Analysis of the *PRPF31* Gene in Spanish Autosomal Dominant Retinitis Pigmentosa Patients: A Novel Genomic Rearrangement

Inmaculada Martin-Merida,[1,2] Rocio Sanchez-Alcudia,[1,2] Patricia Fernandez-San Jose,[1,2] Fiona Blanco-Kelly,[1,2] Raquel Perez-Carro,[1,2] Luciana Rodriguez-Jacy da Silva,[3] Berta Almoguera,[4] Blanca Garcia-Sandoval,[5] Maria Isabel Lopez-Molina,[5] Almudena Avila-Fernandez,[1,2] Miguel Carballo,[6] Marta Corton,[1,2] and Carmen Ayuso[1,2]

[1]Department of Genetics, Instituto de Investigacion Sanitaria–Fundacion Jimenez Diaz University Hospital-Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain
[2]Center for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain
[3]Universidade de Mogi das Cruzes, São Paulo, Brazil
[4]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States
[5]Department of Ophthalmology, Instituto de Investigacion Sanitaria–Fundacion Jimenez Diaz University Hospital–Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain
[6]Molecular Genetics Unit, Hospital de Terrassa, Terrassa, Barcelona, Spain

Correspondence: Carmen Ayuso, Department of Genetics, Instituto de Investigacion Sanitaria-Fundacion Jimenez Diaz University Hospital-Universidad Autonoma de Madrid (IIS-FJD, UAM), Avenue Reyes Católicos No. 2, Madrid 28040, Spain;
cayuso@fjd.es.

MCo and CA contributed equally to the work presented here and should therefore be regarded as equivalent authors.

Submitted: August 12, 2016
Accepted: January 15, 2017

Citation: Martin-Merida I, Sanchez-Alcudia R, Fernandez-San Jose P, et al. Analysis of the *PRPF31* gene in Spanish autosomal dominant retinitis pigmentosa patients: a novel genomic rearrangement. *Invest Ophthalmol Vis Sci*. 2017;58:1045–1053. DOI: 10.1167/iovs.16-20515

Purpose. The aim was to determine the prevalence of *PRPF31* mutations in a cohort of Spanish autosomal dominant retinitis pigmentosa (adRP) families to deepen knowledge of the pathogenic mechanisms underlying the disease and to assess genotype–phenotype correlations.

Methods. A cohort of 211 adRP patients was screened for variants in *PRPF31* by using a combined strategy comprising next-generation sequencing approaches and copy-number variation (CNV) analysis. Quantitative RT-PCR and CNV analysis of the regulatory MSR1 element were also performed to assess *PRPF31* gene expression. Phenotype was assessed by using ophthalmologic examination protocols.

Results. Fifteen different causative mutations and genomic rearrangements were identified, revealing five novel mutations. Prevalence of *PRPF31* mutations, genomic rearrangements, and lack of penetrance were 7.6%, 1.9%, and 66.7%, respectively. Interestingly, we identified a tandem duplication and a partial *PRPF31* deletion in different affected individuals from the same family. *PRPF31* gene expression was significantly decreased in symptomatic cases carrying either *PRPF31* duplication or deletion as compared to controls. The 4 MSR1 allele in *cis* with the *PRPF31* wild-type allele was apparently a protective factor. The mutated phenotype varied from no symptoms to typical retinitis pigmentosa with variable onset and course depending on the kind of mutation, with the duplication case the most severe.

Conclusions. In view of the high genetic heterogeneity of *PRPF31* mutations, the screening must include the entire gene, as well as CNV assays, to detect large rearrangements. This is the first report of a variable phenotype correlation as well as a gross duplication and deletion within the same family.

Keywords: *PRPF31*, adRP, CNV, NGS, MSR1

Inherited retinal dystrophies (RDs) are a group of rare diseases, most of which are degenerative and progressive as a result of photoreceptor cell death. The most common form of RD is retinitis pigmentosa (RP), with a prevalence of 1/4000.[1] Retinitis pigmentosa is characterized by rod photoreceptor degeneration causing night blindness, progressive reduction of visual field, and pigment accumulation in the peripheral retina detected on fundus examination at early stages of the disease.[1,2] Retinitis pigmentosa shows great clinical and genetic heterogeneity, presenting autosomal recessive (35%–50%), autosomal dominant (15%–25%), or X-linked (7%–15%) inheritance patterns, making molecular diagnosis very complex.[2] Approximately 15% of Spanish families with RP present autosomal dominant RP (adRP).[3] To date, 28 genes have been associated with adRP (RetNet; available in the public domain at https://sph.uth.tmc.edu/RetNet/).

One of the disease-causing genes in adRP is *PRPF31* (pre-mRNA processing factor 31; NM_015629), located at 19q13.42. *PRPF31* encompasses 14 exons spanning approximately 18 kb of genomic DNA, and encoding a protein of 499 amino acids. The *PRPF31* gene is ubiquitously expressed including in neuronal tissues, brain, and retina.[4] PRPF31 is highly conserved from mammals to yeast[5] and contains a snoRNA-binding domain (NOP domain). This domain is involved in the

formation of the U4/U6*U5 tri-small nuclear ribonucleoprotein, the spliceosomal key molecule.[6] PRPF31 plays an important role in the excision of introns from RNA transcripts during the splicing process; thus, the associated pathogenic mechanism is clearly due to the disruption of the splicing function.[5] Mutations in this gene have only been associated with adRP. Since the retina contains the highest volume of processed pre-mRNA in the entire body, high demands on functional splicing components are required in retinal photoreceptors.[7] Accordingly, in *PRPF31*-associated RP, haploinsufficiency of this splicing factor leads to photoreceptor death.[7]

The prevalence of *PRPF31* mutations in adRP cohorts ranges from 1% to 8% depending on geographic origin,[8] with a previously reported prevalence of 1.7% in the Spanish population.[9] To date, more than 100 mutations have been identified, including missense and loss-of-function (LOF) mutations, such as nonsense, splice site, and frameshift variants, and also complex genomic rearrangements, as described in Human Gene Mutation Database.[10]

Incomplete penetrance in mutated *PRPF31* families including asymptomatic carriers has been previously reported by McGee et al.,[11] presumably explained by compensation of the high expressivity form of the wild-type allele. The expression of the *PRPF31* gene has a continuous distribution in the general population with a threshold for normal retina function, so asymptomatic *PRPF31* mutation carriers usually have levels of *PRPF31* gene expression above this threshold.[12] It has been postulated that this phenomenon occurs by transcriptional regulation of different modifier genes, as *CNOT*,[13,14] or recently, by a minisatellite repeat element (MSR1)[15] located in the *PRPF31* promoter.

The purpose of this study was to determine the prevalence of mutations in the *PRPF31* gene in a Spanish cohort of adRP families to gain knowledge of new pathogenic mechanisms underlying the disease and to assess further genotype/phenotype correlations.

Subjects and Methods

Subjects and Samples

A total of 211 unrelated Spanish patients with initial clinical diagnosis of adRP were included in this study. These subjects were identified after a search of the Fundación Jiménez Díaz (FJD) University Hospital database, containing 25 years of collected data. Autosomal dominant RP diagnosis was based on complete ophthalmologic examination and genetic classification, as previously described.[3,16]

All subjects provided written informed consent after the nature of the study procedures had been fully explained. The samples belong to the Biobank of the FJD University Hospital. This study was reviewed and approved by the Ethics Committee of our hospital and was performed according to the tenets of the Declaration of Helsinki and subsequent reviews.

Genomic DNA was extracted from peripheral blood leukocytes by using automated DNA extractors: BioRobot EZ1 (QIAGEN, Hilden, Germany) or MagNA Pure Compact system (Roche Applied Science, Penzberg, Germany). Total RNA was isolated from human whole blood with the PAXgene Blood RNA Kit using the automated extractor QIAcube (QIAGEN) following the manufacturer's instructions.

Mutational Analysis of *PRPF31*

Screening of the *PRPF31* gene was performed with a combined strategy of several genetic tools (Fig. 1), depending on the development of different molecular techniques. First, a screening of mutations was performed in 60 of these 211 patients by using single-strand conformation polymorphism (SSCP) and CG-clamped denaturing gradient gel electrophoresis (DGGE), as previously reported by Martinez-Gimeno et al.[9] In addition, in 189 patients, a commercial adRP genotyping microarray (Asper Biotech, Tartu, Estonia) was applied to analyze known mutations in genes previously associated with the disease, as described by Blanco-Kelly et al.[17] Second, resequencing of causative adRP genes was carried out by targeted next-generation sequencing (NGS) in a total of 76 families as previously described,[18,19] and whole exome sequencing (WES) was performed in 6 families, as reported by Almoguera et al.[20] All identified variants were confirmed by Sanger sequencing and further segregated in the respective families when other relatives were available (Supplementary Fig. S1).

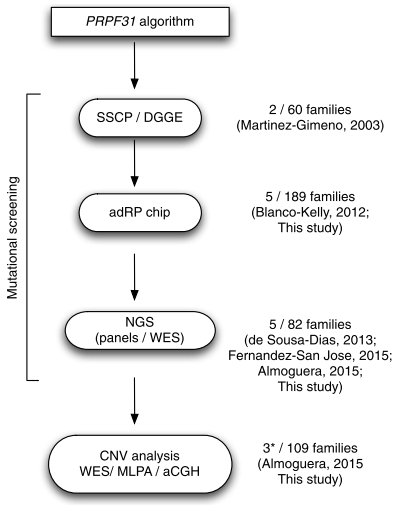

Figure 1. Molecular strategy followed for the screening of *PRPF31* in our cohort of Spanish patients with autosomal dominant retinitis pigmentosa. The workflow shows the number of families characterized versus number of families screened with each technology. *One of the families with two different rearrangements.

In silico pathogenic prediction of new splicing variants was performed with several splicing tools (Human Splicing Finder, MaxEntScan, Splice Site Finder-like, NNSPLICE, GeneSplicer, ESEFinder), all of them in the commercial Alamut software (Interactive Biosoftware, Rouen, France).

Haplotype Analysis

Five microsatellite markers with high heterozygosity (D19S921, D19S572, D19S927, D19S418, and D19S605) flanking 1.9 Mb around *PRPF31* were screened in four families. Six intragenic single-nucleotide polymorphisms (SNPs) (rs57960425, rs2303557, rs1058572, rs11556769, rs76781318, and rs655240) were genotyped in two families with the same causative mutation. Markers were separately amplified and PCR products were separated by capillary electrophoresis on an ABI 3130xl (Applied Biosystems, Foster City, CA, USA) and analyzed with GeneMapper v3.5 software (Applied Biosystems). Single-nucleotide polymorphism genotyping was performed by Sanger sequencing. Cyrillic program (Cyrllic Software, Wallingford, UK) was used for haplotype reconstruction.

Copy-Number Variation (CNV) Analysis

Copy-number variation analysis of *PRPF31* was performed by means of multiplex ligation-dependent probe amplification

(MLPA) analysis in 103 families with negative findings in these previous mutational screening, and genomic rearrangements were further refined by array-based comparative genomic hybridization (aCGH).

MLPA was performed according to manufacturer instructions by using the commercially available P235 Retinitis Pigmentosa Kit (MRC-Holland, Amsterdam, The Netherlands), which contains specific probes for all exons of the *PRPF31* gene. Amplified fragments were separated by capillary electrophoresis through an ABI 3130xl automatic analyzer and analyzed with GeneMapper (Applied Biosystems) and Coffalyser (MRC-Holland) software programs. The data were interpreted by taking the ratio of each probe signal between the control and the patient DNA sample. A ratio of 1.0 indicates the presence of two alleles (normal diploid), while a ratio below 0.6 or above 1.4 suggests deletion or duplication of the target sequence, respectively. In these positive cases, the MLPA study was extended to the other available members of the family to complete segregation analysis.

To delimit genomic rearrangements found in previous MLPA or WES analyses, aCGH was performed by using the commercial aCGH platform, SurePrint G3 CGH+SNP 2X400k (Agilent Technologies, Inc., Santa Clara, CA, USA), which includes 292,097 CGH probes and 119,091 SNP probes with 3.7 kb of mean space. Results were analyzed by Agilent CytoGenomics software v.2.7 using the default analysis method CGH+SNP v2 with the ADM-2 aberration algorithm.

Additionally, to verify whether the duplication region was in tandem, a long PCR strategy was used with the Expand Long Range dNTPack kit (Roche Life Science, Basel, Switzerland), following the manufacturer's instructions. Specific primers for the *PRPF31* exons involved in the breakpoint boundaries were designed (Exon 1F: ACTTGTCCCATCAGGCTCAG; Exon 6-7R: TCCCAAGGTCACAGTGTCAG). Polymerase chain reaction amplicons were checked in a 1% agarose gel, using Lambda DNA/*Hin*dIII as a marker. The amplification of the wild-type allele corresponds with a band of 8.6 kb, while a PCR amplification of 13.7 kb was expected for tandem duplication.

Expression of the *PRPF31* Gene

Total RNA from three individuals carrying *PRPF31* rearrangements and nine wild-type controls were transcribed by using random primers with the High Capacity cDNA Reverse Transcription kit (Applied Biosystems) in a final volume of 20 µL. Quantitative PCR (qPCR) was performed with specific primers (Exon 7-8F: CCAAGATCATGGGTGTGG; Exon 8R: GGGGCAGCACTGAGGTAG) and SYBR Green PCR Master Mix (Roche Life Science) on a LightCycler 480 Real-Time PCR System (Roche Life Science). Quantitative PCR analysis was performed twice and in triplicate. Glyceraldehyde-3-phosphate dehydrogenase was also analyzed in the qPCR experiment as an endogenous gene control. Fluorescence intensity was analyzed by using LightCycler 480 gene scanning software (Roche Life Science), and the ΔΔCt method was used to calculate relative abundances in *PRPF31* mRNA as fold-change values. The Mann-Whitney test was used to calculate statistical significance between different groups.

MSR1 Copy-Number Analysis

The number of copies of the minisatellite repeat element (MSR1) adjacent to the *PRPF31* promoter was studied in 35 symptomatic and 5 asymptomatic patients from 15 families with causative mutation in the *PRPF31* gene, as previously reported by Rose et al.[15] Additionally, a cohort of 123 control samples was also studied to obtain the minor allele frequency (MAF) of the 4 MSR1 allele. Amplified fragments were separated in an ABI 3130xl automatic analyzer and analyzed with GeneMapper. Proportion test or Fisher exact test, when the expected frequency was less than 5, was used to calculate statistical significance.

Results

Screening of *PRPF31* Mutations

Our cohort of 211 Spanish patients was sequentially screened for variants in the *PRPF31* gene, following a combination of different methods (Fig. 1). Table 1 summarizes all the mutations reported in this study. Preliminary studies on the implication of *PRPF31* in adRP were initially performed by using SSCP and DGGE in 60 patients, allowing the genetic characterization of two adRP families (RP-0637 and RP-0368; Supplementary Fig. S1), as previously reported.[9] Further *PRPF31* mutations were identified in 5 of 189 affected adRP families by using the adRP genotyping microarray (RP-2136, RP-0182, RP-0006, RP-1609, and RP-1599; Supplementary Fig. S1). Additionally, comprehensive resequencing of all exons and splicing boundaries of *PRPF31* was performed by means of customized targeted NGS gene panels and WES in 76 and 6 uncharacterized families, respectively,[18,19,20] allowing the identification of causative variants in six additional families (RP-0361, RP-1541, RP-2496, RP-1970, RP-1318, and RP-0777; Supplementary Fig. S1).

Using this combined approach, 12 different causative mutations in the *PRPF31* gene were found in 13 of 211 unrelated families, yielding a mutation detection rate of 6.2%. We found five splicing, three frameshift, one missense, one small in-frame deletion, one large deletion, and one nonsense mutation in two different families (RP-0006 and RP-1609), two of them being novel (Fig. 2; Table 1).

Analysis of *PRPF31* Rearrangements

In 103 index cases with negative results observed through the NGS custom panel approach, the genetic analysis was extended to perform MLPA, finding *PRPF31* rearrangements in two families (RP-0932 and RP-2426) (Table 1).

First, in family RP-0932, a large pedigree in which RP is segregating in nine individuals from two different branches (Fig. 3), two different types of rearrangements in *PRPF31* were identified (Table 1; Supplementary Fig. S1). Within the same family, a duplication from exons 2 to 5 was found in the index case and her affected father (IV:1 and III:1, respectively), whereas a large deletion from exons 1 to 13 was observed in two affected individuals: her paternal aunt and her cousin (III:12 and IV:12, respectively).

In both rearrangements, the breakpoint regions were defined by using a commercial aCGH of 400K probes that cover the *PRPF31* locus with 34 probes. The deletion size is at least 29.7 kb, encompassing not only *PRPF31* partially but also three contiguous upstream genes (*TFPT, NDUFA3*, and part of *OSCAR)*, as detailed in Figure 2 and Table 2. The aCGH analysis enabled us to delimit the duplication with a minimum size of 5.1 kb involving only part of *PRPF31* (Fig. 2; Table 2). Additionally, a specific long PCR allowed us to determine that both patients (IV:1 and III:1) carried a partial tandem duplication of *PRPF31* (Supplementary Fig. S2).

To check the behavior of the deletion and the duplication in the *PRPF31* gene—as this is the first time a large duplication has been reported—the expression analysis of this gene was also performed. The *PRPF31* expression was significantly decreased from patients carrying either the deletion (IV:12) or the duplication (IV:1 and III:1), showing levels ranging from

Table 1. Spectrum of *PRPF31* Mutations in adRP Spanish Families

	PRPF31 Mutations						
Family	Exon	Nucleotide Change	Protein Change	References	Technique	Incomplete Penetrance	Segregation
RP-0932_del	Deletion of *OSCAR* (E1 to E2), *NDUFA3*, *TFPT*, and *PRPF31* (E1 to 13)			**This study**	MLPA	Yes	Yes
RP-0932_dup	Duplication of *PRPF31* (E2 to E5)			**This study**	MLPA	Yes	Yes
RP-2426	Deletion of *PRPF31* (E1)			**This study**	MLPA	No	NA
RP-2496	IVS10-1	c.1074-1G>T	Splice	**This study**	NGS panel	No	NA
RP-0182	IVS11+2	c.1146+2T>A	Splice	**This study**	adRP chip	Yes	Yes
RP-0777	Deletion of *OSCAR* (E1 to E2), *NDUFA3*, *TFPT*, and *PRPF31* (E1 to 14)			**Almoguera et al., 2015**[20]	WES	No	Yes
RP-1318	4	c.322+4_322+7del	Splice	**Zhang et al., 2016**[35]	NGS panel	Yes	NA
RP-0361	5	c.328_330del	p.Ile110del	**de Sousa Dias et al., 2013**[18]	NGS panel	Yes	Yes
RP-2136	IVS6+1	c.527+1G>T	Splice	Chakarova et al., 2006[36]	adRP chip	No	NA
RP-0006	7	c.541G>T	p.Glu181*	Pomares et al., 2010[37]	adRP chip	No	Yes
RP-1609	7	c.541G>T	p.Glu181*	Pomares et al., 2010[37]	adRP chip	Yes	Yes
RP-0637	8	c.770dup	p.Thr258Aspfs*21	Vithana et al., 2001[4]	SSCP/DGGE	Yes	Yes
RP-0368	8	c.828_829del	p.His276Glnfs*2	**Martinez-Gimeno et al., 2003**[9]	SSCP/DGGE	Yes	Yes
RP-1599	9	c.895T>C	p.Cys299Arg	Sullivan et al., 2006[38]	adRP chip	Yes	Yes
RP-1541	9	c.939dup	p.Gly314Argfs*10	**Fernandez-San Jose et al., 2015**[19]	NGS panel	Yes	Yes
RP-1970	IVS11+2	c.1146+2T>C	Splice	Waseem et al., 2007[21]	NGS panel	Yes	Yes

Nucleotide numbering reflects cDNA in the reference sequence NM_015629, according to the guidelines of the Human Genome Variation Society. The initiation codon is codon +1. Different molecular approaches were used to identify pathogenic alleles. References marked in bold show mutations described in our cohort for the first time; NGS panel.[18,19] adRP chip, autosomal dominant RP genotyping microarray; E, exon; del, deletion; dup, duplication; NA, data not available.

52% to 70% as compared with control samples ($P < 0.001$) (Supplementary Fig. S3).

In family RP-2426, aCGH analysis confirmed a small deletion of a minimum size of 227 bp, which encompasses only exon 1 of the *PRPF31* gene, as previously detected by MLPA (Fig. 2; Table 2).

Breakpoint regions were also defined for family RP-0777 by using the aforementioned aCGH. This family, which was initially negative for adRP genotyping microarray, was finally characterized by WES as carrying a gross deletion encompassing the entire *PRPF31* gene along with other adjacent genes[20] (part of the *OSCAR* gene and the entire *NDUFA3* and *TFPT* genes). Thus, we identified a larger deletion of at least 32.2 kb (Fig. 2; Table 2).

Spectrum of *PRPF31* Defects in Our Spanish Population

In a total of 15 families, causative mutations or CNVs in the *PRPF31* gene were found by combining different genetic approaches. In one pedigree, two different gene rearrangements were identified within the same family. Thus, a total of 16 *PRPF31* defects were identified, obtaining a frequency of 7.6% (16/211) in this cohort. The prevalence of coding mutations and genomic rearrangements in the *PRPF31* gene were 5.7% and 1.9%, respectively.

Fifteen different mutations were identified: five splicing mutations (33.3%), three frameshift mutations (20.0%), three large deletions (20.0%), one missense mutation (6.7%), one nonsense mutation (6.7%), one in-frame mutation (6.7%), and one large duplication (6.7%).

Each mutation was found once with the exception of the p.Glu181* mutation, which was recurrent for two unrelated families (RP-0006 and RP-1609). Haplotype analysis using intragenic SNPs and microsatellite markers flanking *PRPF31* was performed in both families. The analysis of the haplotypes showed a common region sharing a minimum of 1.2 Mb and a maximum of 1.4 Mb, suggesting a founder effect rather than a recurrent mutation (Supplementary Fig. S4).

According to the available data on family history, lack of penetrance was observed in up to 66.7% (10/15) of the families.

Contribution of MSR1 Repeat Element to Disease Penetrance of *PRPF31* Mutations

Rose et al.[15] have reported that copy number variation of a cluster of MSR1 elements located upstream to the *PRPF31* promoter could have a major genetic role in determining incomplete penetrance of *PRPF31*-associated adRP by acting as modulator of its transcriptional activity. Considering that incomplete penetrance was identified in up to 66.7% of our families, affecting at least 12% of genetically confirmed *PRPF31* carriers, we have evaluated the possible implication of this MSR1 repeat element in the disease expressivity in our cohort. Thus, CNV analysis was assessed in 40 *PRPF31* mutation carriers, including 35 symptomatic from 15 different families and 5 asymptomatic individuals from 5 of these families (Supplementary Fig. S1). After genotyping of symptomatic individuals, 54% (19/35) carried the 3/4 MSR1 genotype and only 46% (16/35) carried the 3/3 genotype. In this last group, we have also included four hemizygous individuals for the 3 MSR1 allele, owing to the presence of large *PRPF31* deletions affecting the MSR1 region. No affected patients carried the 4/4 genotype. After considering the high proportion of 4 MSR1 allele, the MAF estimation on Spanish population was 0.13. Although the 3/3 (20%), 3/4 (40%), and 4/4 (40%) genotypes were found among the asymptomatic group, the 4 MSR1 allele was significantly overrepresented ($P < 0.001$) with respect to ethnically matched controls. These observations were consistent with the hypothesis that the 4-copy allele provides higher *PRPF31* gene expression and could act as a protective factor in

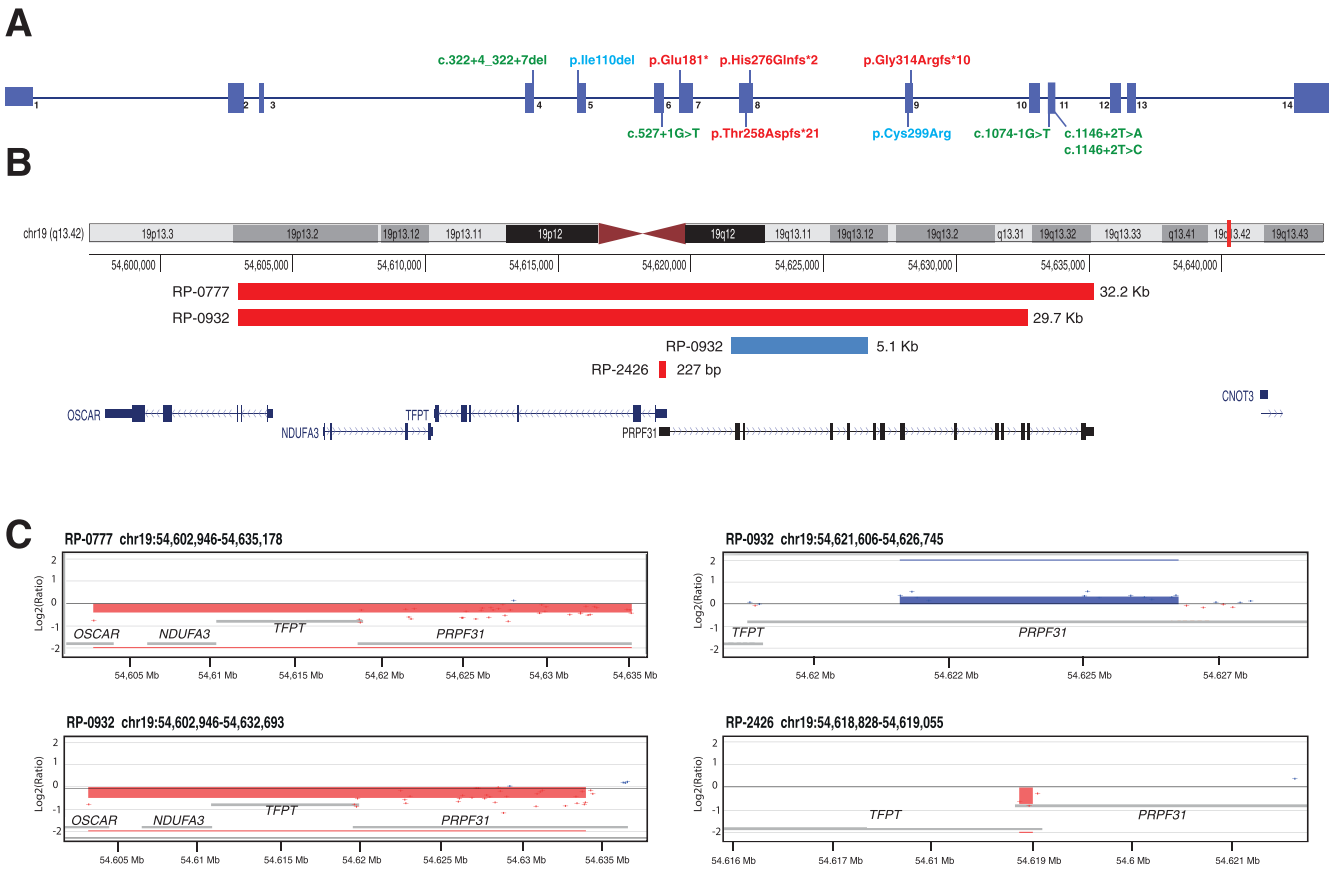

Figure 2. Overview of *PRPF31* mutations and large rearrangements identified in our study. (**A**) Schematic representation of the *PRPF31* gene showing the localization of the coding mutations identified. Exons are indicated by *colored rectangles* that are wider for the coding regions. Nucleotide numbering reflects cDNA in the reference sequence NM_015629, according to the guidelines of the Human Genome Variation Society. The initiation codon is codon +1. Splice site, missense/in-frame, and nonsense/frameshift are written in *green*, *blue*, and *red*, respectively. (**B**) Characterization of *PRPF31* rearrangements and other adjacent genes on chromosome 19 in families RP-0777, RP-0932, and RP-2426 by aCGH analysis. Commercial comparative genomic hybridization array was performed to delimit previous MLPA or WES findings in individuals carrying either deletion or duplication. *Red bars* represent the genomic positions of the deletions in families RP-0777 (chr19:54602946-54635178), RP-0932, patient IV:12, (chr19:54602946-54632693), and RP-2426 (chr19:54618828-54619055). The *blue bar* indicates the genomic positions of the duplication (chr19:54621606-54626745) in patient IV:1 in family RP-0932. Schematic representation of the complete intron–exon structure of *PRPF31* and contiguous genes are shown. Exons are indicated by *rectangles*. (**C**) Comparative genomic hybridization array data for four individuals are shown. The patient versus reference log2-ratio of each probe is plotted. Dots with log2 ratio approximately –1 indicate a heterozygous deletion (*red dots*), log2 ratio 0 indicates normal pattern, and +0.6 indicates heterozygous amplification (*blue dots*). *Shaded areas* indicate significant gains and losses. Genomic coordinates are based on the Human Genome Assembly hg19.

PRPF31-associated disease expressivity. Lack of penetrance was observed in all families with symptomatic 3/4 MSR1 alleles (Supplementary Fig. S1).

Genotype–Phenotype Correlation for *PRPF31*

A summary of clinical and genetic data is presented in Table 3 and Supplementary Table S1. Night blindness was the earliest presenting symptom (age, 11.7 ± 6.6 years), followed by visual field constriction (17.2 ± 9.7 years) and visual acuity loss (23.8 ± 13.4 years). Subcapsular cataracts were found in up to 36% of the patients (range: 21–65 years). Apparently, the most severe symptoms were observed in the patient carrying the tandem duplication (IV:1, RP-0932), in terms of onset and course of the disease. However, since we only studied one patient with a duplication, we could not obtain a general genotype–phenotype correlation. Moreover, the cases with missense/in-frame mutations seem to have an earlier onset and rapid evolution to loss of vision. Deletion of flanking genes (*OSCAR, NDUFA3*, and *TFPT* genes) was not associated with additional obvious signs.

Discussion

This study provided the prevalence not only for the single-nucleotide variants but also for genomic rearrangements in the *PRPF31* gene in a Spanish adRP cohort. In addition, we described two novel mutations, two new large deletions, and a gross duplication associated with adRP, described for the first time in this study. These results revealed a *PRPF31* mutation prevalence of 7.6%, which is higher than that previously reported in another study on the Spanish population (1.7%).[9] This prevalence is similar to that found in a French population (6.7%)[8] and higher than the one observed in the UK,[21] Japanese,[22] and Chinese populations,[23] with frequencies of 5%, 3%, and 1%, respectively. The frequency of genomic rearrangements in the *PRPF31* gene, first reported for the Spanish population, accounts for 1.9%. This prevalence is lower than one previously described[24] in a North American population but higher than in France,[8] where no genomic rearrangements have been found.

Mutational *PRPF31* spectrum in our cohort is very heterogeneous, as we found a total of 16 mutations in 15 different families tested. No hotspot or recurrent mutations

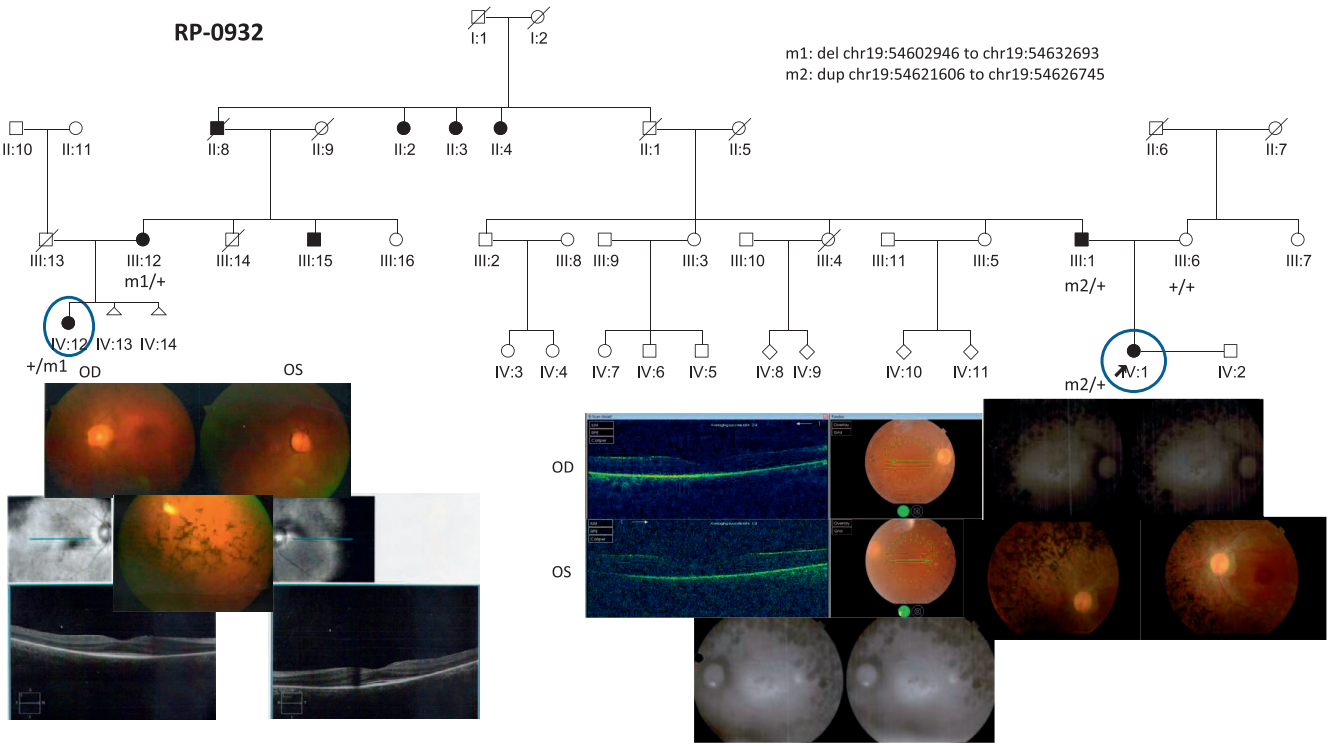

Figure 3. Pedigree of family RP-0932 and retinal imaging from the index case and her cousin, who carried duplication and deletion within the *PRPF31* gene, respectively. Ophthalmoscopic and optic coherence tomography (OCT) examination of patients IV:1 (proband, with a duplication within the *PRPF31* gene) and IV:12 (cousin, with a deletion within the *PRPF31* gene). Patient IV:1, at the age of 41 years, showed midperipheral and posterior (outside the arcades) clumps of black pigment in bone spicule configuration, macular alteration, attenuated retinal vessels, atrophy of the retinal pigment epithelium (RPE) with choroidal transparency, and pallor of the optic nerve. Impression of disruption and thinning of the retina (mainly on OS) on OCT. Patient IV:12, at the age of 38 years, showed midperipheral and posterior (outside the arcades) clumps of black pigment in bone spicule configuration, with macular alteration, attenuated retinal vessels, and pallor of the optic nerve. Slight epiretinal membrane on OD. Thinning of retina on OCT. OD, right eye; OS, left eye.

were found, except for the p.Glu181* mutation, a previously known mutation,[19] which was detected in two of our families. Both families have the same geographic origin and share a 1.2-Mb common haplotype, therefore suggesting a founder effect. Additionally, in the RP-0932 family, two different large rearrangements were found in different individuals within the same family.

Most of the mutations found in our population are LOF mutations (e.g., nonsense, splicing, frameshift mutation, or gross rearrangements), presumably causing premature termination codon in alternative transcripts susceptible to being degraded via nonsense-mediated mRNA decay (NMD), as has been previously described by Rio Frio et al.[25] For this reason, most of the *PRPF31* mutations seem to cause disease by haploinsufficiency.[25] In the RP-0932 family, the *PRPF31* expression was significantly reduced from 30% to 48% as compared to control samples in both deletion and duplication patients. Our finding supports the hypothesis of haploinsufficiency as the pathogenic mechanism associated with *PRPF31*. In the case of the tandem duplication, the most likely cause of haploinsufficiency is the degradation of aberrant *PRPF31* transcripts through NMD. In all the patients carrying *PRPF31* deletions, the 5′UTR region on the noncoding exon 1, which is predicted to contain several transcriptional regulatory sites, is involved, and as a result, *PRPF31* haploinsufficiency could be explained by dysregulation of gene transcription.

In view of the high genetic heterogeneity and complexity of *PRPF31*-associated mutational mechanisms, molecular studies for adRP require not only a comprehensive resequencing of exons and splicing sites of this gene but also a copy-number assay by means of MLPA or aCGH to detect large rearrangements. MLPA has some advantages including ease of use, high throughput, and relatively small quantities of DNA needed, but on the other hand, aCGH allows a more reliable refining of rearrangements and breakpoints, particularly if other adjacent genes seem to be involved. Although a large deletion in family

Table 2. Patients Carrying Large Genomic Rearrangements on 19q13.42, Affecting *PRPF31*

	Mutation			Size, bp	
Family	Kind of Rearrangement	Genomic Coordinates, hg19	Involved Genes	Minimum	Maximum
RP-0777	Deletion	chr19:54,602,946-54,635,178	*OSCAR* (E1 to E2), *NDUFA3*, *TFPT*, and entire *PRPF31*	32,232	111,184
RP-0932	Deletion	chr19:54,602,946-54,632,693	*OSCAR* (E1 to E2), *NDUFA3*, *TFPT*, and *PRPF31* (E1 to E13)	29,747	87,435
RP-0932	Duplication	chr19:54,621,606-54,626,745	*PRPF31* (E2 to E5)	5,139	7,832
RP-2426	Deletion	chr19:54,618,828-54,619,055	*PRPF31* (E1)	227	18,609

Table 3. Age of Onset of Night Blindness, Visual Field Constriction, Visual Acuity Loss, and Cataract in Different Types of Mutations

Mutation Type	NB (No. of Cases)	VFC (No. of Cases)	VAL (No. of Cases)	% of VAL	Cataract (No. of Cases)	% Cataract
Splice site	12.6 ± 5.0 (5)	17.0 ± 11.4 (3)	25.5 ± 13.4 (2)	67%	38.0 (1)	17%
Missense/in-frame	10.3 ± 6.4 (3)	10.0 ± 4.7 (4)	13.3 ± 7.8 (4)	100%	30.0 (1)	20%
FS/nonsense	10.5 ± 6.1 (7)	21.8 ± 10.1 (6)	37.3 ± 10.3 (4)	75%	44.4 ± 16.9 (5)	63%
Tandem duplication	5.0 (1)	7.0 (1)	20.0 (1)	100%	25.0 (1)	50%
Partial deletion	27.0 (1)	21.0 ± 9.6 (3)	13.0 (1)	33%	50.0 (1)	33%
Total	12.3 ± 6.8	14.9 ± 6.1	20.1 ± 10.0		37.5 ± 10.2	

Data shown as average ± standard deviation. FS, frameshift; NB, night blindness; VAL, visual acuity loss; VFC, visual field constriction.

RP-0777 was identified by WES, in general CNV analyses using NGS data are still constrained owing to technical issues such as coverage and design.

We have successfully applied aCGH for refining *PRPF31* rearrangements that were previously detected by other techniques. Here, we used a commercial array that specifically targets genes and loci previously associated with pathogenic chromosomal aberrations, as in the case of *PRFP31*. Although this design allowed for a high coverage of *PRPF31*, its up- and downstream genomic regions are less covered, so in several cases, we could not establish the exact breakpoints for the largest deletions found in our patients. This was the case of contiguous gene deletions in the RP-0932 and RP-0777 families, for which the deletion seems to start at the same point (Table 2). In the RP-0777 family, the detected deletion was longer than that previously reported,[20] including *PRPF31* exon 14 and *OSCAR* exon 2.

Large deletions in *PRPF31* affecting several continuous exons, the whole gene, or even part of *PRPF31* along upstream genes[24,26] have been described previously, as in our cohort. To our knowledge, this is the first time that two different kinds of *PRPF31* rearrangements—a duplication and a deletion—are reported in different members within the same family and associated with adRP. Interestingly, we also identified a small deletion involving only a single exon of *PRPF31*. In family RP-2426, a deletion of the noncoding exon 1 was detected; thus, a haploinsufficiency mechanism was expected in affected patients due to loss of regulatory regions. Unfortunately, *PRPF31* gene expression studies were not available for this family.

The *PRPF31* gene is located on a region rich in repeat elements, especially Alu repeats. Alu elements are a type of short interspersed element that accounts for 10% of the total genome content, but in chromosome 19 accounts for 26.3%, with Alu being the richest chromosome.[27] Alu repeats have been implicated in human genetic diseases by insertion of Alu elements or by nonallelic homologous recombination causing rearrangements. This last mechanism, which is predicted to cause 0.3% of human genetic diseases,[28] could be the most likely explanation for the high prevalence of genomic *PRPF31* rearrangements. The mechanism underlying these complex structural rearrangements in family RP-0932 could be nonallelic homologous recombination between nearly identical Alu repeats.

Incomplete penetrance events were found in a large number of families with mutation in the *PRPF31* gene, in particular in 66.7% of the families of our cohort. The mechanism that explains this phenomenon is allelic imbalance with high expression of the wild-type allele, compensating for the mutant allele in asymptomatic carriers.[29,30] A possible role of *PRPF31* modifier genetic factors, such as *CNOT3* or the MSR1 repeat element upstream of PRPF31,[12] have been hypothesized to explain this overexpression event.[13–15] These modifier genetic factors are involved in transcriptional repression of the *PRPF31* gene.[13–15] However, the exact mechanism is still unclear.

The 3- and 4-copies of the MSR1 minisatellite showed strong differences in the modulation of the *PRPF31* gene transcription, which could be associated with incomplete penetrance in *PRPF31*-associated RP.[15] Consistent with the hypothesis that the 4 MSR1 allele seems to be a major protective factor for disease expressivity,[15] asymptomatic individuals from our cohort had a frequency for 4 MSR1 allele significantly higher than that of the control population. Rose et al.[15] have reported that all symptomatic patients in their study are homozygous for the 3 MSR1 allele, predicted to have a decreased *PRPF31* expression, and thus developing the disease. However, 33% of our families (RP-0932, RP-0368, RP-0637, RP-0361, and RP-1599) had at least one symptomatic individual that was heterozygous 3/4; this fact is contradictory with that seen by Rose et al.,[15] reflecting a higher frequency of 4 MSR1 allele than previously reported. In these heterozygous symptomatic cases, it is expected that the *PRPF3*1 gene expression will be reduced and then symptoms will develop if the wild-type allele is in *cis* with the 3-repeat MSR1 allele and the 4 MSR1 allele is in *cis* with the mutant allele, so the 4-copy does not seem to act as a protective factor. Conversely, in heterozygous asymptomatic individuals, the protective 4-allele should be in *cis* with the wild-type allele in order to maintain a high expression of *PRPF31*, thus compensating the mutated allele. Unfortunately, it was not possible to fully confirm this hypothesis in our study, given that the complete phasing of alleles could not be inferred in most families owing to lack of familial information and/or sample availability. However, we observed two clear exceptions for this hypothesis in two heterozygous 3/4 asymptomatic individuals, the III:6 and III:9 individuals from families RP-0361 and RP-1599, respectively, where the 4 MSR1 allele is necessarily in *cis* with the mutated allele as the haplotype analysis showed (Supplementary Fig. S1). In these two cases, the MSR1 cannot explain the asymptomatic status, so something else must be present outside the 1.9 Mb region around the *PRPF31* gene to explain the incomplete penetrance. This last observation is in agreement with a previous hypothesis of a polygenic control of *PRPF31* transcriptional regulation in the incomplete penetrance of *PRPF31*-associated RP.

An ophthalmic review of 26 patients from 15 different families carrying *PRPF31* mutations was performed, representing the largest phenotypic description for this gene to our knowledge. The clinical signs observed in family RP-0932 seem to be more severe in the patients carrying the duplication than in the patients carrying the deletion in *PRPF31*. Thus, this family represents a good example for phenotypic variability associated with different *PRPF31* defects. In our cohort, patients carrying a missense mutation, in contrast to those patients carrying a nonsense mutation or a frameshift mutation, did not necessarily have a milder phenotype as previously reported by Xu et al.[31] Though these results are preliminary and cannot be used to establish a clear genotype–

phenotype correlation because clinical information for patients carrying *PRPF31* defects was limited in some cases, it seems that the duplication and the missense mutations could produce a more severe phenotype than the other kind of mutations. Patients from families RP-0932 and RP-0777 with deletion in additional upstream genes did not show additional clinical findings apart from RP, which is consistent with previous observations.[32–34]

In conclusion, in light of the high genetic heterogeneity of *PRPF31* mutations, it is necessary to screen all exons and splicing boundaries and also perform copy-number assays, as large duplications or deletions in the *PRPF31* gene can be also identified. The identification of novel pathogenic mechanisms underlying *PRPF31*-related RP and genotype–phenotype correlation is important not only for current diagnostics but also for genetic counseling. In addition, it is particularly important for the development of new gene-dependent therapies and further inclusion of the patients in human gene-specific clinical trials.

Acknowledgments

The authors thank the Genetics and Ophthalmology Departments of the Fundación Jimenez Diaz University Hospital (Madrid) and especially Asunción Giménez, Cristina Nieto, Cristina Villaverde, and Olga Zurita for their technical assistance. They also thank Oliver Shaw for English review and editing the manuscript. They thank all patients and doctors who participated in the study, especially Maria Asunción López Ariztegui from Hospital Universitario Cruces, Bilbao, Guiomar Perez de Nanclares from Hospital Universitario Araba-Txagorritxu, Vitoria, and Cristina Irigoyen Laborra from Hospital Universitario de Donostia, San Sebastian.

Supported by several grants from the Instituto de Salud Carlos III (ISCIII) from the Spanish Ministry of Health, including CIBERER (06/07/0036), FIS – FEDER (European Regional Development Fund) (PI013/00226), and IIS-FJD Biobank PT13/0010/0012. In addition, the Spanish National Organization of the Blind (ONCE) and the Spanish Fighting Blindness Foundation (FUNDALUCE) also supported our work. IM-M and RS-A are sponsored by the Rio Hortega and Sara Borrell Postdoctoral Programs (CD14-00079 and CD12-00676) from ISCIII/FEDER. MCo is supported by the Miguel Servet Program (CP12/03256) from ISCIII. RP-C is supported by Fundación Conchita Rábago (FCR), and LRJdS was supported by the CAPES Foundation, Ministry of Education of Brazil.

Disclosure: **I. Martin-Merida**, None; **R. Sanchez-Alcudia**, None; **P. Fernandez-San Jose**, None; **F. Blanco-Kelly**, None; **R. Perez-Carro**, None; **L. Rodrigues-Jacy da Silva**, None; **B. Almoguera**, None; **B. Garcia-Sandoval**, None; **M.I. Lopez-Molina**, None; **A. Avila-Fernandez**, None; **M. Carballo**, None; **M. Corton**, None; **C. Ayuso**, None

References

1. Hamel C. Retinitis pigmentosa. *Orphanet J Rare Dis*. 2006;1:40.
2. Ayuso C, Millan JM. Retinitis pigmentosa and allied conditions today: a paradigm of translational research. *Genome Med*. 2010;2:34.
3. Ayuso C, Garcia-Sandoval B, Najera C, Valverde D, Carballo M, Antinolo G. Retinitis pigmentosa in Spain: The Spanish Multicentric and Multidisciplinary Group for Research into Retinitis Pigmentosa. *Clin Genet*. 1995;48:120–122.
4. Vithana EN, Abu-Safieh L, Allen MJ, et al. A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol Cell*. 2001;8:375–381.
5. Makarova OV, Makarov EM, Liu S, Vornlocher HP, Luhrmann R. Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J*. 2002;21:1148–1157.
6. Liu S, Li P, Dybkov O, et al. Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science*. 2007;316:115–120.
7. Cao H, Wu J, Lam S, et al. Temporal and tissue specific regulation of RP-associated splicing factor genes PRPF3, PRPF31 and PRPC8—implications in the pathogenesis of RP. *PLoS One*. 2011;6:e15860.
8. Audo I, Bujakowska K, Mohand-Said S, et al. Prevalence and novelty of PRPF31 mutations in French autosomal dominant rod-cone dystrophy patients and a review of published reports. *BMC Med Genet*. 2010;11:145.
9. Martinez-Gimeno M, Gamundi MJ, Hernan I, et al. Mutations in the pre-mRNA splicing-factor genes PRPF3, PRPF8, and PRPF31 in Spanish families with autosomal dominant retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2003;44:2171–2177.
10. Stenson PD, Ball EV, Mort M, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat*. 2003;21:577–581.
11. McGee TL, Devoto M, Ott J, Berson EL, Dryja TP. Evidence that the penetrance of mutations at the RP11 locus causing dominant retinitis pigmentosa is influenced by a gene linked to the homologous RP11 allele. *Am J Hum Genet*. 1997;61:1059–1066.
12. Rose AM, Bhattacharya SS. Variant haploinsufficiency and phenotypic non-penetrance in PRPF31-associated retinitis pigmentosa. *Clin Genet*. 2016;90:118–126.
13. Venturini G, Rose AM, Shah AZ, Bhattacharya SS, Rivolta C. CNOT3 is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance. *PLoS Genet*. 2012;8:e1003040.
14. Rose AM, Shah AZ, Venturini G, Rivolta C, Rose GE, Bhattacharya SS. Dominant PRPF31 mutations are hypostatic to a recessive CNOT3 polymorphism in retinitis pigmentosa: a novel phenomenon of "linked trans-acting epistasis." *Ann Hum Genet*. 2014;78:62–71.
15. Rose AM, Shah AZ, Venturini G, et al. Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa. *Sci Rep*. 2016;6:19450.
16. Fernandez-SanJose P, Blanco-Kelly F, Corton M, et al. Prevalence of rhodopsin mutations in autosomal dominant retinitis pigmentosa in Spain: clinical and analytical review in 200 families. *Acta Ophthalmol*. 2015;93:e38–e44.
17. Blanco-Kelly F, Garcia-Hoyos M, Corton M, et al. Genotyping microarray: mutation screening in Spanish families with autosomal dominant retinitis pigmentosa. *Mol Vis*. 2012;18:1478–1483.
18. de SousaDias M, Hernan I, Pascual B, et al. Detection of novel mutations that cause autosomal dominant retinitis pigmentosa in candidate genes by long-range PCR amplification and next-generation sequencing. *Mol Vis*. 2013;19:654–664.
19. Fernandez-SanJose P, Corton M, Blanco-Kelly F, et al. Targeted next-generation sequencing improves the diagnosis of autosomal dominant retinitis pigmentosa in Spanish patients. *Invest Ophthalmol Vis Sci*. 2015;56:2173–2182.
20. Almoguera B, Li J, Fernandez-SanJose P, et al. Application of whole exome sequencing in six families with an initial diagnosis of autosomal dominant retinitis pigmentosa: lessons learned. *PLoS One*. 2015;10:e0133624.
21. Waseem NH, Vaclavik V, Webster A, Jenkins SA, Bird AC, Bhattacharya SS. Mutations in the gene coding for the pre-mRNA splicing factor, PRPF31, in patients with autosomal dominant retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2007;48:1330–1334.

22. Sato H, Wada Y, Itabashi T, Nakamura M, Kawamura M, Tamai M. Mutations in the pre-mRNA splicing gene, PRPF31, in Japanese families with autosomal dominant retinitis pigmentosa. *Am J Ophthalmol*. 2005;140:537–540.
23. Lim KP, Yip SP, Cheung SC, Leung KW, Lam ST, To CH. Novel PRPF31 and PRPH2 mutations and co-occurrence of PRPF31 and RHO mutations in Chinese patients with retinitis pigmentosa. *Arch Ophthalmol*. 2009;127:784–790.
24. Sullivan LS, Bowne SJ, Seaman CR, et al. Genomic rearrangements of the PRPF31 gene account for 2.5% of autosomal dominant retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2006;47:4579–4588.
25. Rio Frio T, Wade NM, Ransijn A, Berson EL, Beckmann JS, Rivolta C. Premature termination codons in PRPF31 cause retinitis pigmentosa via haploinsufficiency due to nonsense-mediated mRNA decay. *J Clin Invest*. 2008;118:1519–1531.
26. Eisenberger T, Neuhaus C, Khan AO, et al. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PLoS One*. 2013;8:e78496.
27. Grover D, Mukerji M, Bhatnagar P, Kannan K, Brahmachari SK. Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. *Bioinformatics*. 2004;20:813–817.
28. Deininger PL, Batzer MA. Alu repeats and human disease. *Mol Genet Metab*. 1999;67:183–193.
29. Vithana EN, Abu-Safieh L, Pelosini L, et al. Expression of PRPF31 mRNA in patients with autosomal dominant retinitis pigmentosa: a molecular clue for incomplete penetrance? *Invest Ophthalmol Vis Sci*. 2003;44:4204–4209.
30. Rivolta C, McGee TL, Rio Frio T, Jensen RV, Berson EL, Dryja TP. Variation in retinitis pigmentosa-11 (PRPF31 or RP11) gene expression between symptomatic and asymptomatic patients with dominant RP11 mutations. *Hum Mutat*. 2006;27:644–653.
31. Xu F, Sui R, Liang X, Li H, Jiang R, Dong F. Novel PRPF31 mutations associated with Chinese autosomal dominant retinitis pigmentosa patients. *Mol Vis*. 2012;18:3021–3028.
32. Abu-Safieh L, Vithana EN, Mantel I, et al. A large deletion in the adRP gene PRPF31: evidence that haploinsufficiency is the cause of disease. *Mol Vis*. 2006;12:384–388.
33. Rose AM, Mukhopadhyay R, Webster AR, Bhattacharya SS, Waseem NH. A 112 kb deletion in chromosome 19q13.42 leads to retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2011;52:6597–6603.
34. Köhn L, Bowne SJ, Sullivan LS, et al. Breakpoint characterization of a novel approximately 59 kb genomic deletion on 19q13.42 in autosomal-dominant retinitis pigmentosa with incomplete penetrance. *Eur J Hum Genet*. 2009;17:651–655.
35. Zhang Q, Xu M, Verriotto JD, et al. Next-generation sequencing-based molecular diagnosis of 35 Hispanic retinitis pigmentosa probands. *Sci Rep*. 2016;6:32792.
36. Chakarova CF, Cherninkova S, Tournev I, et al. Molecular genetics of retinitis pigmentosa in two Romani (Gypsy) families. *Mol Vis*. 2006;12:909–914.
37. Pomares E, Riera M, Permanyer J, et al. Comprehensive SNP-chip for retinitis pigmentosa-Leber congenital amaurosis diagnosis: new mutations and detection of mutational founder effects. *Eur J Hum Genet*. 2010;18:118–124.
38. Sullivan LS, Bowne SJ, Birch DG, et al. Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: a screen of known genes in 200 families. *Invest Ophthalmol Vis Sci*. 2006;47:3052–3064.

Investigative Ophthalmology & Visual Science