar	26 (35%)
	Sporadic	40 (53%)
	Unknown	9 (12%)

2.2. NGS Sequencing Coverage

Figure S2 shows four examples of pitfalls in the NGS sequencing coverage of *RPGR*[ORF15] exon 15. Table S1 summarizes the NGS sequencing coverage of *RPGR* ORF15 exon in the analyzed samples. As can be seen from Figure 1, NGS sequencing coverage is low, and most of the samples (73 patients) had a sequencing coverage between 50.0% and 65.0%.
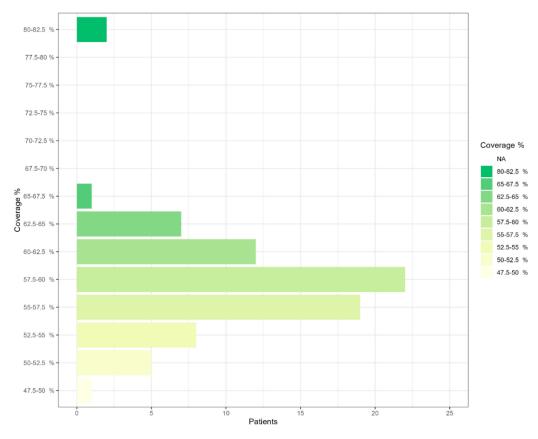
Figure 1. NGS sequencing coverage of RPGR ORF15 exon in the analyzed samples.

2.3. PacBio Sequencing Results

Table 2 reports the genetic variants identified in the *RPGR* gene with PacBio sequencing. On average, 98.12% of reads mapped to the entire genome. Within the region, coverage averaged 3587.44 with a standard deviation of 436.58. PacBio sequencing was able to detect in eight unrelated patients eight genetic variants that were not identified with NGS sequencing, four of which were predicted to be likely pathogenic. Table 3 compares the clinical characteristics between the probands in which *RPGR* variants were identified and the other probands.

Table 2. Genetic variants identified in RPGR gene in patients with PacBio sequencing.

ID	Nucleotide Variant	rsID	Verdict	Zygosity
1	NM_001034853.2:c.2919_2940dup	/	Likely Pathogenic	0/1
2	NM_001034853.2:c.2203_2226del	rs768423834	Uncertain Significance	0/1
3	NM_001034853.2:c.2203_2226del	rs768423834	Uncertain Significance	1/1
4	NM_001034853.2:c.2820_2841dup	/	Likely Pathogenic	0/1
5	NM_001034853:c.1871A>C	/	Uncertain Significance	0/1
6	NM_001034853:c.3423G>T	/	Uncertain Significance	1/1
7	NM_001034853.2:c.3262_3263insA	/	Likely Pathogenic	0/1
8	NM_001034853.2:c.2820_2841dup	/	Likely Pathogenic	0/1

Table 3. Clinical characteristics of the probands in which *RPGR* variants were identified and of the other probands. The pathology was considered familiar if other cases were present in their clinical familiar history. All the analyzed patients were unrelated.

Characteristics		Patients with *RPGR*[ORF15] Mutation (n = 8)	Case Subjects (n = 67)	*p*-Value
Age (years)	Mean	35 ± 11	53 ± 15	<0.05
	Median	36 ± 11	53 ± 15	
Females/Males		4/4 (50%/50%)	31/36 (46%/54%)	-
Diagnosis	Hereditary non-syndromic retinal dystrophies	0 (0%)	5 (7%)	-
	Retinitis pigmentosa	3 (37.5%)	53 (79%)	
	Cone dystrophy	2 (25%)	6 (9%)	
	Macular dystrophy	3 (37.5%)	3 (5%)	
Age of onset (years)	Mean	13 ± 11	27 ± 16	<0.05
	Median	10 ± 11	27 ± 16	
	Unknown	n = 0	n = 8	
Familiarity	Familiar	2 (25%)	24 (36%)	-
	Sporadic	3 (37.5%)	37 (55%)	
	Unknown	3 (37.5%)	6 (9%)	

2.4. Structural Analysis of TTLL5 Core Domain Binding

The RPGR[ORF15] sequence diverges from the default variant, which is not glutamylated, at its C-terminal half, consisting of a region rich in Glu-Gly repeats followed by a C-terminal basic domain (Figure 2). The repeat region contains glutamate-rich motifs (GEEEG) homologous to the a-tubulin C-terminal tail and is glutamylated by the TTLL5 core domain. The basic domain, highly conserved among vertebrates, is crucial for recruiting TTLL5 [7,18].
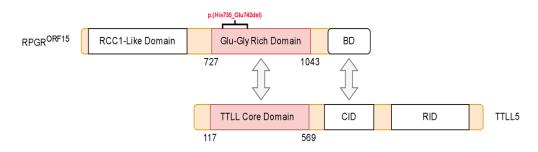

Figure 2. Graphical representation of RPGR[ORF15] and TTLL5 interaction. The basic domain (BD) of RPGR[ORF15] recruits TTLL5 and binds to its noncatalytic CID domain. The in-frame deletion is indicated in red.

The variants we identified are of the stop gained type, leading to the loss of almost the entire length of the ORF region and basic domain, except for the NM_001034853.2:c.2203_2226del p.(His735_Glu742del) (Figure 2). Such a variant leads to the loss of a GEGE tandem repeat but retains the structure of RPGR[ORF15] and was therefore used as a model for the interaction. The model of the TTLL5 core domain was obtained by homology modeling to the most recent evidence of TTLL6 structure, where an initiation analog is bound to the active site. The analog mimics a di-Glu peptide where the donor glutamate is linked to the γ-carboxylic acid of acceptor glutamate through a phosphinate. The di-Glu was split by removing the phosphinate and was transformed into the donor and the acceptor glutamate. The latter was interpreted as Glu737 in RPGR[ORF15], the glutamination site in the deleted peptide segment from the NM_001034853.2:c.2203_2226del p.(His735_Glu742del) variant. The Glu737 was then expanded at each terminus to reproduce the structure of the deleted peptide and then docked into the crevice of the active site by holding Glu737 fixed, working as an

anchor. Such experiment-based positional restraint imposes the right orientation for the peptide to be docked and lowers its conformational space, thus improving the accuracy of the docked structures. The final model shows the peptide GEEEHGE737GEEEE filling the positively charged crevice centered on the active site with each glutamate sidechain closely interacting with at least one lysine/arginine residue (Figure 3). Interestingly, the only non-Glu/non-Gly residue, the His375, finds its sidechain deeply inserted in a sub pocket of the crevice and involved in a π-interaction with Ser307; such interactions have been recently proposed as relevant in protein-protein binding with stabilization energy ranging from −20 kJ mol^{-1} to −40 kJ mol^{-1} [19].
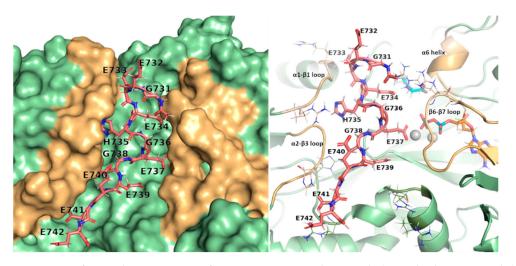

Figure 3. Surface and cartoon view of TTLL5 active site. The peptide (in pink, shown as sticks) fits into the crevice centered at the active site, with Glu737 directed towards ADP. Glutamates align to form electrostatic interactions with the positively charged residues (shown as lines in the cartoon view), stabilizing the peptide into the pocket. The structural components surrounding the peptide chain are shown in light orange (loops α1-β1, α2-β3, β6-β7, and helix α6), with the rest of the protein shown in green; the donor glutamate is shown in cyan; magnesium ions and ADP are shown in gray and orange, respectively.

Molecular dynamics simulations of free and bound TTLL5 were performed to further characterize the binding of the RPGR^{ORF15} glutamate-rich segments. Comparing the average structure of the bound TTLL5 simulations with the experimental structure of TTLL6 bound to the initial analog (PDB entry 6VZW) showed that the β6-β7 loop adopts a different conformation (Figure 4b), in accordance with the results reported in [20] on the importance of this loop to the unique activity of TTLL5. Jointly with loops α1-β1 and α2-β3, it forms the crevice where the glutamate-rich segment can bind (Figure 3). Furthermore, the β6-β7 is also rearranged in the bound simulation, adapting to an open conformation to accommodate the RPGR^{ORF15} segments (Figure 4A), as also shown by the center of mass (COM) distance between the respective loops forming the crevice (Figure S3).

To understand how the binding of different glutamate-rich segments would affect the active site of TTLL5, we performed the modeling and molecular dynamics simulations of two more peptide structures of the same length of the Glu737 centered GEEEHGE737GEEEE: (i) the peptide resulting from the NM_001034853.2:c.2203_2226del p.(His735_Glu742del) variant, centered on E734, EEGGEEE734GDREE; (ii) a peptide reproducing different glutamylation site, centered on E870, EGEGEEE870GEEGE. To assess the structural stability of TTLL5 core domain bound to such different segments, the RMSD of each bound peptide was evaluated. The results showed the two wildtype segments GEEEHGE737GEEEG and EGEGEEE870GEEGE to be stable, while the post-deletion sequence EEGGEEE734GDREE showed significant deviation, as shown in Figure 5.
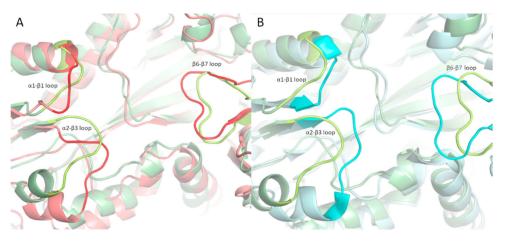
Figure 4. Loop configurations around the binding site. (**A**) Comparison of the binding cavity between the average structure of the bound (light green) and free (light red) TTLL5. The β6-β7 loop is slightly displaced to accommodate the peptide chain. (**B**) Comparison of the binding cavity between the average structure of the TTLL5 bound to the GEEEHGE737GEEEE (light green) and TTLL6 (light cyan). The open configuration is evident in TTLL5, with a notable displacement of the β6-β7 loop. Loops are marked with higher opacity.
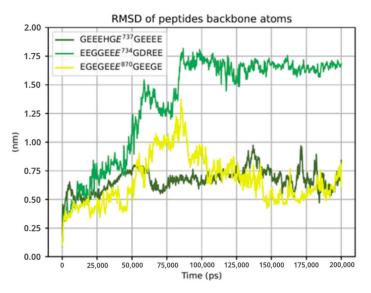

Figure 5. RMSD of each peptide chain bound to TTLL5. While the variant (shown in light green) is characterized by large movements for the first 100 ns, the two wildtype segments (shown in dark green and yellow) are subject to fewer fluctuations, especially the one with Glu737 as the glutamylation site.

3. Discussion

Genetic mutations to *RPGR* are responsible for over 70% of X-linked Retinitis pigmentosa and rode cone dystrophy cases [5,6,21]. Nevertheless, sequencing of *RPGR*[ORF15], one of *RPGR*'s two main isoforms, is still challenging with NGS, and it can be associated with a considerable amount of false results [8,10]. TGS has recently been proposed for diagnostic purposes, considering that its accuracy has highly improved in recent years [22,23]. Thus, this study aimed to propose PacBio for diagnostic sequencing, focusing on *RPGR*[ORF15]. Furthermore, molecular modeling and dynamics of the RPGR GLU-GLY repeats binding to the TTLL5 core domain allowed for the structural evaluation of the variants, providing a way to predict their pathogenicity.

As shown in Figure 1 and Figure S2, NGS is not suitable for sequencing *RPGR*[ORF15] because its coverage highly decreases in exon 15, resulting in possible false results. On the

other hand, in our study, PacBio sequencing identified eight genetic variants of $RPGR^{ORF15}$ that were not detected by NGS sequencing in the analyzed patients (Table 2). Among them, four were predicted to be likely pathogenic. In addition, the comparison of the clinical data of the patients with $RPGR^{ORF15}$ mutations with the clinical data of the other analyzed patients revealed that the age and the age of onset were significantly different ($p < 0.01$), suggesting that $RPGR^{ORF15}$ mutations could lead to an earlier onset pathology (Table 3). These results support the use of PacBio sequencing for genetic diagnosis of $RPGR^{ORF15}$ mutation, considering that NGS did not identify many genetic variants and that these variants seem to be involved in retinal dystrophy onset. NGS of the low-complexity region can also provide false results, which in turn could lead to the inaccurate annotation of genetic variants in dbSNP [24] and ClinVar [25] databases, tools on which HGMD [11] and Ensembl [12] rely, finally resulting in incorrect genetic variant interpretation and impeding an efficient genetic diagnosis.

Considering the occurrence of pathogenic variants in the unstructured ORF15 domain, we modeled the interaction between the RPGR^{ORF15} Glu-Gly rich segments and the TTLL5 core domain to derive structural information about the interaction with the observed variants. Indeed, the results showed that Glu-Gly rich segments of RPGR^{ORF15} similarly bind TTLL5 to that observed in the α-tubulin tail, with loops α1-β1, α2-β3, and β6-β7 adapting to an open conformation and stabilizing the bound segment by forming electrostatic interactions with the negatively charged glutamates.

The main limitation of this study is the small number of patients analyzed with PacBio sequencing. A bigger cohort of patients could support the main aim of the article, proving that TGS is the best solution for sequencing low-complexity regions and that NGS could result in false negative or false positive results. Moreover, further molecular modeling studies on the uncharacterized interaction between RPGR^{ORF15} basic domain and the cofactor interaction domain of TTL5 could yield insight into the pathogenicity of the variants found in this region. Finally, more detailed clinical data could be useful in finding possible correlations between genetic variants in $RPGR^{ORF15}$ and disease severity. Nevertheless, long-read sequencing could be a feasible approach to sequence low-complexity regions for diagnostic purposes, considering NGS as a first option for the other regions. At the same time, molecular dynamics can add useful knowledge of proteins involved in disease onset, increasing our chance to identify pathogenic variants, and searching for possible new therapeutic targets.

4. Materials and Methods

4.1. Subjects and Samples

We analyzed 75 Caucasian subjects diagnosed with retinitis pigmentosa or cone dystrophy. All patients were recruited and underwent detailed clinical examinations by different Italian hospitals. All patients underwent pre-test counseling, during which clinical data—including personal and family history—were collected and evaluated. The patients were informed about the significance of genetic testing. All of them gave their written informed consent in compliance with the Declaration of Helsinki. Genomic DNA was isolated from peripheral blood or saliva using a commercial kit (SaMag Blood DNA Extraction Kit (Sacace Biotechnologies, Como, Italy)) according to the manufacturer's instructions.

4.2. NGS Sequencing

A large custom panel [approximately 2.4 Mb cumulative target length (GRCh38/hg38)], encompassing several genes related to retinitis pigmentosa and cone dystrophy among which *RPGR*, were used for NGS analysis. The DNA probe set was designed to capture the coding exons and flanking regions of each gene of the panels using the Twist Custom Panel Design Technology (Twist Bioscience, South San Francisco, CA, USA https://www.twistbioscience.com/products/ngs accessed on 20 September 2023). The subpanel of analyzed genes were selected on the basis of literature or databases [Human

Gene Mutation Database (HGMD Professional), Online Mendelian Inheritance in Man (OMIM), Orphanet, NCBI GeneReviews, NCBI PubMed and specific database].

Library preparation from genomic DNA samples was performed according to the manufacturer's protocol using the Twist Library Preparation EF Kit and Twist Universal Adapter (UDI) System with Standard Hybridization Target Enrichment (Twist Bioscience). Briefly for library preparation, 50 ng of each genomic DNA was enzymatically fragmented to yield fragments of 450–550 bp, and end repaired and dA-Tailed in the same reaction. Then, the Twist universal adaptor was ligated on fragments, SPRI purified and enriched by 7 PCR cycles with Twist Unique Dual Index Primers. Next, 187.5 ng of each purified Libraries was then hybridized to the Twist oligo probe capture library for 16 hr in a twelve-plex reaction. After hybridization, washing, and elution, the eluted fraction was PCR-amplified with 9 cycles and purified. A 150 bp paired-end reads sequencing was performed on MiSeq personal sequencer (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. A total of 24 pool library samples were loaded on Miseq using MiSeq V3 kit.

4.3. SMRTbell Library Preparation and TGS Sequencing

The SMRTbell library to sequence $RPGR^{ORF15}$ was prepared following the PacBio protocol guidelines Procedure & Checklist—Preparing SMRTbell® Libraries using PacBio® Barcoded Universal Primers for Multiplexing Amplicons—Part Number 101–791-800 Version 02 (April 2020) [26]. All assays were performed as already described in our publication [27].

For the first-round PCR, forward target-specific primer (5′-GACAG-TTACATGGA AGGTGCAA-3′) and reverse target-specific primer (5′-TACCAG-TGCCTCCTATTGTCTT-3′) tailed to F/R universal sequences were designed to amplify the ~1.7 kb DNA fragment spanning the entire $RPGR^{ORF15}$ gene exon 15 sequence (NM_001034853). Primer design was performed using the freely available program Primer3web [28] and, to avoid mispriming, the primer couple was tested in silico using NCBI PrimerBLAST [29] and the UCSC tool In-Silico PCR [30]. First- and second-round PCR were performed using PrimeSTAR GXL Polymerase (TaKaRa Bio, Shiga, Japan). The polymerase was tested and confirmed in our precedent work. Second round PCR products were purified using 0.45X AMPure PB beads (Pacific Biosciences, Menlo Park, CA, USA) and DNA concentrations were read on a Qubit 2.0 Fluorometer using the Qubit dsDNA Broad Range Assay Kit (Invitrogen Life Technologies, Carlsbad, CA, USA).

For an amplicon size ranging from 1 to 3 kb, the input DNA amounts per pool should be between 500–1000 ng. Depending on the recorded concentrations, we proceeded by calculating the necessary microliters of each sample so as to obtain a total amount of 1000 ng. The pooled PCR amplicons were then concentrated using AMPure PB Beads. The next step consisted in the construction of the SMRTbell library, which involves DNA damage repair, End-repair/A-tailing, and adapter ligation. All the steps are described in the PacBio procedure with specific reagent and amounts [26]. The SMRTbell Templates were then purified again with AMPure PB before sequencing.

4.4. Bioinformatics and Genetic Variants Classification

The pathogenicity of all identified variants was evaluated according to the American College of Medical Genetics and Genomics guidelines (ACMG) [31], with the help of Var-Some [32], dbSNP [24], ClinVar [25], and gnomAD [33] databases, thanks to an in-house bioinformatics pipeline [34]. Fastq (forward-reverse) files were obtained after sequencing. Bioinformatic analysis was performed as previously described [35,36]. Briefly, the sequencing reads were mapped to the reference genome (hg38/GRCh38) using Burrow-Wheeler Aligner (version 0.7.17-r1188) software. Duplicates were removed using SAMBAMBA (version 0.6.7) and MarkDuplicates GATK (version 4.0.0.0). The BAM alignment files generated were refined by local realignment and base quality score recalibration, using the RealignerTargetCreator and IndelRealigner GATK tools. Minor allele frequencies (MAF)

were retrieved from the Genome Aggregation Database [33]. Long-read sequencing data were analyzed using an in-house bioinformatics pipeline [37].

4.5. Molecular Modeling of TTLL5-RPGR^{ORF15} Interactions

The model of the TTLL5 core domain was obtained by homology modeling to the TTLL6 structure reported in PDB entry 6VZW, where an initiation analog is bound to the active site of TTLL6. The analog mimics a di-Glu peptide where the donor glutamate is linked to γ-carboxylic acid of acceptor glutamate through a phosphinate (donor and acceptor refer to the free glutamate to be linked and the receiving glutamate, respectively). The main chain residues flanking the di-Glu are replaced by ethylamine at C-term and acetate at N-term. The intermediate's structure and binding geometry were then retained in the modeled TTLL5 and exploited to generate the enzyme-bound form of RPGR^{ORF15}. The di-Glu was split by removing the phosphinate and was transformed into the donor and the acceptor glutamate. The latter was interpreted as Glu737 in RPGR^{ORF15}, and after ethylamine and acetate capping removal, the Glu737 was expanded at each terminus to reproduce the sequence of the deleted peptide plus the flanking amino acids to the extent necessary to cover the active site crevice, leading to the final 12-mer GEEEHGE737GEEEE peptide. The structure of the expanded peptide was docked to the protein structure by holding Glu737 fixed, acting as a covalent anchor to the upstream and downstream segments. The process was repeated for two other peptides, namely EEGGEEE734GDREE and EGEGEEE870GEEGE.

Docking was performed using Autodock Vina 1.2 [38] and MAGI-Dock [39], a PyMol plugin that generates the docking boxes [37] around the active site crevice. Two separate docking runs were performed, one for each of the upstream and downstream segments of Glu737.

Finally, the resulting conformation was energetically refined by molecular dynamics simulations. We used CHARMM-GUI [40] to generate the solvated system and GRO-MACS [41] to run the simulation. The segment's flanking residues were neutralized to prevent unwanted interaction, and the resulting docked structure was used as the simulation starting structure. The protein complex was placed in a triclinic box with a minimum 1.2 nm spacing on each side, and the system was solvated using TIP3P water molecules and neutralized with K^+/Cl^-. Sequential minimization and position-restrained equilibrations in the NVT and NPT ensemble were performed before a 200 ns long production run. Root Mean Squared Deviation (RMSD) and trajectory clustering were analyzed using the GROMACS tools gmx rms and gmx cluster, respectively.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms242316881/s1.

Author Contributions: Conceptualization: A.B. and G.M.; Data curation: G.B., W.C., C.M. and F.C.; Validation: G.B., W.C., A.B., J.K., F.C. and G.M.; Formal Analysis: G.B., W.C., A.B., J.K., F.C. and G.M.; Funding acquisition: G.M. and M.B. (Matteo Bertelli); Investigation: W.C., A.B., J.K., C.M., F.C., G.M., M.C.M., L.C., S.M., G.S., A.P.S., B.F., G.P. (Giorgio Placidi), M.A., G.P. (Grazia Pertile), M.B. (Mario Bengala), F.B., A.P., F.D., B.T., P.L., F.R. and F.V.; Methodology: J.K., A.B., G.M. and M.B.; Project administration: G.M. and M.B. (Matteo Bertelli); Resources: M.C.M., L.C., S.M., G.S., A.P.S., B.F., G.P. (Giorgio Placidi), M.A., G.P. (Grazia Pertile), M.B. (Mario Bengala), F.B., A.P., F.D., B.T., P.L., F.R. and F.V.; Software: W.C., J.K., A.B. and G.M.; Supervision: A.B. and G.M.; Validation: W.C., C.M., F.C. and G.M.; Visualization: G.B., W.C., J.K. and A.B.; Writing—original draft: G.B., W.C., A.B. and J.K.; Writing—review & editing: L.C., S.M., G.S., A.P.S., B.F., G.P. (Giorgio Placidi), M.A., G.P. (Grazia Pertile), M.B. (Mario Bengala), F.B., A.P., F.D., B.T., P.L., F.R., F.V., G.M. and M.B. (Matteo Bertelli). All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethical Committee of Azienda Sanitaria dell'Alto Adige, Italy (Approval No. 132-2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study, which includes consent to use the anonymized genetic results for research.

Data Availability Statement: All data are contained in the manuscript and in the Supplementary Materials.

Acknowledgments: We would like to thank all the patients and the clinicians for their kind collaboration.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hu, T.; Chitnis, N.; Monos, D.; Dinh, A. Next-generation sequencing technologies: An overview. *Hum. Immunol.* **2021**, *82*, 801–811. [CrossRef]
2. van Dijk, E.L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. The Third Revolution in Sequencing Technology. *Trends Genet.* **2018**, *34*, 666–681. [CrossRef]
3. Logsdon, G.A.; Vollger, M.R.; Eichler, E.E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **2020**, *21*, 597–614. [CrossRef] [PubMed]
4. Nanopore Sequencing Accuracy. Available online: https://nanoporetech.com/accuracy (accessed on 21 March 2023).
5. Nguyen, X.-T.-A.; Talib, M.; van Schooneveld, M.J.; Brinks, J.; Ten Brink, J.; Florijn, R.J.; Wijnholds, J.; Verdijk, R.M.; Bergen, A.A.; Boon, C.J. RPGR-Associated Dystrophies: Clinical, Genetic, and Histopathological Features. *Int. J. Mol. Sci.* **2020**, *21*, 835. [CrossRef] [PubMed]
6. Nassisi, M.; De Bartolo, G.; Mohand-Said, S.; Condroyer, C.; Antonio, A.; Lancelot, M.E.; Bujakowska, K.; Smirnov, V.; Pugliese, T.; Neidhardt, J.; 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. [CrossRef] [PubMed]
7. Sun, X.; Park, J.H.; Gumerson, J.; Wu, Z.; Swaroop, A.; Qian, H.; Roll-Mecak, A.; Li, T. Loss of RPGR glutamylation underlies the pathogenic mechanism of retinal dystrophy caused by *TTLL5* mutations. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2925–E2934. [CrossRef] [PubMed]
8. Chiang, J.P.W.; Lamey, T.M.; Wang, N.K.; Duan, J.; Zhou, W.; McLaren, T.L.; Thompson, J.A.; Ruddle, J.; De Roach, J.N. Development of High-Throughput Clinical Testing of RPGR ORF15 Using a Large Inherited Retinal Dystrophy Cohort. *Investig. Opthalmology Vis. Sci.* **2018**, *59*, 4434. [CrossRef] [PubMed]
9. Jin, Z.-B.; Hayakawa, M.; Murakami, A.; Nao-i, N. RCC1-Like Domain and ORF15: Essentials in RPGR Gene. In *Retinal Degenerative Diseases*; Springer: Boston, MA, USA, 2006; pp. 29–33. [CrossRef]
10. 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. [CrossRef]
11. The Human Gene Mutation Database. Available online: https://www.hgmd.cf.ac.uk/ac/index.php (accessed on 14 April 2023).
12. Ensembl. Available online: https://www.ensembl.org/index.html (accessed on 14 April 2023).
13. Yokoyama, A.; Maruiwa, F.; Hayakawa, M.; Kanai, A.; Vervoort, R.; Wright, A.F.; Yamada, K.; Niikawa, N.; Naōi, N. Three novel mutations of the *RPGR* gene exon ORF15 in three Japanese families with X-linked retinitis pigmentosa. *Am. J. Med. Genet.* **2001**, *104*, 232–238. [CrossRef]
14. Oh, J.K.; Del Valle, J.G.V.; de Carvalho, J.R.L.; Sun, Y.J.; Levi, S.R.; Ryu, J.; Yang, J.; Nagasaki, T.; Emanuelli, A.; Rasool, N.; et al. Expanding the phenotype of TTLL5-associated retinal dystrophy: A case series. *Orphanet J. Rare Dis.* **2022**, *17*, 146. [CrossRef]
15. Smirnov, V.; Grunewald, O.; Muller, J.; Zeitz, C.; Obermaier, C.D.; Devos, A.; Pelletier, V.; Bocquet, B.; Andrieu, C.; Bacquet, J.-L.; et al. Novel TTLL5 Variants Associated with Cone-Rod Dystrophy and Early-Onset Severe Retinal Dystrophy. *Int. J. Mol. Sci.* **2021**, *22*, 6410. [CrossRef]
16. Del Pozo-Valero, M.; Riveiro-Alvarez, R.; Martin-Merida, I.; Blanco-Kelly, F.; Swafiri, S.; Lorda-Sanchez, I.; Trujillo-Tiebas, M.J.; Carreño, E.; Jimenez-Rolando, B.; Garcia-Sandoval, B.; et al. Impact of Next Generation Sequencing in Unraveling the Genetics of 1036 Spanish Families with Inherited Macular Dystrophies. *Investig. Ophthalmol Vis. Sci* **2022**, *63*, 11. [CrossRef]
17. Mahalingan, K.K.; Keenan, E.K.; Strickland, M.; Li, Y.; Liu, Y.; Ball, H.L.; Tanner, M.E.; Tjandra, N.; Roll-Mecak, A. Structural basis for polyglutamate chain initiation and elongation by TTLL family enzymes. *Nat. Struct. Mol. Biol.* **2020**, *27*, 802–813. [CrossRef]
18. Megaw, R.D.; Soares, D.C.; Wright, A.F. RPGR: Its role in photoreceptor physiology, human disease, and future therapies. *Exp. Eye Res.* **2015**, *138*, 32–41. [CrossRef] [PubMed]
19. Hussain, H.B.; Wilson, K.A.; Wetmore, S.D. Serine and Cysteine π-Interactions in Nature: A Comparison of the Frequency, Structure, and Stability of Contacts Involving Oxygen and Sulfur. *Aust. J. Chem.* **2015**, *68*, 385. [CrossRef]
20. Natarajan, K.; Gadadhar, S.; Souphron, J.; Magiera, M.M.; Janke, C. Molecular interactions between tubulin tails and glutamylases reveal determinants of glutamylation patterns. *EMBO Rep.* **2017**, *18*, 1013–1026. [CrossRef]
21. Colombo, L.; Maltese, P.E.; Castori, M.; El Shamieh, S.; Zeitz, C.; Audo, I.; Zulian, A.; Marinelli, C.; Benedetti, S.; Costantini, A.; et al. Molecular Epidemiology in 591 Italian Probands with Nonsyndromic Retinitis Pigmentosa and Usher Syndrome. *Investig. Opthalmology Vis. Sci.* **2021**, *62*, 13. [CrossRef] [PubMed]

22. Wu, J.; Xie, D.; Wang, L.; Kuang, Y.; Luo, S.; Ren, L.; Li, D.; Mao, A.; Li, J.; Chen, L.; et al. Application of third-generation sequencing for genetic testing of thalassemia in Guizhou Province, Southwest China. *Hematology* **2022**, *27*, 1305–1311. [CrossRef]
23. Hassan, S.; Bahar, R.; Johan, M.F.; Hashim, E.K.M.; Abdullah, W.Z.; Esa, E.; Hamid, F.S.A.; Zulkafli, Z. Next-Generation Sequencing (NGS) and Third-Generation Sequencing (TGS) for the Diagnosis of Thalassemia. *Diagnostics* **2023**, *13*, 373. [CrossRef] [PubMed]
24. dbSNP. Available online: https://www.ncbi.nlm.nih.gov/snp/ (accessed on 21 March 2023).
25. ClinVar. Available online: https://www.ncbi.nlm.nih.gov/clinvar/ (accessed on 21 March 2023).
26. Procedure & Checklist—Preparing SMRTbell®Libraries using PacBio®Barcoded Universal Primers for Multiplexing Amplicons. Available online: https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-SMRTbell-Libraries-using-PacBio-Barcoded-Universal-Primers-for-Multiplexing-Amplicons.pdf (accessed on 21 March 2023).
27. Mareso, C.; Albion, E.; Cozza, W.; Tanzi, B.; Cecchin, S.; Gisondi, P.; Michelini, S.; Bellinato, F.; Michelini, S.; Michelini, S.; et al. Optimization of long-range PCR protocol to prepare filaggrin exon 3 libraries for PacBio long-read sequencing. *Mol. Biol. Rep.* **2023**, *50*, 3119–3127. [CrossRef]
28. Primer3web. Available online: https://primer3.ut.ee (accessed on 21 March 2023).
29. Primer-BLAST. Available online: https://www.ncbi.nlm.nih.gov/tools/primer-blast (accessed on 21 March 2023).
30. UCSC In-Silico PCR. Available online: https://genome.ucsc.edu/cgi-bin/hgPcr (accessed on 21 March 2023).
31. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–424. [CrossRef]
32. VarSome. Available online: https://varsome.com/ (accessed on 21 March 2023).
33. GnomAD. Available online: https://gnomad.broadinstitute.org/ (accessed on 21 March 2023).
34. Cristofoli, F.; Sorrentino, E.; Guerri, G.; Miotto, R.; Romanelli, R.; Zulian, A.; Cecchin, S.; Paolacci, S.; Miertus, J.; Bertelli, M.; et al. Variant Selection and Interpretation: An Example of Modified VarSome Classifier of ACMG Guidelines in the Diagnostic Setting. *Genes* **2021**, *12*, 1885. [CrossRef] [PubMed]
35. Mattassi, R.; Manara, E.; Colombo, P.G.; Manara, S.; Porcella, A.; Bruno, G.; Bruson, A.; Bertelli, M. Variant Discovery in Patients with Mendelian Vascular Anomalies by Next-Generation Sequencing and Their Use in Patient Clinical Management. *J. Vasc. Surg.* **2018**, *67*, 922–932.e11. [CrossRef] [PubMed]
36. Marceddu, G.; Dallavilla, T.; Guerri, G.; Zulian, A.; Marinelli, C.; Bertelli, M. Analysis of Machine Learning Algorithms as Integrative Tools for Validation of next Generation Sequencing Data. *Eur. Rev. Med. Pharmacol. Sci.* **2019**, *23*, 8139–8147. [CrossRef] [PubMed]
37. Sorrentino, E.; Albion, E.; Modena, C.; Daja, M.; Cecchin, S.; Paolacci, S.; Miertus, J.; Bertelli, M.; Maltese, P.E.; Chiurazzi, P.; et al. PacMAGI: A pipeline including accurate indel detection for the analysis of PacBio sequencing data applied to RPE65. *Gene* **2022**, *832*, 146554. [CrossRef]
38. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898. [CrossRef]
39. MAGI-Dock. Available online: https://github.com/gjonwick/MAGI-Dock (accessed on 29 May 2023).
40. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865. [CrossRef]
41. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. [CrossRef]