Likely pathogenic structural variants in genetically unsolved patients with retinitis pigmentosa revealed by long-read sequencing

Yusuke Sano,1,2 Yoshito Koyanagi,1,3 Jing Hao Wong,2 Yusuke Murakami,1 Kohta Fujiwara,1 Mikiko Endo,4 Tomomi Aoi,4 Kazuki Hashimoto,5 Toru Nakazawa,5,6 Yuko Wada,7 Shinji Ueno,8 Dan Gao,9 Akira Murakami,9 Yoshihiro Hotta,10 Yasuhiro Ikeda,11 Koji M Nishiguchi,8 Yukihide Momozawa,4 Koh-Hei Sonoda,1 Masato Akiyama,3,12 Akihiro Fujimoto2

INTRODUCTION

Retinitis pigmentosa (RP; OMIM:268000) is a prevailing form of inherited retinal dystrophy (IRD) and a major cause of blindness worldwide. RP is inherited in autosomal dominant, autosomal recessive or X-linked inheritance patterns following Mendel’s law of inheritance, with very few exceptions.1 So far, more than 80 genes have been reported as genetic causes of RP.2 According to previous studies, 40%–70% of patients with RP do not have a molecular diagnosis.1,4–6 Nevertheless, advances in sequencing technology have enabled scientists to reveal the genetic causes of RP.3,13–15 Previous studies have demonstrated that EYS, USH2A, RPGR and RHO are the frequent causative genes of RP across ethnicities.1,4,7 For Japanese patients with RP, EYS, which is the cause of autosomal recessive RP (ARRP), has been reported as the most frequent causative gene.9,10

We previously reported the targeted short-read sequencing of 83 RP-associated genes in 1204 Japanese patients with RP.11 In that study, custom-made multiplex PCR-based targeted resequencing was performed using a next-generation sequencer. As a result, causative genes were determined in 29.6% of patients; however, the remaining cases remained unsolved. Importantly, the study indicated that more than a quarter of unsolved patients had one pathogenic variant in EYS,11 suggesting that EYS is a promising candidate as the causative gene.11 Although the exact reason for this observation has not been clarified, it is possible that variants are difficult to identify by short-read targeted resequencing, such as structural variants (SVs) and intronic variants, are another causative variant. Indeed, SVs, intronic variants and hypomorphic variants have been uncovered as a second pathogenic variant in patients with IRD.9,12–15 Therefore, we considered whether carriers of pathogenic variants in EYS have second causative variants that are not detectable by short-read target sequencing.11

Long-read sequencing technology has several advantages compared with short-read sequencing, such as accurate SV detection.16,17 In order to increase the number of solved patients with RP and to deepen our knowledge about the genetic basis of RP here we conducted a sequencing study of 15 patients with RP who are carriers of one pathogenic variant in EYS using long-read sequencing and examined the frequencies of the identified SVs from an independent patient group in our previous study (n=1189).11

MATERIALS AND METHODS

Study patients

The clinical diagnosis of RP was based on the patient’s history, visual field and electroretinography outcomes, as well as ophthalmological findings by trained ophthalmologists. We selected 15 patients with RP, EYS, which is the cause of autosomal recessive RP (ARRP), has been reported as the most frequent causative gene.9,10

We previously reported the targeted short-read sequencing of 83 RP-associated genes in 1204 Japanese patients with RP.11 In that study, custom-made multiplex PCR-based targeted resequencing was performed using a next-generation sequencer. As a result, causative genes were determined in 29.6% of patients; however, the remaining cases remained unsolved. Importantly, the study indicated that more than a quarter of unsolved patients had one pathogenic variant in EYS,11 suggesting that EYS is a promising candidate as the causative gene.11 Although the exact reason for this observation has not been clarified, it is possible that variants are difficult to identify by short-read targeted resequencing, such as structural variants (SVs) and intronic variants, are another causative variant. Indeed, SVs, intronic variants and hypomorphic variants have been uncovered as a second pathogenic variant in patients with IRD.9,12–15 Therefore, we considered whether carriers of pathogenic variants in EYS have second causative variants that are not detectable by short-read target sequencing.11

Long-read sequencing technology has several advantages compared with short-read sequencing, such as accurate SV detection.16,17 In order to increase the number of solved patients with RP and to deepen our knowledge about the genetic basis of RP here we conducted a sequencing study of 15 patients with RP who are carriers of one pathogenic variant in EYS using long-read sequencing and examined the frequencies of the identified SVs from an independent patient group in our previous study (n=1189).11

For validation purposes, we
of search for the following types of SVs: (1) SVs in the coding region of EYS genes, (2) SVs which overlap human retina-specific exons of EYS that may play important roles in retinal diseases,20 (3) SVs within 500 bp of the exon boundaries of EYS that could affect splicing or promoter functions and (4) SVs in the RefSeq coding regions of other RP genes that could be causative variants. Variant classification was performed for the detected variants according to the guidelines of the American College of Medical Genetics and Genomics (ACMG).21

To prioritise pathogenic SVs, the allele frequencies (AFs) of the SVs were compared with those in our in-house long-read whole-genome sequencing data (53 Japanese healthy control samples), Genome Aggregation Database (gnomAD SVs v2.1) and dbVar database.22 We removed variants using the following criteria: (1) variants found in the 53 Japanese control samples, and (2) variants with an AF > 0.5% for recessive genes and > 0.01% for dominant genes in gnomAD SVs or dbVar.

Among the SVs overlapping the EYS region, we excluded three large SVs (110 Mb inversion, 82 Mb inversion and 97 Mb translocation) to avoid variant-calling errors using the following criteria: (1) the size of the SVs was too large, (2) there were only two reads supporting the breakpoints of the SVs and (3) the 1 kb sequence around the breakpoint was aligned with multiple regions of the human genome and likely to be a repeat region. We noted that the excluded SVs were less likely to be pathogenic, since their breakpoints did not overlap with the regions of RP-associated genes. We tried to identify SNVs from long-read data and compared the results with our previous study11 (online supplemental table S3). Long-read sequencing identified a much larger number of SNVs than short-read sequencing, suggesting a higher error rate in the former. Therefore, we did not take SNVs into consideration in the current study.

To validate the identified SVs, we amplified the SV junction regions by PCR using KOD Multi&EpEnzyme (TOYOBO), and amplicons were subjected to Sanger sequencing.

To examine the frequencies of the SVs in the independent patient group (n=1189),11 we performed multiplex PCR-based targeted sequencing. Three types of PCR primer sets were designed for each of the identified SVs to amplify the upstream breakpoint regions, downstream breakpoint regions and deleted regions of the large deletions (online supplemental figure 1 and table S4). The details of the sequencing method have been described previously.18

RESULTS

We selected 15 patients with RP with one pathogenic variant in EYS from our previous study11 and confirmed that all of them had typical RP. We performed whole-genome sequencing with the Oxford Nanopore long-read sequencer. The average number of reads and their lengths were 6623354.7 and 8136.7 bp, respectively (online supplemental figure 2, table S1), and 93.7% of the reads were aligned to the human reference genome (GRCh38) (online supplemental table S5). The mean depth of the EYS region was 16.5. The most the EYS region (94.1%) was covered by > 10 reads (online supplemental figure 3, table S6).

We confirmed that the pathogenic variants of EYS detected in our previous targeted resequencing study11 could be found by long-read sequencing all patients (online supplemental figure 4, table S7). We then focused on identified SVs. In the 15 patients with RP sequenced, 46899 SVs consisting of 22786 deletions, 22790 insertions, 89 inversions, 765 duplications and 469 translocations were identified across the genome. We observed 176 exon-overlapping SVs per patient on average, which included
## Structural variation

**Table 1** Summary of SVs identified in RP-associated genes

<table>
<thead>
<tr>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Length (bp)</th>
<th>Type of SV</th>
<th>Gene</th>
<th>Overlap with coding regions of RP-associated genes</th>
<th>Distance (bp)*</th>
<th>Frequency in in-house data</th>
<th>Frequency in gnomAD</th>
<th>Frequency in dbVar</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr6</td>
<td>63,957,115</td>
<td>63,958,454</td>
<td>1340</td>
<td>Deletion</td>
<td>EYS</td>
<td>+</td>
<td>–</td>
<td>14%</td>
<td>–</td>
<td>–</td>
<td>OPH17, OPH217, OPH641, OPH690, OPH693, OPH861, OPH556</td>
</tr>
<tr>
<td>chr6</td>
<td>65,001,113</td>
<td>65,005,820</td>
<td>4718</td>
<td>Deletion</td>
<td>EYS</td>
<td>+</td>
<td>–</td>
<td>21%</td>
<td>–</td>
<td>–</td>
<td>OPH17, OPH217, OPH690, OPH693, OPH556</td>
</tr>
<tr>
<td>chr6</td>
<td>65,350,144</td>
<td>65,352,738</td>
<td>1995</td>
<td>Deletion</td>
<td>EYS</td>
<td>+</td>
<td>–</td>
<td>81%</td>
<td>–</td>
<td>–</td>
<td>All samples</td>
</tr>
<tr>
<td>chr6</td>
<td>65,683,153</td>
<td>65,694,734</td>
<td>562</td>
<td>Deletion</td>
<td>EYS</td>
<td>+</td>
<td>–</td>
<td>31%</td>
<td>–</td>
<td>–</td>
<td>All samples</td>
</tr>
<tr>
<td>chr6</td>
<td>63,942,754</td>
<td>64,337,844</td>
<td>395,091</td>
<td>Deletion</td>
<td>EYS</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>OPH661</td>
</tr>
<tr>
<td>chr6</td>
<td>64,423,168</td>
<td>64,798,962</td>
<td>375,795</td>
<td>Deletion</td>
<td>EYS</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>OPH661</td>
</tr>
<tr>
<td>chr6</td>
<td>65,454,074</td>
<td>64,504,754</td>
<td>305</td>
<td>Inversion</td>
<td>EYS</td>
<td>+</td>
<td>–</td>
<td>17%</td>
<td>–</td>
<td>–</td>
<td>OPH25, OPH31, OPH64, OPH66, OPH693, OPH861, OPH661, OPH663</td>
</tr>
<tr>
<td>chr6</td>
<td>64,296,539</td>
<td>64,296,332</td>
<td>94</td>
<td>Deletion</td>
<td>EYS</td>
<td>–</td>
<td>66</td>
<td>–</td>
<td>56%</td>
<td>–</td>
<td>OPH666, OPH663, OPH31, OPH64, OPH66, OPH661, OPH663</td>
</tr>
<tr>
<td>chr6</td>
<td>65,204,982</td>
<td>65,205,044</td>
<td>63</td>
<td>Deletion</td>
<td>EYS</td>
<td>–</td>
<td>157</td>
<td>–</td>
<td>2.4%</td>
<td>–</td>
<td>OPH664, OPH31, OPH64, OPH66, OPH661</td>
</tr>
<tr>
<td>chr6</td>
<td>65,564,961</td>
<td>65,565,284</td>
<td>324</td>
<td>Deletion</td>
<td>EYS</td>
<td>–</td>
<td>154</td>
<td>80%</td>
<td>–</td>
<td>–</td>
<td>All samples</td>
</tr>
<tr>
<td>chr6</td>
<td>64,295,413</td>
<td>64,296,413</td>
<td>110</td>
<td>Inversion</td>
<td>EYS</td>
<td>–</td>
<td>148</td>
<td>2%</td>
<td>–</td>
<td>–</td>
<td>All samples</td>
</tr>
<tr>
<td>chr6</td>
<td>65,278,329</td>
<td>65,278,329</td>
<td>59</td>
<td>Inversion</td>
<td>EYS</td>
<td>–</td>
<td>59</td>
<td>–</td>
<td>2.7%</td>
<td>–</td>
<td>All samples</td>
</tr>
<tr>
<td>chr8</td>
<td>10,607,821</td>
<td>10,608,503</td>
<td>682</td>
<td>Deletion</td>
<td>RP1L1</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>OPH690</td>
</tr>
<tr>
<td>chr8</td>
<td>10,610,165</td>
<td>10,610,165</td>
<td>3</td>
<td>Inversion</td>
<td>RP1L1</td>
<td>+</td>
<td>–</td>
<td>15%</td>
<td>–</td>
<td>–</td>
<td>All samples</td>
</tr>
<tr>
<td>chr19</td>
<td>7,450,130</td>
<td>7,451,150</td>
<td>1021</td>
<td>Duplication</td>
<td>ARHGEF7</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>7.2%</td>
<td>–</td>
<td>OPH661, OPH31, OPH66, OPH661, OPH663, OPH690</td>
</tr>
</tbody>
</table>

*The shortest distance between the exon boundary and SV breakpoint.

RP, retinitis pigmentosa; SV, structural variant.
In this study, we identified likely pathogenic SVs in two previously unsolved RP cases by long-read sequencing. The proportion of patients with a new molecular diagnosis was 2 out of 15, which is consistent with previous reports using short-read whole-genome sequencing (4.8%–12.5%). According to the ACMG guidelines, the PVS1 and PM2 criteria were applied to these SVs, and they were considered likely pathogenic. We also confirmed the pathogenicity of SNVs identified in our previous study (Supplementary Note). Considering that two causative variants were present in the ARRP causative gene, we regarded EYS as the causative gene for both OPH641 and OPH861. Clinical information for these two patients is provided in the Supplemental material (online supplemental figure 5). Both probands were sporadic cases according to the medical interviews. Although we were unable to investigate the segregation of the SVs with phenotypes in the pedigrees, this family history is consistent with the ARRP gene (EYS) being the causative gene. For the other thirteen candidate SVs for the ARRP gene, three control data (in-house, gnomAD SVs, dbVar) were used to prioritise variants by AF. A 683 bp heterozygous deletion in RP1L1 found in OPH690 remained, and the other 12 candidates were excluded. According to the ACMG guidelines, the PVS1 and PM2 criteria were applied to this SV, and it was considered likely pathogenic. However, the other pathogenic variants in RP1L1 were not detected in OPH690 by our previous targeted resequencing (online supplemental table S9). Therefore, we did not consider RP1L1 as the causative gene for this patient.

Finally, we examined the frequencies of the two large SVs in EYS in other patients with RP. We conducted multiplex PCR-based targeted sequencing to evaluate these SVs in 1189 independent Japanese patients with RP. However, the two SVs were not detected, suggesting that they are not founder mutations.

**DISCUSSION**

In this study, we identified likely pathogenic SVs in two previously unsolved RP cases by long-read sequencing. The proportion of patients with a new molecular diagnosis was 2 out of 15, which is consistent with previous reports using short-read whole-genome sequencing (4.8%–12.5%).

Short-read and long-read sequencing can be applied to whole-genome sequencing. Although the former has the advantage of detecting SVs and short indels, functional experiments are generally required to assess the causality of intronic variants. On
the other hand, long-read sequencing may identify pathogenic SVs, such as exon-overlapping SVs. Indeed, our study identified two likely pathogenic SVs. Although the utility of long-read sequencing for Mendelian diseases is not well established, this study indicates that long-read sequencing contributes to the molecular diagnosis of Mendelian diseases. Our previous and current studies have comprehensively investigated SNVs and short indels in the coding regions and SVs in EYS for 13 cases. However, 87% (13/15) of the cases still remain unsolved. Considering that variants in coding regions were well investigated with sufficient coverage (online supplemental table S6), variants in non-coding regions could be the cause for these patients.

This study did not analyse intronic SNVs for three reasons: (1) pathogenic deep intronic EYS variants have not been reported previously, (2) the pathogenicity of intronic variants requires functional validation and (3) long-read sequencing does not have sufficient accuracy for SNV identification. While long-read sequencing has contributed to molecular diagnosis in some cases, an inadequate search for variants in deep introns may result in the low detection rate of pathogenic variants. Therefore, further evaluation of non-coding regions in patients with unresolved RP by short-read sequencing and functional validation should increase the number of genetically solved patients. Additionally, an analysis of other candidate genes is recommended.

Another limitation of this study is that the haplotypes of the pathogenic variants in the patients could not be examined. A segregation analysis of the patients observed in the present study will further clarify the impact of the identified SVs.

In conclusion, we identified likely pathogenic SVs in an ARRP gene by long-read sequencing. Our results imply that searching for SVs and the comprehensive evaluation of non-coding regions in genetically unsolved cases will contribute to the molecular diagnosis of RP.

Author affiliations
1Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Fukuoka, Japan
2Department of Human Genetics, The University of Tokyo, Graduate School of Medicine, Bunkyo-ku, Tokyo, Japan
3Institute for Statistical Analysis and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan
4Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan
5Department of Ophthalmology, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan
6Department of Advanced Ophthalmic Medicine, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan
7Yuko WADA Eye Clinic, Sendai, Japan
8Department of Ophthalmology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan
9Department of Ophthalmology, Juntendo University Graduate School of Medicine, Bunkyo-ku, Tokyo, Japan
10Department of Ophthalmology, Hamamatsu University School of Medicine, Hamamatsu, Shizuoka, Japan
11Department of Ophthalmology, Faculty of Medicine, University of Miyazaki, Miyazaki, Miyazaki, Japan
12Department of Ocular Pathology and Imaging Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

Acknowledgements
The super-computing resource was provided by the Human Genome Centre, Institute of Medical Science, the University of Tokyo.

Contributors
MA and AF designed the study. YS performed long-read sequencing. YS, YK, and AF analysed the data. YS, MA, and AF contributed to data interpretation. ME, TA, and Y Momoza made Multiplex PCR-based target sequencing. YK, Y Murakami, KE, KH, TN, YY, SU, DG, AM, YH, YI, KMN and KHS collected the samples. YS, JHW, MA and AF contributed to the manuscript preparation and editing. All authors approved the final manuscript.

Funding
This research was supported by AMED under Grant Number JP21km09000101 (A.F.) and by the Japanese Retinitis Pigmentosa Society (M.A.).

Competing interests
None declared.

Patient consent for publication
Obtained.

Ethics approval
The ethics committees of Kyushu University Hospital, the University of Tokyo, and all collaborating hospitals have approved this study. All patients have provided written consent to participate in the study. This study was conducted following the principles of the Declaration of Helsinki.

Provenance and peer review
Not commissioned; externally peer reviewed.

Supplemental material
This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Yusuke Sano http://orcid.org/0000-0003-2023-7630
Yoshito Koyanagi http://orcid.org/0000-0001-5559-6305
Akihiro Fujimoto http://orcid.org/0000-0002-0075-0800

REFERENCES
Structural variation


20 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 2013;14:486.


