Prevalence of 

RPGR-Mediated Retinal Dystrophy in an Unselected Cohort of Over 5000 Patients

Sari Tuupanen1,*, Kimberly Gall2, Johanna Sistonen1, Inka Saarinen1, Kati Kämpjärvi1, Kirsty Wells1, Katja Merkkiniemi1, Pernilla von Nandelstadh1, Laura Sarantaus1, Johanna Känsäkoski1, Emma Mårtenson1, Hanna Västinsalo1, Jennifer Schleit2, Eeva-Marja Sankila3, Annakarin Kere1, Heidi Junnila1, Pauli Siivonen1, Margarita Andreevskaya1, Ville Kytölä1, Mikko Muona1, Pertteli Salmenperä1, Samuel Myllykangas1, Juha Koskenvuo1, and Tero-Pekka Alastalo2

1 BlueprintGeneticsOY, Keilaranta, Espoo, Finland
2 BlueprintGeneticsInc, Seattle, WA, USA
3 Helsinki University Eye Hospital, Outpatient Clinic for Hereditary Eye Diseases, Helsinki, Finland

Correspondence: Sari Tuupanen,
Blueprint Genetics OY, Keilaranta 16 A-B, Espoo, Finland.
email: sari.tuupanen@blueprintgenetics.com

Received: January 13, 2021
Accepted: November 22, 2021
Published: January 5, 2022

Keywords: retinitis pigmentosa; RPGR; retinal dystrophy; next-generation sequencing; genetic testing


Introduction

Retinitis pigmentosa (RP) is the most common form of inherited retinal dystrophy (IRD), affecting up to 1 in 3000 people worldwide.1 RP is heterogeneous, both clinically and molecularly, and the inheritance is complex with autosomal dominant, autosomal recessive, and X-linked mechanisms described. Determination of the mode of inheritance is complicated, because many cases are simplex (the only affected individual in a family) and some families with X-linked
or autosomal recessive RP seem to have autosomal dominant RP. Molecular genetic studies are the most reliable way to determine or confirm the diagnosis, provide prognostic information, determine the inheritance pattern, identify at-risk individuals, and allow for family planning.

X-linked RP (XLRP) has one of the most severe retinal phenotypes and accounts for up to 20% of all patients with RP. Variants in RPGR and RP2 cause the majority of XLRP, whereas variants in OFD1 contribute rarely. RPGR is the most commonly implicated gene and variants in this gene account for more than 70% of XLRP. Further, RPGR variants account for approximately 13% of simplex cases and 7.8% of dominant RP. Molecular genetic studies are the most reliable way to determine or confirm the diagnosis, provide prognostic information, determine the inheritance pattern, identify at-risk individuals, and allow for family planning.

Isoform C (RefSeq NM_001034853.1) of RPGR is primarily expressed in the retina and encodes an 1152-amino acid protein comprising 3459 nucleotides. It contains the ORF15 exon, which encodes 567 amino acids at the C-terminus, is rich in glutamic acid and glycine residues, and is highly repetitive. ORF15 encompasses nucleotides c.1754–3459. Further, ORF15 harbors a central, difficult-to-sequence region defined as c.2470–3230 (RefSeq NM_001034853.1). ORF15 is recognized as a hotspot for disease-causing variants, the most common being small deletions leading to frameshifts. The repetitive nature of the ORF15 exon makes sequencing both difficult and prone to error. Traditionally, clinical testing of ORF15 relied on Sanger sequencing, an expensive and time-consuming method. More recently, long-range polymerase chain reaction plus next-generation sequencing (NGS) strategies have been used. However, because more than one-half of RPGR disease-causing variants are located in ORF15, there is a need for a scalable, high-throughput, and reliable method that can effectively tackle the challenges of this exon.

Further driving the need for molecular diagnoses in families affected by IRD is the promising race toward personalized medicine. Gene therapy trials are underway for a number of different retinal dystrophies, and many rely on a confirmed molecular diagnosis (https://clinicaltrials.gov). RPGR is an attractive target given its frequency and the severity of disease. Additionally, the size of RPGR is amenable to gene replacement therapies.

Given the recent developments in gene therapy and the importance of RPGR in IRD, there is a need for an analytically validated, high-throughput, NGS-based diagnostic assay with the ability to resolve ORF15 that detects both sequence and copy number variants (CNVs). In this study, we retrospectively reviewed NGS multi-gene panel testing results from 5201 unselected individuals with suspected retinal dystrophy to assess the contribution of RPGR to the diagnostic yield, the variant characteristics and the impact of high-resolution CNV detection.

### Methods

This retrospective study included 5201 consecutive patients referred and consented for IRD genetic testing in addition to aggregate sharing of the information such as publication. Patients were referred from the United States (86%), Canada (9%), Europe (4.5%), Latin America (0.2%), the Middle East (0.4%), and the South Pacific (0.2%). The majority of patients were tested as part of the My Retina Tracker program in collaboration with the Foundation Fighting Blindness. Detailed clinical information was not available to the authors. Given the lack of detailed phenotypic information, the goal of this study is to report the frequency and characteristics of RPGR variants identified in this large cohort.

All identifying information was removed from the data and this work received an exemption determination after review by the Western Institutional Review Board (WCG IRB Work Order 1-1378008-1).

Genetic testing was performed at Blueprint Genetics, a College of American Pathologists– and Clinical Laboratory Improvement Amendments–certified laboratory, using a 266-gene retinal dystrophy NGS panel (Supplementary Table S1). In addition to the coding regions, the panel targeted 20 base pairs at the intron/exon boundaries and 81 noncoding variants previously reported as disease causing in association with IRD. The panel was carved out of an in-house tailored whole exome sequencing assay. The CNV analysis was performed bioinformatically concurrently for all patients from the NGS data using a commercially available bioinformatic pipeline CNVkit and a proprietary, in-house developed deletion caller based on read depth to improve the detection of small CNVs (Supplementary Methods). Sample sex was estimated based on sequencing coverage of the sex chromosomes.

The mean sequencing depth was more than 150× and more than 99.4% of target nucleotides were covered with more than 20× the sequencing depth.
Specifically, the mean sequencing depth for \textit{RPGR} was $82 \times$ for males and $159 \times$ for females and 99.4\% and 100\% of target nucleotides were covered with more than $20 \times$ for males and females, respectively (Supplementary Fig. S1). Bidirectional Sanger sequencing and quantitative polymerase chain reactions were used for confirmation of sequence variants and CNVs, respectively (Supplementary Methods).

Variants were classified according a point-based modification of the Association for Molecular Pathology/American College of Molecular Genetics and Genomics guidelines (www.blueprintgenetics.com), with evidence from population and gene- and disease-specific databases, in silico prediction tools (including PolyPhen, SIFT, and Mutasser), our in-house variant database, multiple publicly and commercially available mutation databases and appropriate scientific literature as the foundation for scoring. All reported variants were de-identified and shared with the ClinVar database. A test result was considered diagnostic when the patient was found to have one or two pathogenic or likely pathogenic variants in a single gene, depending on the gene and mode of inheritance. \textit{RPGR} variant nomenclature is based on NM_001034853.1 from the RefSeq database.

**Results**

A total of 5201 patients underwent sequencing with the clinically validated 266-gene retinal dystrophy panel, including sequence and CNV analysis (Supplementary Table S1). Demographic data were available for 5074 individuals (97.6\%) in the cohort. The median age at referral was 45 years (range, <1 year to 93 years). Males represented 51.6\% ($n = 2620$) of the group while females represented 47.8\% ($n = 2425$). Sex was not provided for 0.02\% of the cohort ($n = 29$). The majority of patients tested were referred from the United States (86\% [$n = 4538$]); however, patients were referred from an additional 20 countries from North America, Europe, South America, the Middle East, and the South Pacific.

Diagnostic variants were identified in 51.1\% of patients in 139 unique genes. The median age of patients who received a diagnosis was 32 years (range, <1 year to 93 years). Males represented 50.3\% of the diagnosed cohort and females represented 48.8\%. Age and sex were not available for 4 and 23 patients, respectively, in the diagnosed cohort. Variants in \textit{ABCA4} were responsible for the most diagnoses (14.5\%), followed by \textit{USH2A} (12.0\%), \textit{RPGR} (8.7\%), \textit{RHO} (5.2\%), and \textit{PRPH2} (4.6\%) (Supplementary Table S2). Interestingly, 11 patients had disease-causing variants in two genes, suggestive of dual diagnoses. CNVs made up 4.7\% of all diagnostic variants with \textit{USH2A}, \textit{EYS}, and \textit{PRPF31} being the most frequently implicated (Supplementary Table S2). Of the genes in which diagnostic variants were identified, 117 genes contributed to less than 1.0\% each of the diagnostic burden, but cumulatively contributed to 25.9\%.

Of the 5201 patients tested, 257 (5\%) had a rare variant in \textit{RPGR}; 252 (98\%) had sequence variants and 5 (1.9\%) had CNVs. In 96 out of 252 patients (38\%), the identified sequence variant was confirmed with Sanger sequencing, whereas four of five CNVs were confirmed with quantitative polymerase chain reaction (data not shown). A total of 158 unique \textit{RPGR} variants were detected in these 257 patients (Supplementary Table S3). Of the 257 diagnostic or potentially diagnostic variants, 157 (61\%) were classified as pathogenic, 72 (28\%) as likely pathogenic, and 23 (9\%) as variants of uncertain significance. The \textit{RPGR} variant was the diagnostic finding in 234 of the 5201 patients (4.5\%).

Of the 234 patients with diagnostic \textit{RPGR} variants, 177 (76\%) were male and 57 (24\%) were female (Table). All male patients with a diagnostic \textit{RPGR} variant ($n = 177$) were reported to have visual symptoms. Of the women, 52 (91\%) were reported to have visual symptoms, 4 (7\%) were reported to have ocular findings consistent with being a carrier of an X-linked ocular disease, and 1 (1.7\%) was tested because of a family history consistent with X-linked ocular disease (data not shown). The clinical information provided revealed that the majority of patients with a disease-causing \textit{RPGR} variant were suspected to have RP or rod–cone dystrophy (73\% of males and 61\% of females) (Table). A clinical diagnosis of cone, cone–rod, or macular dystrophy was reported for 11.3\% of males and 5.3\% of females. A subset of patients (2.6\%) was referred with a diagnosis of unspecified retinal dystrophy, and symptoms only (no diagnosis) were provided for 15.4\%. Chorioretinal atrophy, fundus albipunctatus, or Oguchi disease were the suspected clinical diagnoses provided at the time of referral for genetic testing for four female patients.

The median age at testing was 34 years in males (range, 4–81 years) and 46 in females (range, 5–88). There was limited information regarding the age of symptom onset in our cohort. A positive family history of ocular disease was reported in 52\% of males and in 56\% of females, whereas 12.4\% and 14\%, respectively, were reported to have a negative family history (Table). No family history information was provided for the remainder of the cohort (35.6\% of males; 30\% of females) (Table).

Of the patients with diagnostic \textit{RPGR} results, 74\% (173/234) had a variant in ORF15.
Table. Demographics of Patients With a Diagnostic RPGR Variant

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>177 (75.6)</td>
<td>57 (24.4)</td>
<td>234</td>
</tr>
<tr>
<td>Reported diagnosis at time of referral for genetic testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP or rod cone dystrophy</td>
<td>129 (73)</td>
<td>35 (61)</td>
<td>164 (70)</td>
</tr>
<tr>
<td>Cone/cone–rod dystrophy or macular dystrophy</td>
<td>20 (11.3)</td>
<td>3 (5.3)</td>
<td>23 (8.5)</td>
</tr>
<tr>
<td>Chorioretinal atrophy</td>
<td>0</td>
<td>2 (3.5)</td>
<td>2 (0.9)</td>
</tr>
<tr>
<td>Fundus albipunctatus</td>
<td>0</td>
<td>1 (1.8)</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Oguchi disease</td>
<td>0</td>
<td>1 (1.8)</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Retinal dystrophy, unspecified</td>
<td>5 (2.8)</td>
<td>1 (1.8)</td>
<td>6 (2.6)</td>
</tr>
<tr>
<td>Only symptoms listed</td>
<td>23 (13)</td>
<td>13 (23)</td>
<td>36 (15.4)</td>
</tr>
<tr>
<td>Family history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>92 (52)</td>
<td>32 (56)</td>
<td>124 (53)</td>
</tr>
<tr>
<td>Negative</td>
<td>22 (12.4)</td>
<td>8 (14)</td>
<td>30 (13)</td>
</tr>
<tr>
<td>Not mentioned</td>
<td>63 (35.6)</td>
<td>17 (30)</td>
<td>80 (34)</td>
</tr>
</tbody>
</table>

Values are number (%).

disease-causing variants were located in the central, most difficult-to-sequence region in 68 of the 234 patients (29%) and were outside the central region in 105 (45%). The remaining 61 diagnostic variants (26%) were located within exons 1–14 (Fig. a). Diagnostic variants were frameshifts (161 [69%]), nonsense (47 [20%]), missense (14 [6%]), splice (7 [3%]), and CNVs (5 [2%]) (Fig. b). There was no difference in variant distribution or variant type between male and female patients (Fig.). All diagnostic missense variants were located within exons 3–10, affecting the regulator of chromosome condensation (RCC1) domain (amino acids 54–367) (Supplementary Table S2). Eight missense variants and three in-frame deletions affecting conserved amino acid residues within the RCC1 domain, classified as variants of uncertain significance, were detected.

The diagnostic RPGR ORF15 variants outside of the central region consisted of frameshift (87/105 [83%]) and nonsense variants (18/105 [17%]). The most commonly observed pathogenic variant was c.2403_2406del, p.(Glu802Glyfs*12), followed by c.2236_2237del, p.(Glu746Argfs*23) and c.2426_2427del, p.(Glu809Glyfs*25). All 20 male patients with a clinical diagnosis of cone, cone–rod, or macular dystrophy had an ORF15 variant; 95% were located toward the 3’ end at, or downstream of, position c.3027 (c.3092_3093del and c.3096_3097del were both seen in four patients). Four diagnostic CNVs were detected in five patients, including a deletion of exon 2 (in-frame), a deletion of exons 2–3 (in-frame), a deletion of exons 8–9, and a whole gene deletion (out-of-frame) (Supplementary Table S2).

Discussion

This study presents the molecular diagnoses in an unselected cohort of 5201 patients with IRD referred for NGS-based genetic testing with a focus on the contribution and characteristics of diagnostic RPGR variants. Importantly, RPGR was the third most common genetic etiology of IRD in this cohort, after ABCA4 and USH2A. The distribution of the most common causative genes is similar to those previously reported by other groups.16,17 Interestingly, more than one-quarter of the diagnoses in our cohort were due to 1 of 117 genes that contributed less than 1.0% each of the diagnostic burden, highlighting the significant genetic heterogeneity in the IRD population and the value of a broad multigene panel approach to maximize diagnostic usefulness.

Our study is consistent with previous NGS-based screening studies on IRD cohorts reporting RPGR diagnostic yields of 4.8% (48/1000)16 and 4% (21/500).18 Notably, the NGS assays used in these studies did not capture ORF15, and alternate technologies were required to detect ORF15 variants making them less optimal for routine high-throughput genetic diagnostics for patients with IRD because they require multiple laboratory steps. To the best of our knowledge, our study is the largest, and first, patient cohort published to date to evaluate the prevalence and characteristics of RPGR variants analyzed using only an NGS-based assay with laboratory methods and bioinformatic analyses that cover all of ORF15, include CNV analysis, and demonstrate similar diagnostic
Figure. RPGR variant characteristics in 234 diagnostic cases. (a) Distribution of variants by location in the gene. ORF15 central region encompasses c.2470-3230, p.824–1077. (b) Variant type.
yields to those studies with additional steps required to cover ORF15 sufficiently.

Studies on the prevalence of RPGR-associated disease in large cohorts of female patients with IRD are scarce. A study describing the phenotypic spectrum in 125 females with a disease-causing RPGR variant observed complete expression of RP or cone–rod dystrophy in 23% of the cases, again supported by our data. Interfamilial and intrafamilial variability of disease penetrance in female patients is observed and is at least partially attributed to random X chromosome inactivation. This clinical variability in females can confound the determination of the mode of inheritance in RPGR families as the family history may seem to be consistent autosomal dominant disease rather than an X-linked disease. Importantly, female patients with complete disease expression may also benefit from ongoing and future therapeutic trials, further highlighting the value of a molecular diagnosis.

In addition, it is important to note that there are many genes implicated in other isolated X-linked retinopathies, including NYX and CACNA1F (congenital stationary night blindness), OPNILW and OPNILMW (blue cone monochromacy), RSI (X-linked retinoschisis), and CHM (choroideremia). Although these conditions are typically distinct when resources such as electroretinography and expert examination are available, there are phenotypic overlaps in individuals with these conditions and XLRP, so ensuring that these genes are included when selecting testing for these patients is prudent.

Our findings are consistent with the previously reported variant characteristics and genotype–phenotype associations of RPGR-associated IRD:

1. ORF15 is a mutation hotspot (Fig.a). In previous studies, based on 135 and 47 XLRP patients, variants in ORF15 accounted for 71.4% and 80.0% of the diagnoses, respectively.
2. Disease-causing missense variants are located within the RCC1 domain in exons 3–10.
3. The cone, cone-rod, or macular dystrophy phenotype is associated with variants located toward the 3’ end of ORF15, as previously described by Branham et al.

Notably, our study demonstrates an important contribution of small CNVs among RPGR diagnostic variants. All detected CNVs were deletions, four out of five were small (one or two exons), and none had been described previously in the literature. Published NGS-based studies have previously reported only a few gross deletions in RPGR. Whole exome sequencing on 60 patients with IRD described deletions affecting exons 4–11 and 2–5, whereas targeted panel analysis on 500 unrelated patients with IRD reported only one hemizygous 18.7-kb deletion affecting exon 15. The Human Gene Mutation Database reports 10 partial gross deletions in RPGR/ORF15, which represent 2% of all disease-causing variants in the database.

In our cohort of patients with a diagnostic RPGR variant, four female patients were referred for genetic testing with a clinical diagnosis of chorioretinal atrophy, fundus albipunctatus, or Oguchi disease. These diagnoses are not associated with RPGR-related disease and these cases demonstrate one of the important benefits of genetic testing, that is, refining, or even changing, the initial clinical diagnosis.

Although the provision of detailed clinical information to diagnostic laboratories is always recommended for the most robust interpretation of variants, it is rarely mandatory outside of research or clinical trial protocols. As such, the major limitation of our study is the limited clinical information provided for the patients. This factor hampers detailed clinical characterization or the ability to determine what characteristics increase the likelihood of a patient having RPGR-associated disease before performing genetic testing. As genetic testing continues to become more common in the primary health care setting, comprehensive genetic testing panels are advantageous for patients with IRD to maximize the likelihood of making a molecular diagnosis. However, the value of involving a retinal specialist in the care of these patients cannot be understated. Benefits include expert phenotyping, clinical correlation with molecular results, ensuring appropriate management, access to therapies, and referral to clinical trials when appropriate.

In this unselected and heterogeneous cohort, RPGR variants are a significant cause of IRD. High-throughput, clinically validated NGS diagnostics for IRD, including RPGR ORF15 in combination with high-resolution CNV detection, are largely lacking. In this study, 29% of the diagnostic variants were detected in the most difficult-to-sequence central region of ORF15 (c.2470-3230) and CNVs accounted for 2% of RPGR diagnoses. These findings highlight the importance of including deep and uniform sequencing depth of ORF15 and high-resolution, sensitive CNV analysis in diagnostic NGS-panels. Comprehensive IRD molecular testing is critical for patient management, family counselling, management, and selection for ongoing clinical trials. There are multiple gene therapy trials ongoing for inherited retinal dystrophies, including XLRP caused by RPGR variants and RPGR-associated retinal dystrophy. RPGR is ideally suited for gene therapy given its prevalence and severe clinical presentation. Our findings confirm the importance of including high-quality RPGR gene analysis.
as part of primary genetic testing for patients with IRD.

Acknowledgments

The authors express their gratitude to patients and medical providers who participated in this study.

Author Contributions

ST, JK, and TPA contributed to the conception and design of the study.

ST, KG, IS, and TPA compiled the clinical and variant data and performed the analyses.

ST, JSI, KK, KW, KM, PVN, LS, JK, EM, JSC, TPA, and JK analyzed individual patient data and performed variant interpretation.

JSI, AK, HJ, PSI, MA, VK, MM, PSA, SM, TPA, and JK developed and implemented the methodology used for DNA analysis, and variant interpretation.

ST, KG, JSI, and TPA drafted and wrote the manuscript.

All authors reviewed and approved the final version for submission.

Disclosure: S. Tuupanen, Blueprint Genetics (E); K. Gall, Blueprint Genetics (E); J. Sistonen, Blueprint Genetics (E); I. Saarinen, Blueprint Genetics (E); K. Kämpjärvi, Blueprint Genetics (E); K. Merkkinen, Blueprint Genetics (E); P. von Nandelstad, Blueprint Genetics (E); L. Sarantaus, Blueprint Genetics (E); J. Künsäkoski, Blueprint Genetics (E); E. Märtenson, Blueprint Genetics (E); H. Västinsalo, Blueprint Genetics (E); J. Schleit, Blueprint Genetics (E); E.-M. Sankila, Blueprint Genetics (C); Novartis (C, R); Biogen (C, R) and Bayer (C, R); A. Kere, Blueprint Genetics (E); H. Junnila, Blueprint Genetics (E); P. Siivonen, Blueprint Genetics (E); M. Andreevskaia, Blueprint Genetics (E); V. Kytölä, Blueprint Genetics (E); M. Muona, Blueprint Genetics (E); P. Salmenperä, Blueprint Genetics (E); S. Myllykangas, Blueprint Genetics (E); J. Koskenuo, Blueprint Genetics (E); T.-P. Alastalo, Blueprint Genetics (E)

References

11. Talevich E., et al. CNVkit: genome-wide copy number detection and visualization from targeted


