Supplementary Information

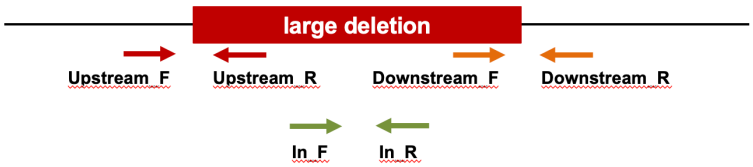
Identification of pathogenic structural variants in genetically unsolved patients with retinitis pigmentosa by long-read sequencing

Yusuke Sano, Yoshito Koyanagi, Jing Hao Wong, Yusuke Murakami, Kohta Fujiwara, Mikiko Endo, Tomomi Aoi, Kazuki Hashimoto, Toru Nakazawa, Yuko Wada, Shinji Ueno, Dan Gao, Akira Murakami, Yoshihiro Hotta, Yasuhiro Ikeda, Koji M Nishiguchi, Yukihide Momozawa, Koh-Hei Sonoda, Masato Akiyama, and Akihiro Fujimoto

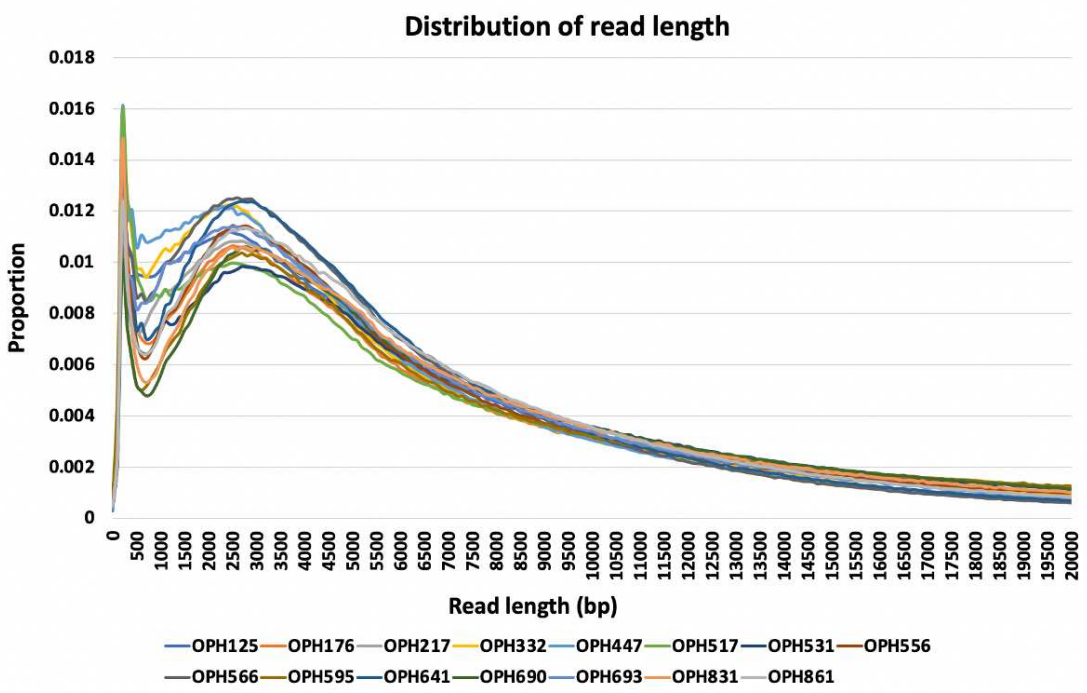
Supplementary Figure 1: Schematic diagram of PCR primer sets for short-read sequencing
Supplementary Figure 2: Distribution of read length
Supplementary Figure 3: Distribution of read depth in *EYS* region
Supplementary Figure 4: IGV visualization of pathogenic variants detected in our previous study
Supplementary Figure 5: Clinical features of the two EYS-RP patients with large SVs

Supplementary Table 1: Summary of whole-genome sequencing
Supplementary Table 2: 89 genes associated with RP
Supplementary Table 3: Comparison of the number of SNVs called by short-read sequencing and long-read sequencing in the *EYS* region
Supplementary Table 4: Primer information
Supplementary Table 5: Summary of mapping
Supplementary Table 6: Mean depth and coverage rate in *EYS* region
Supplementary Table 7: Pathogenic variants detected in previous study
Supplementary Table 8: Number of exon-overlapping SVs by class
Supplementary Table 9: *RP1L1* variants in OPH690 detected using the targeted approach in our previous study.

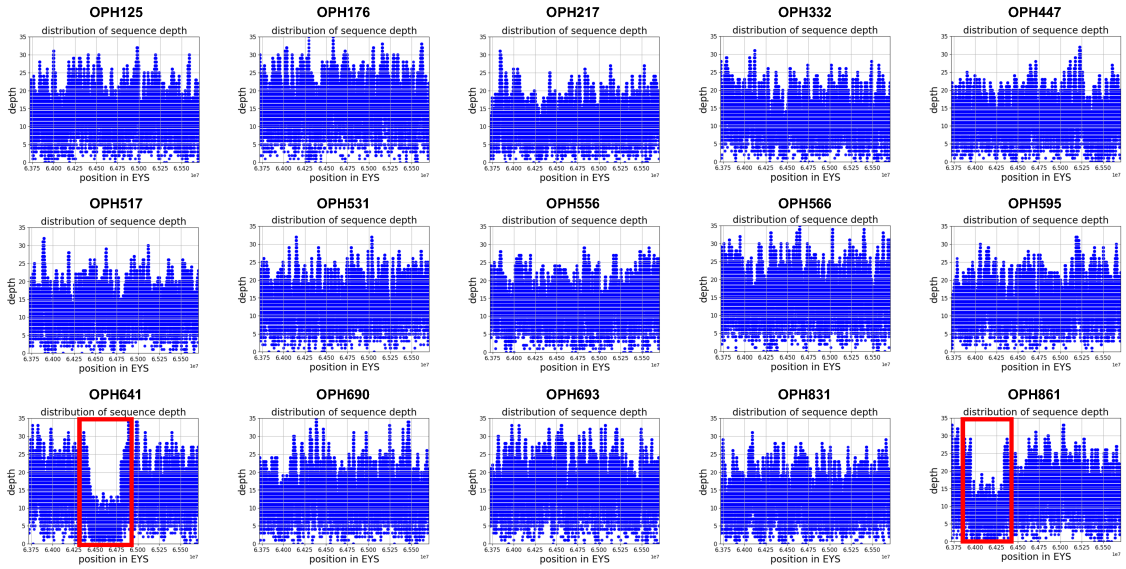
Supplementary note: The adaptation of the ACMG guidelines for heterozygous pathogenic SNVs identified in our previous study



Supplementary Figure 1
Schematic diagram of PCR primer sets for short-read sequencing. Primer sets: flanking the upstream breakpoint (Upstream_ F and Upstream_R), flanking the downstream breakpoint (Downstream_ F and Downstream_R), and amplifying the inside region of the large deletion (In_ F and In_R) were designed. In_ F and In_R are designed not to react with Downstream_R and Upstream_ F, respectively. The pair of Upstream_F and Downstream_R reacts only in the presence of the deletion, when the product length becomes suitable for amplification.

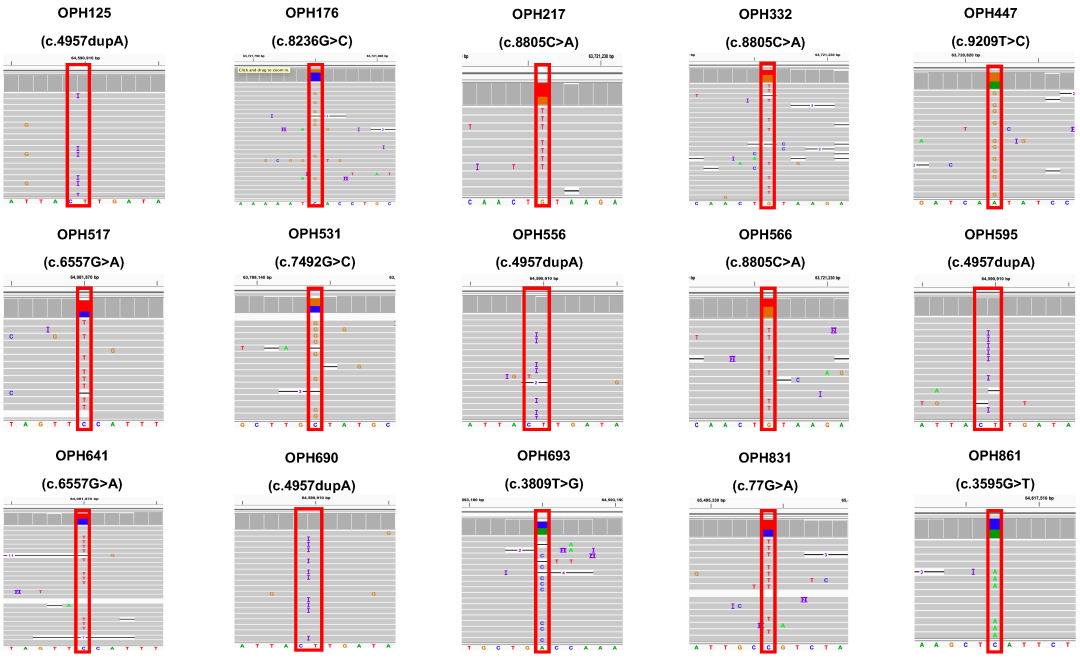


Supplementary Figure 2
Distribution of read length. A large proportion of the reads were 2.7 kb in length; the average read length of the 15 RP cases was 8.1 kb.



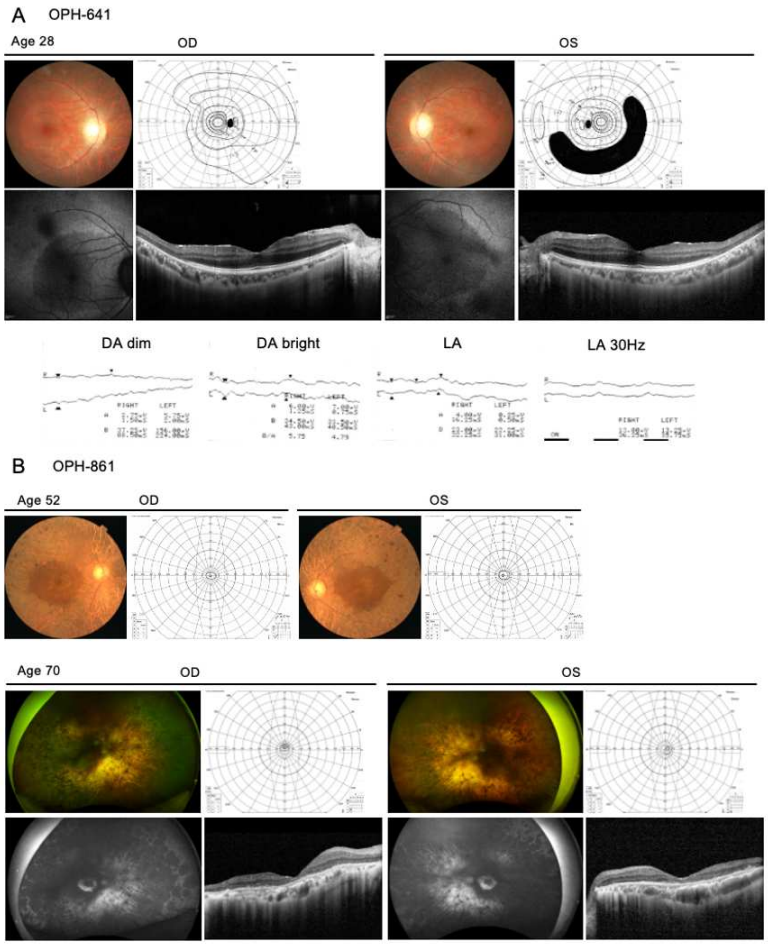
Supplementary Figure 3

Distribution of read depth in *EYS* region. The sequence depth at each locus in the *EYS* region is plotted for each case. The red rectangles indicate the large deletion regions identified in this study.



Supplementary Figure 4

IGV visualization of pathogenic variants detected in our previous study. The *EYS* pathogenic variants detected by the previous targeted short-read sequencing study were also successfully identified by the long-reads. The red rectangles indicate the pathogenic SNVs and short indels.



Supplementary Figure 5
Clinical features of the two *EYS*-RP patients with large SVs.
A. Color fundus photographs (CFP), fundus autofluorescence (FAF), Goldmann visual fields (GVF), optical coherence tomography (OCT), and electroretinogram (ERG) of OPH-641 at age 28. OD - Right eye, OS - Left eye. Visual acuity was 20/20 in both eyes. In CFP, mild retinal degeneration associated with attenuated retinal vessels were observed along the vascular arcades. FAF exhibited the hyperautofluorescent ring surrounding the relatively preserved macula. Horizontal OCT images showed disrupted ellipsoid zone in the temporal macula. Mild visual field constriction (OD) and pericentral ring scotoma (OS) were observed with GVF. ERG demonstrated non-recordable rod response and severely impaired cone response.
B. CFP and GP at age 52 (upper column) and ultra-wide field CFP/FAF, GVF, and OCT at age 70 (lower column) of OPH-861. OD - Right eye, OS - Left eye. At age, 52, visual acuity was 20/2000 in both eyes. CPF showed severe retinal degeneration associated with bone spicule pigmentation extending from the vascular arcades to the peripheral retina. Granular pigmentary changes were also observed in the macula. GVF showed severe concentric contraction. At age 70, visual acuity was dropped to hand motion in both eyes. The eye examinations demonstrated massive retinal degeneration involving the macula and further constriction of central vision.

Supplementary table 1: Summary of whole-genome sequencing

Sample ID	Number of runs	Total number of reads	Mean read length (bp)	Total yield (Gb)
OPH125	2	6,871,891	7,680	53.1
OPH176	3	6,420,990	9,470	61.1
OPH217	2	5,661,856	7,900	45
OPH332	2	6,848,271	7,200	49.7
OPH447	2	6,714,364	7,050	47.7
OPH517	2	5,732,203	8,730	50.3
OPH531	2	5,556,967	9,050	50.6
OPH556	2	6,244,530	8,300	52.2
OPH566	2	9,063,508	6,580	60.1
OPH595	2	5,265,214	10,030	53.1
OPH641	2	7,984,124	7,040	56.6
OPH690	2	5,927,670	9,160	54.6
OPH693	2	7,701,357	7,520	58.3
OPH831	2	5,937,667	8,630	51.5
OPH861	2	7,419,708	7,710	57.5
Mean	2.1	6,623,355	8,137	53.4

Supplementary table 2: 89 genes associated with RP

causative genes	inheritance pattern
ADIPOR1	autosomal dominant
ARL3	autosomal dominant
CRX	autosomal dominant
FSCN2	autosomal dominant
GUCA1B	autosomal dominant
HK1	autosomal dominant
IMPDH1	autosomal dominant
IMPG1	autosomal dominant
KLHL7	autosomal dominant
PRPF3	autosomal dominant
PRPF4	autosomal dominant
PRPF6	autosomal dominant
PRPF8	autosomal dominant
PRPF31	autosomal dominant
PRPH2	autosomal dominant
RDH12	autosomal dominant
ROM1	autosomal dominant
RP9	autosomal dominant
RP17	autosomal dominant
SEMA4A	autosomal dominant
SNRNP200	autosomal dominant
SPP2	autosomal dominant
TOPORS	autosomal dominant
BEST1	autosomal dominant/recessive
NR2E3	autosomal dominant/recessive
NRL	autosomal dominant/recessive
RHO	autosomal dominant/recessive
RP1	autosomal dominant/recessive
RPE65	autosomal dominant/recessive
SAG	autosomal dominant/recessive
ABCA4	autosomal recessive
AGBL5	autosomal recessive
AHR	autosomal recessive
ARHGEF18	autosomal recessive

ARL6	autosomal recessive
ARL2BP	autosomal recessive
BBS1	autosomal recessive
BBS2	autosomal recessive
C2orf71	autosomal recessive
C8orf37	autosomal recessive
CERKL	autosomal recessive
CLCC1	autosomal recessive
CLRN1	autosomal recessive
CNGA1	autosomal recessive
CNGB1	autosomal recessive
CRB1	autosomal recessive
CYP4V2	autosomal recessive
DHDDS	autosomal recessive
DHX38	autosomal recessive
EMC1	autosomal recessive
EYS	autosomal recessive
FAM161A	autosomal recessive
GPR125	autosomal recessive
HGSNAT	autosomal recessive
IDH3B	autosomal recessive
IFT140	autosomal recessive
IFT172	autosomal recessive
IMPG2	autosomal recessive
KIAA1549	autosomal recessive
KIZ	autosomal recessive
LRAT	autosomal recessive
MAK	autosomal recessive
MERTK	autosomal recessive
MVK	autosomal recessive
NEK2	autosomal recessive
NEUROD1	autosomal recessive
PDE6A	autosomal recessive
PDE6B	autosomal recessive
PDE6G	autosomal recessive
POMGNT1	autosomal recessive

PRCD	autosomal recessive
PROM1	autosomal recessive
RBP3	autosomal recessive
REEP6	autosomal recessive
RGR	autosomal recessive
RLBP1	autosomal recessive
RP1L1	autosomal recessive
SAMD11	autosomal recessive
SLC7A14	autosomal recessive
SPATA7	autosomal recessive
TRNT1	autosomal recessive
TTC8	autosomal recessive
TULP1	autosomal recessive
USH2A	autosomal recessive
ZNF408	autosomal recessive
ZNF513	autosomal recessive
OFD1	X-linked
RP2	X-linked
RPGR	X-linked

Supplementary table 3: Comparison of the number of SNVs called by short-read sequencing and long-read sequencing in the *EYS* region

sample	SNV called by LRS in EYS region	SNV called by LRS in non-coding region of EYS	SNV called by LRS in coding region of EYS	SNV called by SRS in coding region of EYS	SNVs matched by LRS and SRS	Proportion of SNVs called by SRS and LRS among those called by LRS
OPH_125	94166	93739	427	23	19	0.044
OPH_176	97756	97103	653	21	19	0.029
OPH_217	108392	107422	970	11	9	0.009
OPH_332	109816	109217	599	6	5	0.008
OPH_447	97003	96391	612	23	22	0.036
OPH_517	103333	102815	518	12	10	0.019
OPH_531	88725	88360	365	13	11	0.03
OPH_556	95270	94829	441	19	17	0.039
OPH_566	95073	94651	422	21	20	0.047
OPH_595	85435	84876	559	22	19	0.034
OPH_641	125153	124444	709	9	6	0.008
OPH_690	101482	101020	462	22	21	0.045
OPH_693	100287	99762	525	13	10	0.019
OPH_831	107743	107194	549	13	11	0.02
OPH_861	114873	114368	505	22	21	0.042
mean	101633.8	101079.4	554.4	16.7	14.7	0.029

Supplementary table 4: Primer information

OPH641

	forward	reverse
Primer sequences designed to span the upstream breakpoint (chr6: 64,423,168)	TTACCCCAGCCTAAGTAAAGTAGTC	AAGTGAGGGAGTTAAATGGTCA
Primer sequences designed to span the downstream breakpoint (chr6: 64,798,957)	ACTGATTTTTCTGCAAAGCA	GCATGTGACATCTTACTTGAAATGA
Primer sequences designed entirely inside the deletion	TTCACGAGAGGATTTTATTTCAG	GGCTCAAGAGGATATTGAAGAAT

OPH861

	forward	reverse
Primer sequences designed to span the upstream breakpoint (chr6: 63,942,752)	CATCTTAATCCCAAAACCAGGA	ACTGAATTTTGCCAAATACTTTCT
Primer sequences designed to span the downstream breakpoint (chr6: 64,337,822)	CAGTGGAAGAATAATTGGTACC	TGAAAGTGATCAGAATCAAGG
Primer sequences designed entirely inside the deletion	CAGCTCAGATCCTGTTTGTAAGT	TCCTATTTAGAACTGCCCTTTTT

Supplementary table 5: Summary of mapping

Sample ID	Number of mapped reads	Number of unmapped reads	Number of split reads	Mapping rate	Total length of aligned bases (Gb)	Proportion of aligned bases
OPH125	6364455	507436	687222	0.926	50	0.931
OPH176	5978972	442018	719595	0.931	58.1	0.93
OPH217	5342357	319499	603280	0.944	43.1	0.932
OPH332	6410503	437768	692553	0.936	47.5	0.925
OPH447	6205007	509357	759909	0.924	44.8	0.929
OPH517	5322419	409784	656072	0.929	47.7	0.927
OPH531	5241981	314986	599672	0.943	48.6	0.935
OPH556	5903551	340979	688175	0.945	50.2	0.93
OPH566	8581230	482278	1030869	0.947	57.6	0.929
OPH595	5002027	263187	650390	0.950	51.2	0.935
OPH641	7549073	435051	933530	0.946	54	0.926
OPH690	5578458	349212	666428	0.941	52.2	0.929
OPH693	7211072	490285	859454	0.936	55.6	0.927
OPH831	5490637	447030	635377	0.925	48.6	0.928
OPH861	6925489	494219	817663	0.933	54.8	0.929
Mean	6207202	416206	733346	0.937	50.9	0.929

Supplementary table 6: Mean depth and coverage rate in *EYS* region

sample	mean depth (LRS)	coverage with sequence depth >= 1 (LRS)	coverage with sequence depth >= 5 (LRS)	coverage with sequence depth >= 10 (LRS)	coverage with sequence depth >= 15 (LRS)	coverage with sequence depth >= 20 (previous study)
OPH125	16.475	0.999	0.996	0.963	0.664	1
OPH176	19.293	0.999	0.999	0.988	0.869	1
OPH217	14.000	0.999	0.996	0.875	0.435	1
OPH332	14.984	0.996	0.995	0.920	0.522	1
OPH447	15.028	0.999	0.998	0.928	0.550	1
OPH517	15.455	0.999	0.998	0.951	0.586	1
OPH531	16.413	1.000	1.000	0.967	0.682	1
OPH556	16.241	0.998	0.998	0.963	0.652	1
OPH566	18.544	0.999	0.999	0.985	0.825	0.998
OPH595	17.028	1.000	0.999	0.972	0.706	1
OPH641	16.404	0.996	0.979	0.837	0.665	1
OPH690	16.714	0.999	0.998	0.964	0.666	1
OPH693	18.214	0.994	0.993	0.976	0.784	1
OPH831	15.879	1.000	0.999	0.938	0.615	1
OPH861	16.262	0.998	0.997	0.893	0.638	0.998
Mean	16.462	0.998	0.996	0.941	0.657	1.000

LRS: long-read sequencing

The previous study performed short-read sequencing.

Supplementary table 7: Pathogenic variants detected in previous study

sample	Causative gene	Chromosome	Position (bp)	Nucleotide change	Amino acid change
OPH125	*EYS*	6	64,590,909	c.4957dupA	p.(Ser1653fs)
OPH176	*EYS*	6	63,721,795	c.8236G>C	p.(Asp2746His)
OPH217	*EYS*	6	63,721,226	c.8805C>A	p.(Tyr2935*)
OPH332	*EYS*	6	63,721,226	c.8805C>A	p.(Tyr2935*)
OPH447	*EYS*	6	63,720,822	c.9209T>C	p.(Ile3070Thr)
OPH517	*EYS*	6	64,081,870	c.6557G>A	p.(Gly2186Glu)
OPH531	*EYS*	6	63,789,144	c.7492G>C	p.(Ala2498Pro)
OPH556	*EYS*	6	64,590,909	c.4957dupA	p.(Ser1653fs)
OPH566	*EYS*	6	63,721,226	c.8805C>A	p.(Tyr2935*)
OPH595	*EYS*	6	64,590,909	c.4957dupA	p.(Ser1653fs)
OPH641	*EYS*	6	64,081,870	c.6557G>A	p.(Gly2186Glu)
OPH690	*EYS*	6	64,590,909	c.4957dupA	p.(Ser1653fs)
OPH693	*EYS*	6	64,593,185	c.3809T>G	p.(Val1270Gly)
OPH831	*EYS*	6	65,495,334	c.77G>A	p.(Arg26Gln)
OPH861	*EYS*	6	64,617,507	c.3595G>T	p.(Glu1199*)

Supplementary table 8: Number of exon-overlapping SVs by class

sample	deletion	insertion	inversion	duplication	translocation
OPH125	108	44	10	10	11
OPH176	116	45	8	10	7
OPH217	84	54	6	10	11
OPH332	97	37	7	8	9
OPH447	84	40	6	8	8
OPH517	105	39	7	7	9
OPH531	100	38	6	10	10
OPH556	95	45	2	10	15
OPH566	111	39	6	13	13
OPH595	104	51	8	10	12
OPH641	117	49	6	10	9
OPH690	101	44	8	11	20
OPH693	94	42	6	10	11
OPH831	106	42	5	7	12
OPH861	116	53	8	12	16
Mean	102.5	44.1	6.6	9.7	11.53

Supplementary table 9: *RP1L1* variants in OPH-690 detected using the targeted approach in our previous study.

Chr.	Position (bp)*	rsID	REF	ALT	Reference sequence	Nucleotide change	Amino acid change	Type of variant	Zygosity	MAF (gnomAD)	ClinVar	HDMG	SIFT	PolyPhen	CADD	ACMG guideline
8	10607245	rs55642448	C	T	NM_178857	c.G6853A	p.G2285R	missense variant	Het	64055/ 151948	Benign	Notregistered	deleterious	benign	16.43	Benign
8	10607375	rs56382513	T	C	NM_178857	c.6723A>G	p.S2241=	synonymous variant	Homo	85188/ 151910	Benign	Notregistered	-	-	0.058	Benign
8	10608238	rs11783478	T	C	NM_178857	c.5860A>G	p.T1954A	missense variant	Homo	80928/ 143550	Benign	Notregistered	tolerated low confidence	benign	0.102	Benign
8	10608432	rs28446662	T	A	NM_178857	c.5666A>T	p.D1889V	missense variant	Homo	33948/ 148070	Benign	Notregistered	tolerated	benign	0.34	Benign
8	10608480	rs200622636	T	C	NM_178857	c.5618A>G	p.D1873G	missense variant	Het	1/ 144422	Benign	Notregistered	tolerated	benign	0.024	Likely Benign
8	10609614	rs4841399	G	C	NM_178857	c.4484C>G	p.P1495R	missense variant	Het	42430/ 152026	Benign	Notregistered	tolerated	benign	0.258	Benign
8	10610117	rs112656102	T	C	NM_178857	c.3981A>G	p.T1327=	synonymous variant	Het	11508/ 140004	Benign	Notregistered	-	-	1.59	Benign
8	10610119	rs146656804	T	TTTC	NM_178857	c.3978_3979insGAA	p.T1327_E1328insE	inframe insertion	Het	14837/ 74880	Benign	Notregistered	-	-	1.73	Benign
8	10610127	rs4240659	T	C	NM_178857	c.3971A>G	p.E1324G	missense variant	Het	19483/ 119966	Conflicting interpretations of pathogenicity	Notregistered	tolerated	benign	1.52	Benign
8	10610142	rs4840501	G	C	NM_178857	c.3956C>G	p.A1319G	missense variant	Het	35483/ 87176	Benign	Notregistered	tolerated	benign	0.046	Benign
8	10616532	rs4388421	T	G	NM_178857	c.665A>C	p.H222P	missense variant	Homo	27432/ 152082	Benign	Notregistered	deleterious	possibly damaging	22.6	Benign

* Chromosomes are based on Human GRCh38. REF, reference allele; ALT, alternative allele; gnomAD, Genome Aggregation Database; MAF, Minor allele frequency.

Supplementary note: The adaptation of the ACMG guidelines for heterozygous pathogenic SNVs identified in our previous study

In OPH641, according to the ACMG guidelines, the PM2, PP1, PP3, PP4, and PP5 criteria were applied to p.(Gly2186Glu) (**online supplemental table S7**) identified in previous study because it was reported to be infrequent in 1000 Genomes control data, segregated with the phenotype,[1] the pathogenicity is supported by computational evidence such as polyphen and SIFT, the phenotype of the patient is consistent with RP, which is Mendelian disease, and that it is considered pathogenic by ClinVar. Therefore, the SNV was considered likely pathogenic. In OPH861, according to the ACMG guidelines, the PVS1, PM2 and PM6 criteria were applied to p.(Glu1199*) (**online supplemental table S7**) identified in the previous study because it is a nonsense variant, have a low frequency in the 1000 Genomes control data, has not been registered in the SNV database, and is considered a de novo variant. Therefore, the SNV was considered pathogenic.

Sample	Gene	Variant	Pathogenicity	Evidence criteria	Reference
OPH641	*EYS*	p.(Gly2186Glu)	Likely Pathogenic	PM2	1000 Genomes (https://www.internationalgenome.org/)
				PP1	Iwanami M. et al. *Mol Vis.* 2019
				PP3	PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2), SIFT (http://sift.jcvi.org)
				PP4	online supplemental figure 5
				PP5	ClinVar (http://www.ncbi.nlm.nih.gov/clinvar)
OPH861	*EYS*	p.(Glu1199*)	Pathogenic	PVS1	ANNOVAR (V.3.4) (https://annovar.openbioinformatics.org/en/latest/)
				PM2	1000 Genomes (https://www.internationalgenome.org/)
				PM6	dbSNP (http://www.ncbi.nlm.nih.gov/snp)

REFERENCE

1 Iwanami M, Oishi A, Ogino K, Seko Y, Nishida-Shimizu T, Yoshimura N, Kato S. Five major sequence variants and copy number variants in the EYS gene account for one-third of Japanese patients with autosomal recessive and simplex retinitis pigmentosa. *Mol Vis* 2019;**25**:766–79.