Contribution of non-coding mutations to *RPGRIP1*-mediated inherited retinal degeneration

Running Title: Non-coding *RPGRIP1* mutations

Farzad Jamshidi M.D., Ph.D.[1], Emily M. Place M.S.[1], Sudeep Mehrotra M.S.[1], Daniel Navarro-Gomez B.Sc.[1], Mathew Maher M.S.[1], Elise Valkanas B.Sc.[2], Timothy J. Cherry Ph.D.[3], Monkol Lek Ph.D.[2,4], Daniel MacArthur Ph.D.[2,4], Kinga M. Bujakowska Ph.D.[1], Eric A. Pierce M.D., Ph.D.[1]

[1]Ocular Genomics Institute, Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA.

[2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA.

[3] Center for Developmental Biology and Regenerative Medicine, Seattle Children's Hospital, University of Washington, Seattle, WA.

[4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA.

Correspondence:

Eric A. Pierce M.D., Ph.D.

Solman and Libe Friedman Professor of Ophthalmology

Massachusetts Eye and Ear Infirmary, Harvard Medical School

243 Charles Street, Boston, MA 02114

Phone: 617-573-6917, Email: eric_pierce@meei.harvard.edu

Conflict of interest notification page

The authors declare no conflict of interest related to the work presented in this manuscript.

The grants funding this work are listed in the acknowledgements section.

Abstract

Purpose: With the advent of gene therapies for inherited retinal degenerations (IRDs), genetic diagnostics will have an increasing role in clinical decision-making. Yet a significant portion of IRDs remains unsolved. We hypothesized that non-coding mutation solve a significant portion of such cases. Focusing on *RPGRIP1*, we examined the significance of non-coding mutations.

Methods: A cohort of IRD families underwent targeted panel sequencing. Unsolved cases were explored by whole exome and whole genome sequencing to look for copy number and intronic mutations. The candidates were then validated by Sanger sequencing, quantitative PCR, and *in vitro* splicing assays in two cell lines analyzed through amplicon sequencing.

Results: Among 1722 families, three had biallelic loss of function mutations in *RPGRIP1* and were considered solved while seven had a single disruptive coding mutation. Whole exome and whole genome sequencing revealed non-coding candidates in these seven families. In six, the non-coding mutations were shown to lead to loss of function *in vitro*. Additionally, an unannotated exon of *RPGRIP1* was discovered and the efficacy of amplicon sequencing in assessing splicing mutations highlighted.

Conclusion: *RPGRIP1*–mediated IRDs are more common than previously thought. Addition of copy number and 30 base pair flanking intronic sequences will increase the sensitivity of diagnostic panels.

Inherited retinal degeneration, non-coding mutations, *RPGRIP1*, intronic mutations, whole-genome sequencing

INTRODUCTION

Inherited retinal degenerations (IRDs) are a group of predominantly monogenic diseases that are the most common cause of blindness in the working age population[1]. About 260 genes have been associated with IRDs with functions spanning almost every aspect of cellular function: from splicing machinery, to microtubular transport and phototransduction[1]. State-of-the-art clinical diagnostics, using next-generation sequencing (NGS) based targeted panel sequencing of known IRD genes, successfully identifies the causal mutation in only 50 to 70% of cases[2,3]. Although copy number changes[4] and intronic mutations[5] contribute to disease, they are not routinely assessed and likely comprise a significant portion of unsolved cases. With the advent of successful gene therapies for IRDs[6], understanding such non-coding mutations and developing assays to evaluate them is of increasing importance. One example is the autosomal recessive *RPGRIP1*-associated disease which is an attractive candidate for gene therapy with already established success in murine[7] and canine[8] models. Yet, almost all of the mutations in *RPGRIP1* have been described in the coding-region[9].

RPGRIP1 plays a critical role in opsin trafficking, outer-segment disc organization and photoreceptor survival[10,11]. While it primarily localizes to the transition zone of rods and cones, various of its isoforms can be found in the outer segment, along the microtubules as well as in the amacrine cells of the inner plexiform layer[12,13]. Its largest transcript variant, NM_020366, is composed of 3861 coding base pairs distributed over 24 exons[14]. This encodes a 1287 amino acid protein that interacts with a variety of other IRD proteins such as Retinitis Pigmentosa GTPase Regulator (RPGR), SPATA7 and NPHP4[15]. The expression of *RPGRIP1* is limited to the retina and testis[16].

Confident genetic diagnosis with *RPGRIP1* as the causal gene will be crucial for

effective clinical trials. In our analysis of IRD families with targeted panel sequencing of coding regions of IRD-associated genes[17], we repeatedly noted identification of singular likely pathogenic variants in *RPGRIP1*. To evaluate these families further, we performed whole exome and genome sequencing to look into the contribution of non-coding mutations which led to identification of a novel un-annotated exon and the development of a new approach to assess splicing.

MATERIALS AND METHODS

Ethical guidelines

The study was approved by the institutional review board at the Massachusetts Eye and Ear (Human Studies Committee MEE in USA) and adhered to the Declaration of Helsinki. Informed consent was obtained from all individuals on whom genetic testing and further molecular evaluations were performed.

Clinical evaluation

All the patients in this study underwent expert clinical assessment by an ophthalmologist sub-specializing in inherited retinal degenerations. The clinical characteristics are outlined in Table 1.

Sequencing

DNA was extracted from venous blood using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). All samples underwent GEDi sequencing as described previously[17]. Whole exome and genome sequencing were done at the Center for Mendelian Genomics at the Broad Institute of MIT and Harvard using methodology described previously[18]. Sanger sequencing was performed on ABI 3730xl using BigDye Terminator v3.1 kits (Life Technologies, Carlsbad, CA) and using PCR primers indicated in Supplementary. When PCR products were sequenced, they were cleaned prior sequencing (ExoSap-IT, Affymetrix, Santa Clara, CA). Gel bands that were

Sanger sequenced had DNA extracted via the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, CA).

Bioinformatics

Analysis was done as described previously[17,19]. BWA was used for alignment. SAMtools and custom programs were used for single nucleotide polymorphism and insertion/deletion calls[19]. Variants of interest were limited to polymorphisms with less than 0.005 allelic frequency in the gnomAD (http://gnomad.broadinstitute.org/) and ExAC (http://exac.broadinstitute.org/) databases[18]. Whole genome copy number analysis, with consideration of structural changes, was done via Genome STRiP 2.0[20]. For the analysis of splicing patterns from amplicon-sequencing, STAR (version 2.5.3a) aligner[21] was used to generate an index of the human genome (GRCh37.75.dna.primary_assembly.fa) and to align the reads. IGV[22] was used to load the aligned sequences (BAM files) and for data visualization with Sashimi plots.

Polymerase chain reactions (PCR), cloning and site-directed mutagenesis

PCR was performed using PfuUltra II Fusion polymerase (Agilent Technologies, Santa Clara, CA) on genomic DNA of patients harboring the mutations of interest (primers are listed in the Supplementary). The PCR products were cloned into pENTR Directional TOPO vector (Thermo Fisher, Waltham, MA) and used to transform chemically competent *Escherichia coli* (One Shot TOP10, Thermo Fisher, Waltham, MA). Plasmid DNA from single colonies was extracted with miniprep kits (ZymoPURE, Zymo Research) and analyzed by restriction enzyme digestion with BsrGI (NE Biolabs, Ipswich, MA) and Sanger sequencing. Essential splice-site mutations were introduced by site-directed mutagenesis (QuickChange II Site Directed mutagenesis kit, Agilent Technologies) and verified by Sanger sequencing. Colonies with the correct sequence and restriction enzyme pattern were then sub-cloned into the pCS2+GW vector (kind gift from Dr

Erica Davis) (Figure S7) via Gateway LR clonase II (Thermo Fisher) and similar analyses as before was done to isolate vectors with the appropriate inserts for transfection experiments.

Quantitative polymerase chain reactions (qPCR)

Five nanograms (ng) of genomic DNA, 200 nM of each primer and 10 µl of Fast SYBR Green Master Mix (Life Technologies, Grand Island, NY) were used for qPCR reactions which was performed on Stratagene Mx3000P (Agilent Technologies) using the standard thermocycling program (95 °C for 3 min, 40 cycles of 95 °C for 20 s, and 60 °C for 1 min, followed by a melting curve). The ddCT method was used for the analysis of results where *ZNF80* was used as a reference gene and an in-house DNA sample with wild type *RPGRIP1* (OGI200) used for normalization. Each sample was done in triplicate and the average value was used. Standard deviation with error propagation was used to calculate up and down errors.

Cell culture and transfections

Human embryonic kidney (HEK293T) and retinoblastoma (WERI-Rb1) cells purchased from American Type Culture Collection (ATCC, Manassas, VA) and maintained in RPMI medium supplemented with 10% Fetal Bovine Serum-1640 (Thermo Fisher). 2ml of 5 x 10^5 cells/ml were plated into each well of a 6-well plate (Corning Inc., Corning, NY) 12hrs prior to transfections. 1-5 µg of vector DNA per well was used for transfections using a commercial reagent (Lipofectamine 2000, Invitrogen, Carlsbad, CA). Cells were harvested for RNA extraction 48 hours after transfection.

RNA isolation and cDNA synthesis

Cells were lysed with TRIzol (Thermo Fisher). After BCP or chloroform (Sigma-Aldrich, St. Louis, MO) treatment, the aqueous phase was transferred to mRNeasy columns with DNase I digestion performed on-column (Qiagen). Quantification was performed via NanoDrop (Thermo

Fisher) and 500ng of RNA was converted to cDNA using oligo(dT) primers and SuperScript II (Thermo Fisher).

Splicing assay and amplicon sequencing

The mutant, control and wild-type vectors described above were transfected in to HEK293T and WERI-Rb1 cells. Two days post-transfection, cDNA was generated as described and RT-PCR performed amplifying the flanking exons of the point mutation of interest. The PCR products were cleaned by DNA-clean and concentrator kits (Zymo Research). Amplicon sequencing was then performed at the Massachusetts General Hospital Center for Computational and Integrative Biology, where the PCR products were fragmented using sonication and sequenced with the standard NGS pipeline. Visualization and analysis of the data were performed as described under bioinformatics.

RESULTS

Genetic analysis of 1722 IRD probands that underwent targeted exon sequencing of known IRD genes[17] revealed three patients with bi-allelic loss-of-function (LoF) mutations in *RPGRIP1* and seven with only one LoF change in this gene (Table 1, Figure S1). In the latter seven families no other significant mutations in *RPGRIP1* or other IRD genes were identified. Since they were all diagnosed with an early onset IRD, a characteristic presentation of *RPGRIP1* disease[23], further testing was performed to find non-coding variants and copy number changes in *RPGRIP1* or other IRD genes through whole exome and whole genome sequence analysis. Among these seven families, three of the second-allele mutations were predicted to be copy number changes and three were intronic mutations (Table 1). In one family, OGI-578, we did not find likely pathogenic variants in *RPGRIP1* (Figure S1).

OGI-1797 had a predicted deletion of exon 19 (Table 1), which was confirmed by qPCR

(Figure S2). Exon 19 has 139 nucleotides, and thus its deletion leads to a frameshift resulting in a premature stop codon and subsequent nonsense mediated decay (NMD)[24].

In OGI-281, the gain of a copy of exon 2 was detected through copy number analysis of whole genome sequencing which included structural analysis[20] in addition to coverage-based predictions. The structural change led to misalignment of paired-end reads (Figure 1A and 1C), which indicated a tandem duplication. This mutation was validated by qPCR (Figure 1D), and the predicted breakpoint confirmed through PCR and Sanger sequencing (Figure 1E and 1F). Sanger sequencing also identified 135bp of missing DNA upstream of the breakpoint, suggesting a possible complex rearrangement as the causal event[25]. Given that exon 2 is 133bp, its tandem duplication would lead to out-of-frame translation, an early stop codon and NMD.

OGI-237 similarly had a tandem copy number gain but in both exons 1 and 2 detected by whole exome sequencing (Figure 2B). The 5' breakpoint was mapped 2Kb upstream of exon 1 (Figure 2C and S3). However, given that the second copy would maintain 2Kb upstream of the exon 1, which would include a proximal presumed promoter (Figure 2C), we questioned whether a second transcriptional start site (TSS) would exclude the mutant exons 1' and 2' thus leading to normal transcription. Therefore, we hypothesized that perhaps a critical *RPGRIP1* regulatory domain exists outside of 2Kb region upstream of exon 1. Reviewing transcriptome data of normal human retina[26], we indeed identified an additional exon upstream of the annotated transcript, which was revealed by numerous split reads as well as high sequence coverage approximately 8Kb upstream of *RPGRIP1* exon 1 (Figure 2D). We further assessed this using ATAC-seq and ChIP-seq data, which suggested open chromatin and transcription factor binding to this region and not to the previously annotated *RPGRIP1* transcription start site (TSS) (Figure 2E). These results suggest that the retinal TSS of *RPGRIP1* and its regulatory regions are 8Kb

upstream of the currently annotated exon 1. This also explains the likely pathogenic effect of the mutation seen in OGI-237. The tandem exon 1 and 2 duplication excludes a TSS and will lead to a premature stop-codon 29 codons past the start of the duplicated 5'UTR of exon 1.

Aside from the copy number changes, three families had deep intronic mutations that segregated in the families and had low frequencies in the gnomAD database[18] (Table 1). We hypothesized that these mutations could potentially lead to aberrant splicing leading to frameshift, pre-mature stop codons and NMD as well. To assess the effect of these changes on splicing, midi-gene assays were performed in HEK293T cells with mutant and wild-type (WT) constructs, as well as essential splice-site changes serving as positive controls. In OGI-827, the c. 2367+23delG change led to a significant retention of intron 15 in mutant vs. wildtype comparable to the findings in the c. 2367+1G>A positive control (Figure 3A). While the ratio of splicing between exons 15 to 16 to that between exons 14 and 15 was 6090/6207 or 0.98 in wild-type, this proportion was altered to 2209/3982 or 0.55 in mutant. This suggests a relative reduction of 43% of the appropriate splicing at this locus. The aberrant splice isoform leads to a premature stop-codon 19 codons into intron 15.

In OGI-601, the c.1611+27G>A mutation led to a splicing pattern that became apparent only through amplicon sequencing (Figure 3B and 5S). In the WT allele, three splicing signals of various strengths were detected between exons 12 and 13 (Figure 3B). The two prominent splice variants have been annotated for *RPGRIP1*, the main event (9598 reads) being present in the NM_020366 transcript and indicating the canonical splice junction between exon 12 and 13. The second splice donor occurring 104bp into intron 12 (8499 reads) is part of 5'UTR predicted transcript model (XM_005267881.3, Figure 3B). The third splice donor site occurs 227bp into intron 12 and has not been reported. In both the positive control, c.1611+1G>A, and the

mutation under study, c.1611+27G>A, the primary splicing event that would lead to full-length protein failed to occur whereas the alternate transcript donor site, 104 past the NM_020366 exon 12, was preferred (Figure 3B). This variant would lead to a pre-mature stop codon 12 codons after the exon 12 of the full-length transcript (i.e. NM_020366).

In OGI-949, the c.1468-263G>C mutation resulted in an activation of a cryptic splice acceptor site 13bp past the mutation (Figure 3C). The resultant cryptic exon 11', although 117bp and in-frame, has a premature stop codon four amino acids past the cryptic acceptor site. Thus the aberrant transcript will lead to NMD.

In order to account for potential differences in splicing between the neuronal and non-neuronal cell types[27], all the splicing experiments were repeated with a retinoblastoma cell line WERI-Rb1, derived from photoreceptor cells[28]. The splicing changes resulting from the mutations assessed, led to similar patterns in both HEK293T as well as WERI-Rb1 lines (Figure S6). Overall, we found six novel non-coding mutations, six unreported coding mutations in *RPGRIP1* (Figure 4) and we have corrected the transcript model for *RPGRIP1* in the retina with a new TSS and 5' exon 8kb upstream of the annotated exon 1.

DISCUSSION

In our analysis of nine families with biallelic *RPGRIP1* mutations, six had a second non-coding mutation leading to a LoF allele. This not only highlights the importance of non-coding mutations in pathogenesis of recessive IRDs, but also implies a greater prevalence of *RPGRIP1* mediated disease than previously thought. Three among the 1722 IRD families had bilallelic coding mutations in *RPGRIP1*, which corresponds roughly to the previously reported rates in the literature (0.17% vs 0.25%[14,29-31]). However, when taking into account non-coding mutations, our rate increased from 0.17% to 0.51%. This significant increase in frequency of *RPGRIP1*

mediated disease is still likely an underestimate given the bias of choosing samples with one LoF coding sequence mutation. Considering the possibility of bialleleic non-coding LoF alleles, the rate of *RPGRIP1* mediated disease is likely to be much greater than what was previously thought.

In the six families solved with the addition of structural and intronic mutation analyses, we validated the second LoF mutations *in vitro* showing aberrations in the reading-frame caused by splicing or copy number changes. While deletions are simpler to interpret given clear interruption of reading frame, copy number gains can be just as disruptive. In both cases of copy number gain, OGI-237 and OGI-281, the mutations were tandem duplications leading to premature stop codons. While investigating the potentially pathogenic effect of the duplication in the first two annotated *RPGRIP1* exons in OGI-237, we discovered a novel exon and a retinal TSS 8Kb upstream of the previously annotated transcriptional start site. This is important as mutations in regulatory regions and 5'UTR have been shown to be disease-causing, such as in the case of *PRPF31* and *OPN1MW*[9]. We propose the inclusion of this *RPGRIP1* exon, chr14:21748265-21748318 (hg19), in future IRD diagnostic panels. On a separate note, knowledge of the true regulatory regions of *RPGRIP1* can be crucial in gene therapy models where biologically-relevant expression rather than over-expression is the therapeutic goal.

We also detected three mutations causing intron retentions or cryptic exons, two of which were in the flanking 30bp of annotated exons. The traditional way of assessing for the effect of intronic mutations includes use of exon-trapping vectors with a small (mini-gene) or medium (midi) sized insert harboring the mutation under study[32,33]. Subsequent transfection into HEK293T cells, and analysis of the processed RNA via RT-PCR allows one to assess the alterations via gel electrophoresis. However, as seen in the case of OGI-601 (Figure 3B), there

can be a multitude of splicing signals, some of which are not annotated for the genes of interest thus making the interpretation of the electrophoresis results difficult. We found that amplicon sequencing of the RT-PCR product can increase the sensitivity of detection (OGI-601 Figure 3B and S5) and also add a quantitative value in assessment of splicing changes.

Splicing is a tissue-specific phenomenon and neuronal tissues have a unique splicing machinery that can lead to unique exonic retentions[27]. Thus the splicing patterns we identified in HEK293T cells were confirmed in a retina relevant line. We chose WERI-Rb1, a retinoblastoma line shown to possess the neuronal specific splicing machinery[27]. Although we did not detect a significant difference between the two lines (Figure S6), we foresee a benefit in using a tissue relevant cell line, such as WERI-Rb1 for IRDs, when assessing splicing mutations.

Despite the decreasing costs of whole genome sequencing, it is still not feasible for routine clinical assessment. Hence, assessing for non-coding mutations in *RPGRIP1* and other recessive IRD genes still demands a step-wise approach to reduce screening costs. Our recommendation is to include and analyze at least 30bp of the flanking exonic regions and to include copy number analysis[4] in targeted panel sequencing of recessive IRD genes. In this study, we noted only one significant mutation outside of these criteria (c.1468-263G>C). The remaining unsolved cases can then be submitted for whole genome sequencing to identify novel genes and mutations.

As demonstrated above, the empirical validation of second-mutant alleles in recessive diseases when one mutant allele is identified can be time consuming and expensive. One might argue that such stringency may not be necessary if the gene identified with a single mutation matches the presenting phenotype. Statistical models predict a small false positive diagnosis rate when a single mutation is identified in a disease-specific gene such as *MYO7A* in Usher Syndrome Type I[34]. Yet when a large number of potential genes can lead to the phenotype of

interest, such as in early-onset IRD, the calculations will not yield as small a false positive rate[34]. Thus we strongly believe that in order to make the correct diagnosis, it is important to identify and characterize both mutant alleles.

Accurate genetic diagnostics for inherited retinal degenerations is increasingly critical with the emergence of gene therapy. The results of the first FDA approved human gene therapy randomized clinical trial, for the treatment of REP65-mediated IRDs, has been published[6]. Furthermore, more than 10 clinical trials for retinal degeneration have either been completed or initiated[35-40]. A multitude of additional candidates for gene therapy with already established animal models exist and *RPGRIP1* is one such example. Evidence of biallelic gene mutations are critical for inclusion of patients in trials when dealing with autosomal recessive diseases. In this paper we highlighted the importance of including non-coding mutations to establish the final diagnosis but also to increase the diagnostic sensitivity.

Acknowledgments

This work was supported by grants from the National Eye Institute [RO1EY012910 (EAP), R01EY026904 (KMB/EAP) and P30EY014104 (MEEI core support)], and the Foundation Fighting Blindness (USA, EAP). Sequencing and analysis was provided by the Center for Mendelian Genomics at the Broad Institute of MIT and Harvard and was funded by the National Human Genome Research Institute, the National Eye Institute, and the National Heart, Lung and Blood Institute grant UM1 HG008900 to Daniel MacArthur and Heidi Rehm. The authors would like to thank the patients and their family members for their participation in this study and the Ocular Genomics Core members for their experimental assistance. The authors would like to thank the Exome Aggregation Consortium, the Genome Aggregation Database (gnomAD) and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at http://exac.broadinstitute.org/about and http://gnomad.broadinstitute.org/about. The authors declare no conflict of interest related to the work presented in this manuscript.

References

1. Farrar GJ, Carrigan M, Dockery A, et al. Toward an elucidation of the molecular genetics of inherited retinal degenerations. *Hum Mol Genet.* Aug 1 2017;26(R1):R2-R11.
2. Ellingford JM, Barton S, Bhaskar S, et al. Molecular findings from 537 individuals with inherited retinal disease. *J Med Genet.* May 11 2016.
3. Riera M, Navarro R, Ruiz-Nogales S, et al. Whole exome sequencing using Ion Proton system enables reliable genetic diagnosis of inherited retinal dystrophies. *Sci Rep.* Feb 9 2017;7:42078.
4. Bujakowska KM, Fernandez-Godino R, Place E, et al. Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. *Genetics in medicine : official journal of the American College of Medical Genetics.* Jun 2017;19(6):643-651.
5. Liquori A, Vache C, Baux D, et al. Whole USH2A Gene Sequencing Identifies Several New Deep Intronic Mutations. *Human mutation.* Feb 2016;37(2):184-193.
6. Russell S, Bennett J, Wellman JA, et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *The Lancet.* 2017;390(10097):849-860.
7. Pawlyk BS, Bulgakov OV, Liu X, et al. Replacement gene therapy with a human RPGRIP1 sequence slows photoreceptor degeneration in a murine model of Leber congenital amaurosis. *Human gene therapy.* 2010;21(8):993-1004.
8. Lhériteau E, Petit L, Weber M, et al. Successful gene therapy in the RPGRIP1-deficient dog: a large model of cone–rod dystrophy. *Molecular therapy.* 2014;22(2):265-277.
9. Human Gene Mutation Database (HGMD), Institute of Medical Genetics, Cardiff University, United Kingdom, 2017. http://www.hgmd.cf.ac.uk/.
10. Shu X, Fry A, Tulloch B, et al. RPGR ORF15 isoform co-localizes with RPGRIP1 at centrioles and basal bodies and interacts with nucleophosmin. *Human Molecular Genetics.* 2005;14(9):1183-1197.
11. Zhao Y, Hong D-H, Pawlyk B, et al. The retinitis pigmentosa GTPase regulator (RPGR)-interacting protein: subserving RPGR function and participating in disk morphogenesis. *Proceedings of the National Academy of Sciences.* 2003;100(7):3965-3970.
12. Mavlyutov TA, Zhao H, Ferreira PA. Species-specific subcellular localization of RPGR and RPGRIP isoforms: implications for the phenotypic variability of congenital retinopathies among species. *Human molecular genetics.* 2002;11(16):1899-1907.
13. Castagnet P, Mavlyutov T, Cai Y, Zhong F, Ferreira P. RPGRIP1s with distinct neuronal localization and biochemical properties associate selectively with RanBP2 in amacrine neurons. *Human molecular genetics.* 2003;12(15):1847-1863.
14. Gerber S, Perrault I, Hanein S, et al. Complete exon-intron structure of the RPGR-interacting protein (RPGRIP1) gene allows the identification of mutations underlying Leber congenital amaurosis. *European journal of human genetics: EJHG.* 2001;9(8):561.

15. Patnaik SR, Raghupathy RK, Zhang X, Mansfield D, Shu X. The Role of RPGR and Its Interacting Proteins in Ciliopathies. *Journal of ophthalmology.* 2015;2015:414781.
16. Boylan JP, Wright AF. Identification of a novel protein interacting with RPGR. *Human molecular genetics.* 2000;9(14):2085-2093.
17. Consugar MB, Navarro-Gomez D, Place EM, et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible and more sensitive for variant detection than exome sequencing. *Genetics in medicine: official journal of the American College of Medical Genetics.* 2015;17(4):253.
18. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-291.
19. Falk MJ, Zhang Q, Nakamaru-Ogiso E, et al. NMNAT1 mutations cause Leber congenital amaurosis. *Nature genetics.* 2012;44(9):1040-1045.
20. Handsaker RE, Van Doren V, Berman JR, et al. Large multiallelic copy number variations in humans. *Nature genetics.* 2015;47(3):296-303.
21. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* Jan 1 2013;29(1):15-21.
22. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol.* Jan 2011;29(1):24-26.
23. Li T. Leber congenital amaurosis caused by mutations in RPGRIP1. *Cold Spring Harbor perspectives in medicine.* 2015;5(4):a017384.
24. Baker KE, Parker R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Current opinion in cell biology.* 2004;16(3):293-299.
25. Usher CL, McCarroll SA. Complex and multi-allelic copy number variation in human disease. *Briefings in functional genomics.* Sep 2015;14(5):329-338.
26. Farkas MH, Grant GR, White JA, Sousa ME, Consugar MB, Pierce EA. Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics.* Jul 18 2013;14:486.
27. Calarco JA, Superina S, O'Hanlon D, et al. Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell.* Sep 4 2009;138(5):898-910.
28. McFall RC, Sery TW, Makadon M. Characterization of a new continuous cell line derived from a human retinoblastoma. *Cancer research.* Apr 1977;37(4):1003-1010.
29. Dryja TP, Adams SM, Grimsby JL, et al. Null RPGRIP1 alleles in patients with Leber congenital amaurosis. *The American Journal of Human Genetics.* 2001;68(5):1295-1298.
30. Foxman SG, Heckenlively JR, Bateman JB, Wirtschafter JD. Classification of congenital and early onset retinitis pigmentosa. *Archives of ophthalmology.* Oct 1985;103(10):1502-1506.
31. Kaplan J, Bonneau D, Frezal J, Munnich A, Dufier JL. Clinical and genetic heterogeneity in retinitis pigmentosa. *Human genetics.* Oct 1990;85(6):635-642.
32. Sangermano R, Bax NM, Bauwens M, et al. Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T-->C Mutation in Stargardt Disease. *Ophthalmology.* Jun 2016;123(6):1375-1385.

33. Schneider B, Koppius A, Sedlmeier R. Use of an exon-trapping vector for the evaluation of splice-site mutations. *Mammalian genome : official journal of the International Mammalian Genome Society.* Sep 2007;18(9):670-676.

34. Kimberling WJ. Estimation of the frequency of occult mutations for an autosomal recessive disease in the presence of genetic heterogeneity: application to genetic hearing loss disorders. *Human mutation.* 2005;26(5):462-470.

35. Zallocchi M, Binley K, Lad Y, et al. EIAV-based retinal gene therapy in the shaker1 mouse model for usher syndrome type 1B: development of UshStat. *PLoS One.* 2014;9(4):e94272.

36. Banin E, Gootwine E, Obolensky A, et al. Gene augmentation therapy restores retinal function and visual behavior in a sheep model of CNGA3 achromatopsia. *Molecular Therapy.* 2015;23(9):1423-1433.

37. Komáromy AM, Alexander JJ, Rowlan JS, et al. Gene therapy rescues cone function in congenital achromatopsia. *Human molecular genetics.* 2010;19(13):2581-2593.

38. Bainbridge JW, Smith AJ, Barker SS, et al. Effect of gene therapy on visual function in Leber's congenital amaurosis. *New England Journal of Medicine.* 2008;358(21):2231-2239.

39. MacLaren RE, Groppe M, Barnard AR, et al. Retinal gene therapy in patients with choroideremia: initial findings from a phase 1/2 clinical trial. *The Lancet.* 2014;383(9923):1129-1137.

40. Maguire AM, Simonelli F, Pierce EA, et al. Safety and efficacy of gene transfer for Leber's congenital amaurosis. *New England Journal of Medicine.* 2008;358(21):2240-2248.

Figure legends

Figure 1 Exon 2 duplication in OGI-287. (a) Intergrated genome viewer (IGV) view of reads spanning exon 2 of *RPGRIP1* in OGI281-608. **(b)** Pedigree of OGI-281. **(c)** Demonstration of the indication of tandem duplication from the spanning reads showin in (a). **(d)** qPCR confirmation of copy number gain in family members. Error bars are standard deviations (n=3). **(e)** PCR across the predicted breakpoint using primers outside of the spanning reads as shown in (c). A separate region of DNA was amplified as control for DNA quality and PCR conditions. **(f)** Sanger sequencing identifying the exact breakpoint using OGU-608 PCR product from (e).

Figure 2 Exon 1 and 2 duplication in OGI-237 and identification of a novel exon. (a) Pedigree of OGI237 shoeing allele segregation. **(b)** qPCR of OGI237 family members, showing inheritance of the duplicated allele from the mother. **(c)** Mapping of the breakpoint and demonstration of the mutant allel (M4) using the current longest RPGRIP1 transcript model (NM_020366). Coding regions of exons are shown in long and 5'UTR in shorter rectangles, with introns demonstrated as arrowed lines. **(d)** Identification of the novel *RPGRIP1* exon. Red indicates mRNA reads, and green lines are split reads covering exon to exon junctions that are unannotated. Blue lines are split reads across annotated exons. There are 550 reads spanning *RPGRIP1* exon 1 to the newly discovered exon 8Kb upstream of exon 1. **(e)** ChIP-Seq of various transcription markers and ATAC-Seq confirm a regulartory region in vicinity of the novel exon shown in (d).

Figure 3 Sashimi plots of the amplicon sequencing of splicing assays. (a) OGI-827. **(b)** OGI-601. **(c)** OGI-949. The annotated NM_020366 exons are indicated on the bottom of each plot in dark blue. Panel (b) also includes the transcript variant XM_005267881.3 with 5'UTR shown as a thinner line compared to coding exons. The sashimi plots are generated using amplicon

sequencing on RT-PCR products across the indicated regions in HEK293T cells. Filter of split-read counts greater than 2200 reads was used to generate these figures. The peaks on the 5' and 3' of the amplicons are secondary to lack of complete shearing of the product resulting in an expected over-representation of the first and last 142bp from paired-end sequencing.

Figure 4 Summary of *RPGRIP1* mutations and gene model. Novel findings are indicated in red with non-coding mutations listed below the NM_020366 transcript model and coding mutations listed above. The RPGRIP1 protein and corresponding domains is shown in the bottom.

Supplementary figure legends

Figure S1 Overview of the approach. Of the 1722 families that underwent GEDi sequencing, ten had at least one loss of function (LoF) mutation in *RPGRIP1*. This included three families that were solved with biallelic LoF mutations and seven families that underwent further investigation.

Figure S2 OGI1797. Pedigree of the family is shown on top, with copy number validation shown on the bottom as qPCR results. Error bars are standard deviations (n=3).

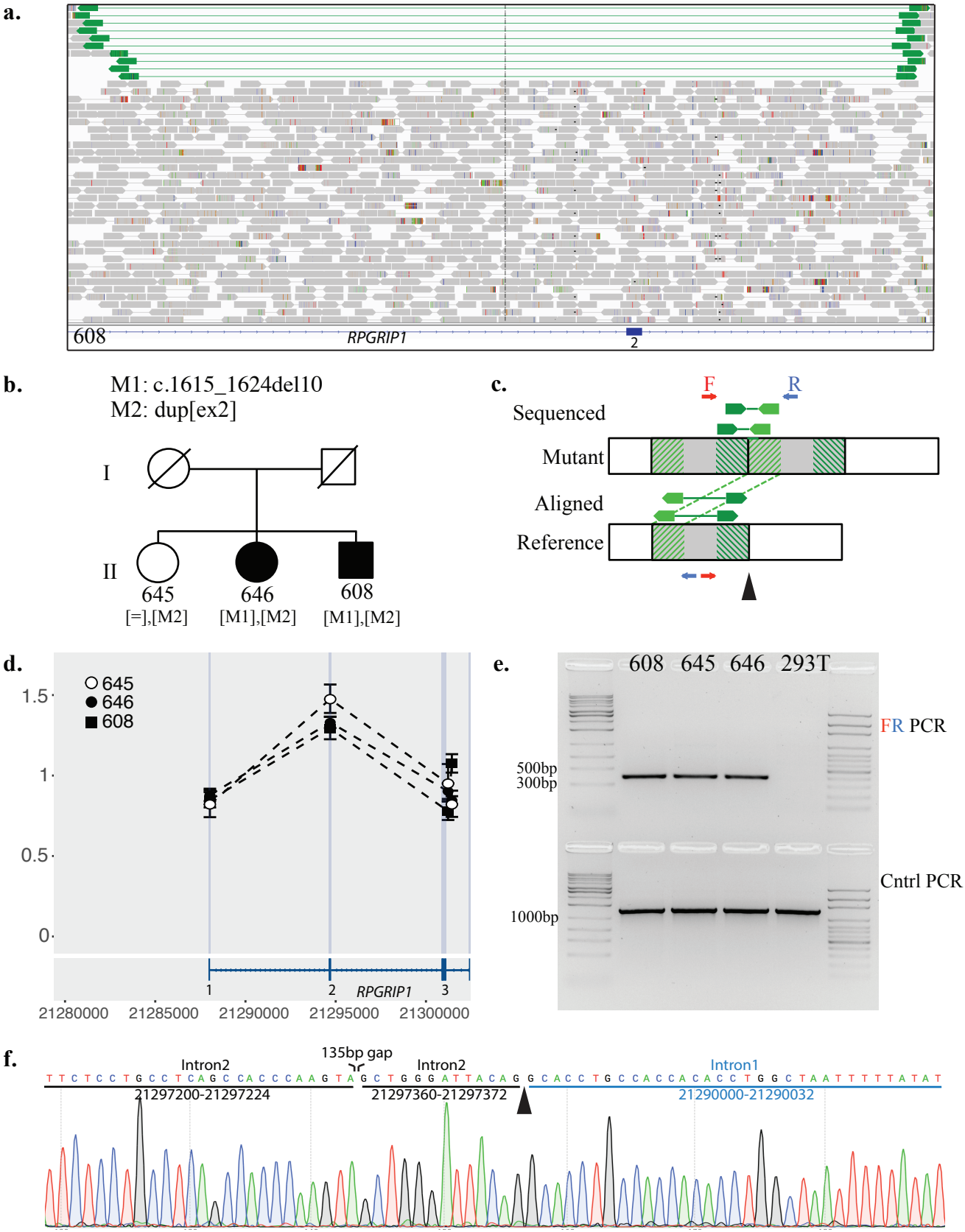
Figure S3 Confirmation of tandem duplication in OGI-237. Top: the model of the mutant allele with the breakpoint of tandem duplication indicated via arrow-head. Bottom: PCR confirmation of the tandem duplication using primers across the breakpoint. Primers at a distant *RPGRIP1* locus (exon 14-15) were uses as control for PCR reaction and DNA quality.

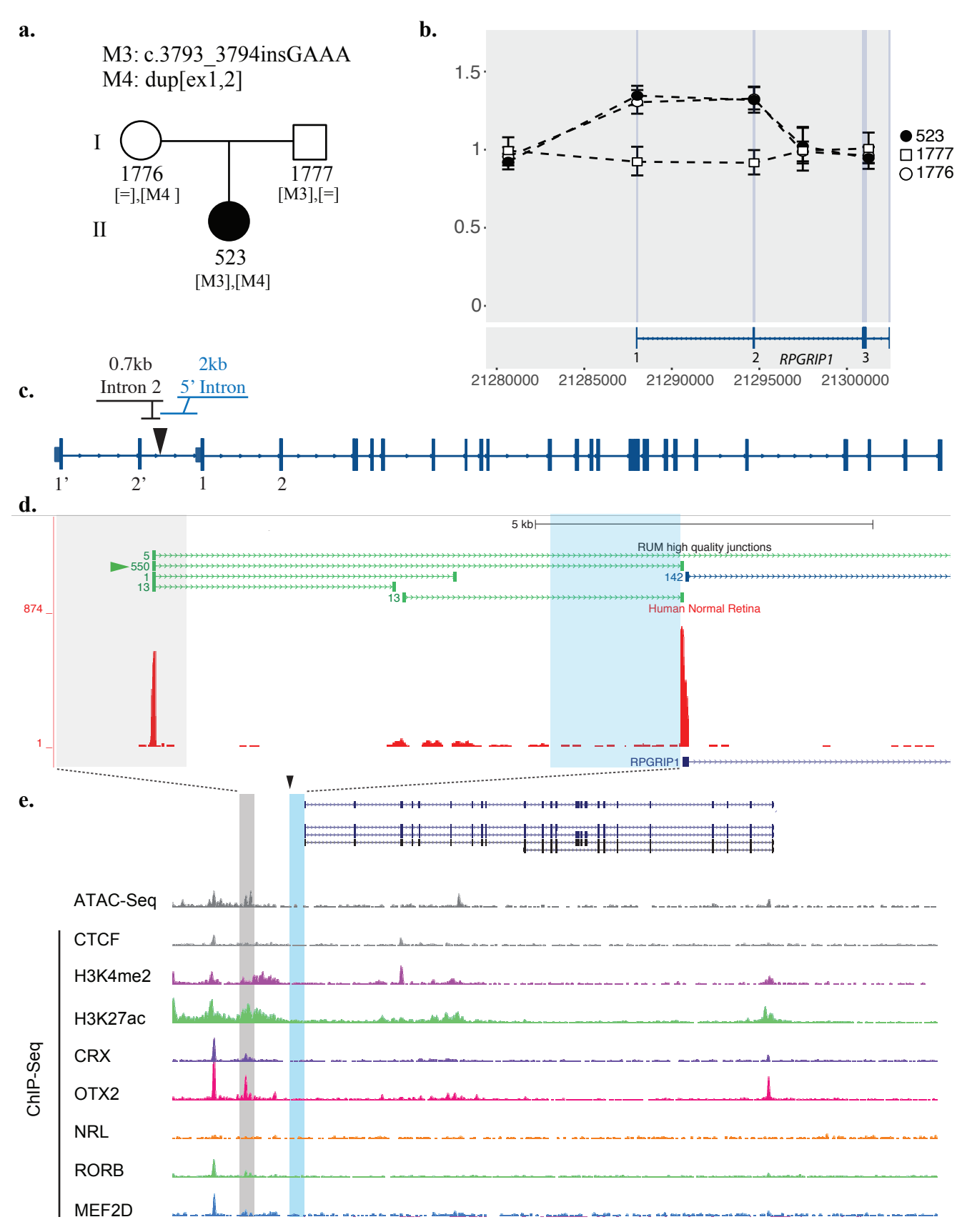
Figure S4 Transfection conditions for WERI-Rb1. Images of the cell line with GFP control plasmid transfection are shown. Objective lens magnifications are shown on the left. The top right inserts are trypan blue staining for cell counting on a hemocytometer. The condition used for transfections was 5ug Vector, 6ul of Lipofectamine 2000 per well in a 6-well plate.

Figure S5 Midi-gene splicing assay of intronic mutation. Gel images of the RT-PCR reactions are shown to the right of each pedigree.

Figure S6 Amplicon-sequencing showing similar splicing pattern in HEK293T and WERI-Rb1 lines. Minimum split read counts of 1500-5000 were used.

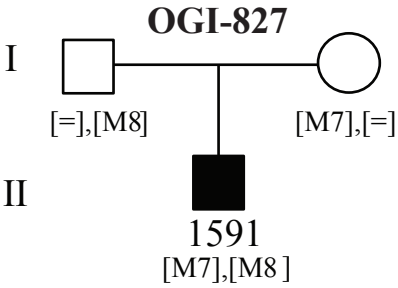
Figure S7 Map of the PCS2 vector with *RPGRIP1* exons 11-16 insert. This vector was used for the midi-gene splicing assay.

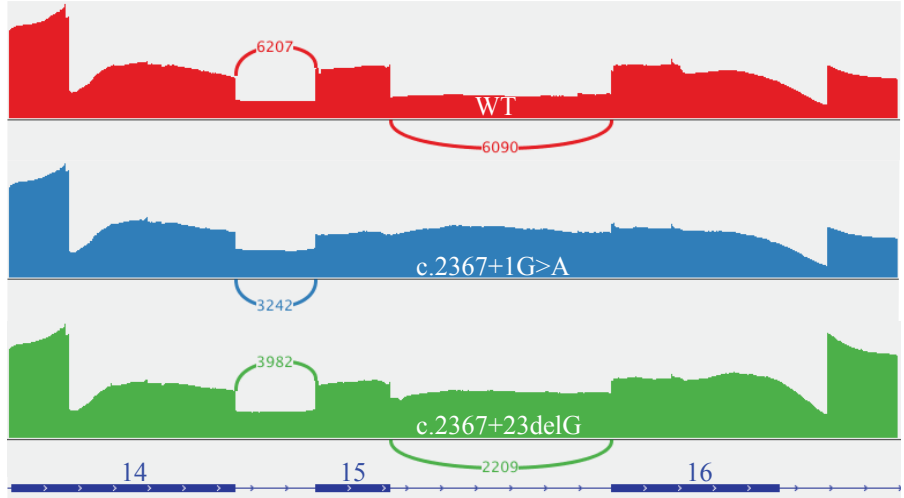

a.
608
RPGRIP1
2
b.
M1: c.1615_1624del10
M2: dup[ex2]
I
II
645
[=],[M2]
646
[M1],[M2]
608
[M1],[M2]
c.
F
R
Sequenced
Mutant
Aligned
Reference
d.
645
646
608
1.5
1
0.5
0
1
2
RPGRIP1
3
21280000
21285000
21290000
21295000
21300000
e.
608
645
646
293T
FR PCR
500bp
300bp
Cntrl PCR
1000bp
f.
135bp gap
Intron2
Intron2
Intron1
21297200-21297224
21297360-21297372
21290000-21290032


a.
M3: c.3793_3794insGAAA
M4: dup[ex1,2]
I
1776
[=],[M4]
1777
[M3],[=]
II
523
[M3],[M4]
b.
1.5
1
0.5
0
523
1777
1776
1
2
RPGRIP1
3
21280000
21285000
21290000
21295000
21300000
c.
0.7kb
Intron 2
2kb
5' Intron
1'
2'
1
2
d.
5 kb
RUM high quality junctions
5
550
1
13
13
142
874
Human Normal Retina
1
RPGRIP1
e.
ATAC-Seq
ChIP-Seq
CTCF
H3K4me2
H3K27ac
CRX
OTX2
NRL
RORB
MEF2D

a.

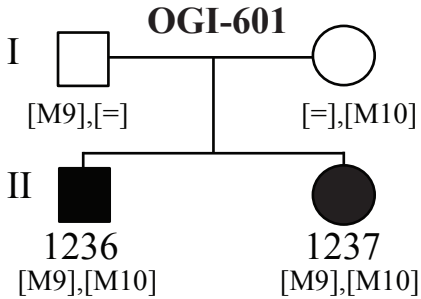
M7: c.895_896delGA
M8: c.2367+23delG

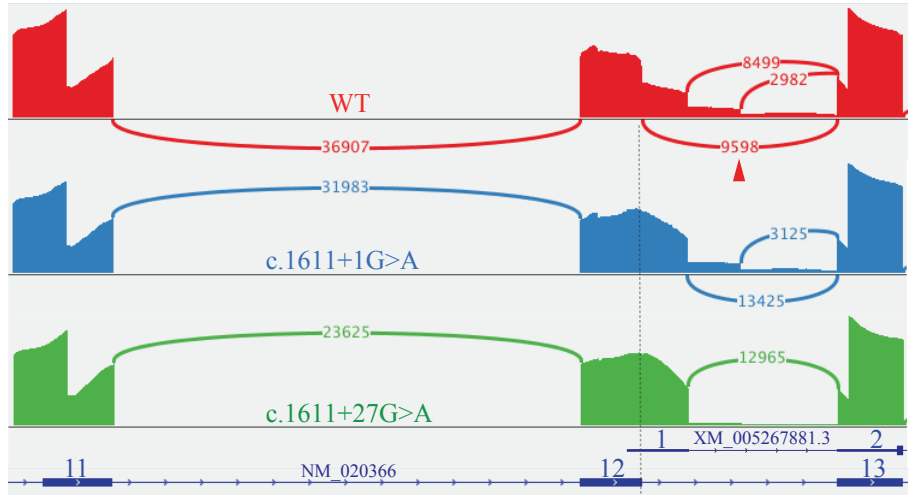




b.

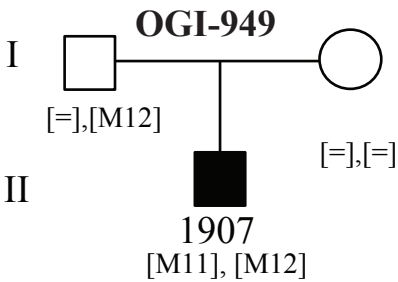
M9: c.3238+1G>A
M10:c.1611+27G>A

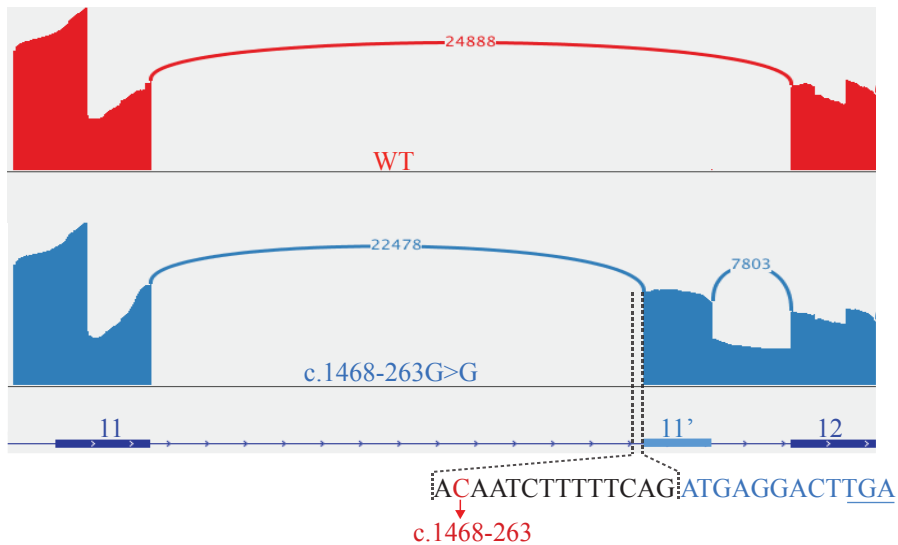


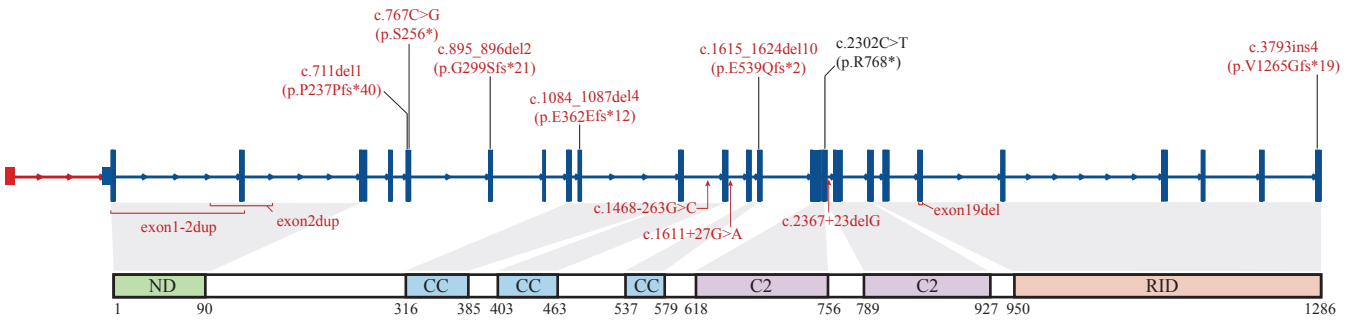


c.

M11:c.3618-1_3621del5
M12:c.1468-263G>C






c.767C>G
(p.S256*)
c.895_896del2
(p.G299Sfs*21)
c.711del1
(p.P237Pfs*40)
c.1084_1087del4
(p.E362Efs*12)
c.1615_1624del10
(p.E539Qfs*2)
c.2302C>T
(p.R768*)
c.3793ins4
(p.V1265Gfs*19)
exon1-2dup
exon2dup
c.1468-263G>C
c.1611+27G>A
c.2367+23delG
exon19del
ND
CC
CC
CC
C2
C2
RID
1
90
316
385
403
463
537
579
618
756
789
927
950
1286