Targeted Next-Generation Sequencing Improves the Diagnosis of Autosomal Dominant Retinitis Pigmentosa in Spanish Patients

Patricia Fernandez-San Jose,[1,2] Marta Corton,[1,2] Fiona Blanco-Kelly,[1,2] Almudena Avila-Fernandez,[1,2] Miguel Angel Lopez-Martinez,[1,2] Iker Sanchez-Navarro,[1,2] Rocio Sanchez-Alcudia,[1,2] Raquel Perez-Carro,[1,2] Olga Zurita,[1,2] Noelia Sanchez-Bolivar,[1] Maria Isabel Lopez-Molina,[2,3] Blanca Garcia-Sandoval,[2,3] Rosa Riveiro-Alvarez,[1,2] and Carmen Ayuso[1,2]

[1]Department of Genetics, Instituto de Investigacion Sanitaria-Fundacion Jimenez Diaz University Hospital (IIS-FJD, UAM), Madrid, Spain
[2]Center of Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain
[3]Department of Ophthalmology, Instituto de Investigacion Sanitaria-Fundacion Jimenez Diaz University Hospital (IIS-FJD, UAM), Madrid, Spain

Correspondence: Carmen Ayuso, IIS-Fundacion Jimenez Diaz University Hospital (IIS-FJD, UAM), Av. Reyes Católicos No. 2, Madrid, 28040 Spain;
cayuso@fjd.es.

Submitted: December 1, 2014
Accepted: February 11, 2015

Citation: Fernandez-San Jose 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. DOI:10.1167/iovs.14-16178

Purpose. Next-generation sequencing (NGS) has been demonstrated to be an effective strategy for the detection of mutations in retinal dystrophies, a group of inherited diseases that are highly heterogeneous. Therefore, the aim of this study is the application of an NGS-based approach in a Spanish cohort of autosomal dominant retinitis pigmentosa (RP) patients to find out causative mutations.

Methods. Index cases of 59 Spanish families with initial diagnosis of autosomal dominant RP and unsuccessfully studied for mutations in the most common RP causal genes, were selected for application of a NGS-based approach with a custom panel for 73 genes related to retinal dystrophies. Candidate variants were select based on frequency, pathogenicity, inherited model, and phenotype. Subsequently, confirmation by Sanger sequencing, cosegregation analysis, and population studies, was applied for determining the implication of those variants in the pathology.

Results. Overall 31 candidate variants were selected. From them, 17 variants were considered as mutations causative of the disease, 64% (11/17) of them were novel and 36% (6/17) were known RP-related mutations. Therefore, applying this technology16 families were characterized rendering a mutation detection rate of 27% (16/59). Of them, 5% (3/59) of cases displayed mutations in recessive or X-linked genes (*ABCA4*, *RPGR*, and *RP2*) allowing a genetic and clinical reclassification of those families. Furthermore, seven novel variants with uncertain significance and seven novel variants probably not causative of disease were also found.

Conclusions. This NGS strategy is a fast, effective, and reliable tool to detect known and novel mutations in autosomal dominant RP patients allowing genetic reclassification in some cases and increasing the knowledge of pathogenesis in retinal dystrophies.

Keywords: target NGS, retinitis pigmentosa, autosomal dominant, Spanish families

Retinitis pigmentosa (RP), with a prevalence of 1:4000,[1] is the most common form of inherited retinal dystrophy (RD) affecting mainly rod photoreceptors and inducing a progressive vision loss.[2] The disease is characterized by an extraordinary genetic and clinical heterogeneity. Clinical symptoms include night blindness, progressive loss of peripheral vision, decreased visual acuity in late stages, abnormal retinal electrophysiology, and ocular fundus alterations.

Genetic heterogeneity of this disorder includes classical monogenic mode of inheritance (autosomal dominant, autosomal recessive, and X-linked transmission) as well as digenic and mitochondrial inherited patterns. Autosomal dominant retinitis pigmentosa (adRP) accounts for approximately 15% of Spanish RP families.[3] To date, more than 70 genes have been associated with nonsyndromic RP and 25 of them have been identified related to adRP (Retnet, updated November 2014; available in the public domain, https://sph.uth.edu/retnet/).

Genetic analysis of these patients is complicated because of the large number of genes involved in adRP disease, the presence of incomplete penetrance in some families and a high inter- and intrafamilial phenotypic heterogeneity exhibited by some of them.[2,4]

Recently, targeted next-generation sequencing (NGS) has become an effective tool for determining the causal mutation in adRP as demonstrated by different reports.[5–11] Therefore, we have developed a specific strategy based on DNA capture of coding and noncoding regions of 73 genes associated with RP

and Leber congenital amaurosis (LCA) that we are applying to the diagnosis of uncharacterized RD patients.

Because it has been estimated that approximately 80% of adRP families have mutations in known genes,[12] the current study proposes the use of a customized NGS panel of RD genes in a large cohort of adRP Spanish families unsuccessfully studied with classical tools.

Methods

Subjects and Samples

Among 116 uncharacterized index patients out of 195 families with initial diagnosis of RP and likely autosomal dominant inheritance, 59 were selected to be included in the proposed NGS-based approach. Diagnosis of adRP was determined according to a dominant mode of inheritance in patients with night blindness, peripheral vision loss, and ocular fundus alteration and reduced scotopic response.[3] Clinical evaluation was assessed as previously described.[13]

Additionally, 150 healthy unrelated Spanish individuals without familial history of retinal dystrophy were screened as controls to evaluate the frequency of the novel missense and in-frame variants.

All families were recruited at the Fundacion Jimenez Diaz University Hospital (Madrid, Spain). This study was reviewed and approved by the ethics committee of the Hospital following the tenets of the Declaration of Helsinki and reviews. The participating subjects, or their legal guardians, signed a written informed consent form after the nature of procedures had been explained.

Genomic DNA was extracted from peripheral blood leukocytes using automated DNA extractors: BioRobot EZ1 (Qiagen, Hilden, Germany) and MagNA Pure Compact system (Roche Applied Science, Penzberg, Germany).

Molecular Methods

All 59 index patients had been previously screened using a combined strategy of several molecular tools: single-strand conformation polymorphism (SSCP), CG-clamped denaturing gradient gel electrophoresis (DGGE), genotyping ADRP Chip (AsperBiotech, Tartu, Estonia), long-range PCR followed by NGS, and Sanger sequencing of the most prevalent adRP genes[13–16] (Supplementary Table S1).

Design of RD Panel

A customized RD_NGS_Panel of 73 previously known RD genes reported in Retnet (available in the public domain at https://sph.uth.edu/retnet/) and literature was developed using the Haloplex capture technology (Agilent Technologies, Inc., Santa Clara, CA, USA; Supplementary Table S2).

First, genomic coordinates of coding and noncoding exons for all RefSeq transcripts (hg19) were obtained from the University of California Santa Cruz (UCSC; Santa Cruz, CA, USA) Genome Browser using the Galaxy software.[17]

Amplicons for all coding and noncoding exons, including 20 bp of flanking 5′ and 3′ intronic sequence were designed using the SureDesign tool (available in the public domain at https://earray.chem.agilent.com/suredesign/; Agilent Technologies, Inc.). Additionally four intronic regions for previously known deep intronic mutations were also included in the final design (Supplementary Table S3).

In total, 1127 regions were targeted comprising 352 kb of target sequence. The final design covered 99.1% of the requested target regions, as showed in Supplementary Table S2.

Sequence Capture and Next-Generation Sequencing

Target enrichment was performed according to HaloPlex Enrichment System for Illumina Sequencing protocol (version D3, December 2012; Agilent Technologies, Inc.) with some modifications: (1) genomic DNA starting material (450 ng) was digested in eight restriction reactions, (2) digested fragments were hybridized to the complementary HaloPlex probe capture library. In this step, index sequences for multiplexing were incorporated into the targeted fragments, (3) DNA-HaloPlex probe hybrids containing biotin were captured on streptavidin beads and ligation of fragments was performed, (4) captured target libraries were amplified by PCR, and (5) amplified products were purified twice with AMPure XP beads (Beckman Coulter, Brea, CA, USA) and additionally, the number of washes with EtOH was doubled. Its concentrations were determined using a 2100 Bioanalyzer (Agilent Technologies, Inc.). Captured target libraries from 59 probands were pooled and finally sequenced using MiSeq v2 reagents kit (Illumina, San Diego, California, USA) on six runs in an Illumina MiSeq system to obtain 150 bp paired-end reads. The base calling and quality control were performed using Illumina RTA sequence analysis pipeline.

Bioinformatic Analysis

A specific custom pipeline for HaloPlex kits on Illumina implemented into the commercial DNAnexus platform (available in the public domain at https://www.dnanexus.com/) was used for the bioinformatic analysis (Fig. 1). After trimming the adapter sequence, reads from the Illumina MiSeq instrument were mapped against the hg19 human reference genome using Burrows-Wheeler Aligner (BWA)-MEM version 0.7.5a with default parameters to generate the mappings.[18] Genome Analysis Toolkit(GATK)-lite version 2.3-9 was applied for local realignment, recalibration of base quality scores and variant calling.[19,20] Reads and coverage statistics were produced using PRINSEQ-lite version 0.20, Picard CalculateHsMetrics version 1.97 (available in the public domain at www.picard.sourceforge.net), and BedTools Coverage version 2.17.0 tools.[21,22] Finally, GATK Variant Annotator and ENSEMBL Variant Effect Predictor v72 tools were used for variant annotation and creation of vCard files[23] (Fig. 1).

Appraisal of Sensitivity

Twenty-nine known variants previously detected by Sanger sequencing in 16 of 59 cases were used as control samples to evaluate the technology applied (Supplementary Table S4).

Variant Priorization and Pathogenicity Assessment

The following criteria were considered for variant filtering (Fig. 1): (1) variants were selected focusing on known, novel, and rare variants in exome target excluding intergenic or intronic sequence. The former were considered directly as pathogenic using the Human Gene Mutation Database (HGMD Professional; available in the public domain at www.biobase-international.com/product/hgmd) and rare variants were filtered by a minor allele frequency (MAF) less than or equal to 0.001 in the 1000 Genomes (available in the public domain at www.1000genomes.org/data), dbSNP132 (in the public domain, http:// www.ncbi.nlm.nih.gov/projects/SNP), and Exome Variant Server (EVS; available in the public domain at http://evs.gs.washington.edu/EVS/) databases, (2) novel and rare variants causing protein truncation such as nonsense and frameshift were considered as pathogenic, (3) then, variants in silico predicted as having

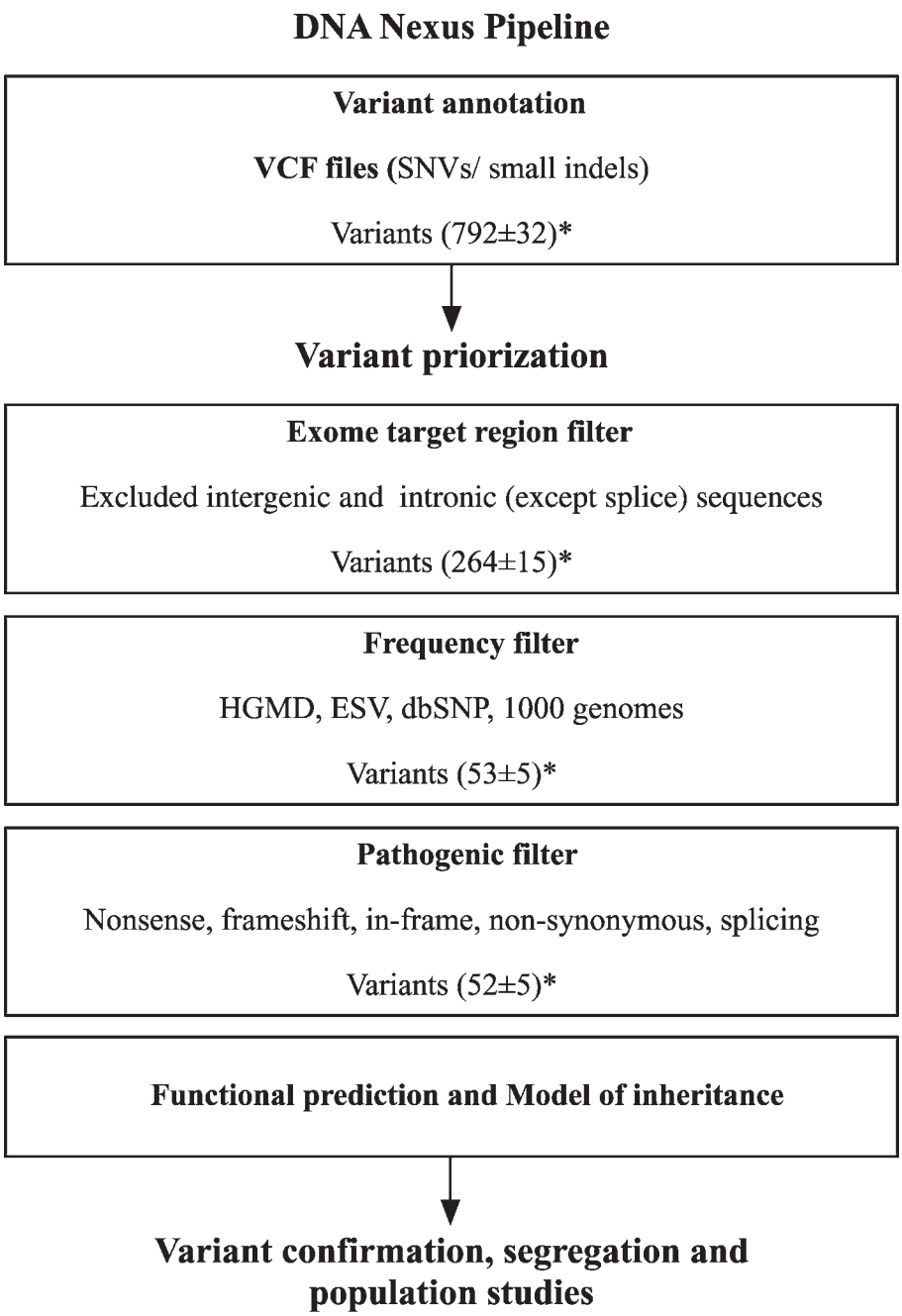

FIGURE 1. Pipeline used to identify and validate candidate variants. *Average and SD per case.

deleterious effect on mRNA splicing as analyzed by the Human Splicing Finder version 2.4 (available in the public domain at www.umd.be/HSF) software tool were also incorporated, (4) nonsynonymous single nucleotide variants (SNVs) predicted as damaging in at least two of three in silico tools (SIFT,[24] Mutation Taster,[25] and PolyPhen-2[26]) used for pathogenicity prediction were also included; and (5) heterozygous variants in genes previously associated with adRP were considered but when no candidate variants were found in the above steps, variants associated with different RD patterns and/or clinical presentations were also included.

All candidate causal variants identified by NGS were further verified by Sanger sequencing on an ABI 3130 Genetic Analyzer sequencer (Applied Biosystems, Foster City, CA, USA). Cosegregation analysis of each variant that was suspected as causal RP phenotype was performed using available DNA from affected and unaffected family members assuming in most cases an autosomal dominant trait model with no phenocopies and complete penetrance. All candidate variants were screened in 150 alleles from 75 in-house whole exomes Spanish control individuals (available in the public domain at http://bioinfo.cipf.es/apps-beta/exome-server/beta/). Additionally, novel, nonsynonymous, and in-frame variants were screened in another 150 Spanish control chromosomes by Sanger sequencing and high resolution melting (HRM) as previously reported.[27]

Finally, candidate causal variants were classified into three categories according to their demonstrated level of pathogenicity, as shown in Figure 2: (1) Variants causative of the disease, including known RD mutations, novel loss of function

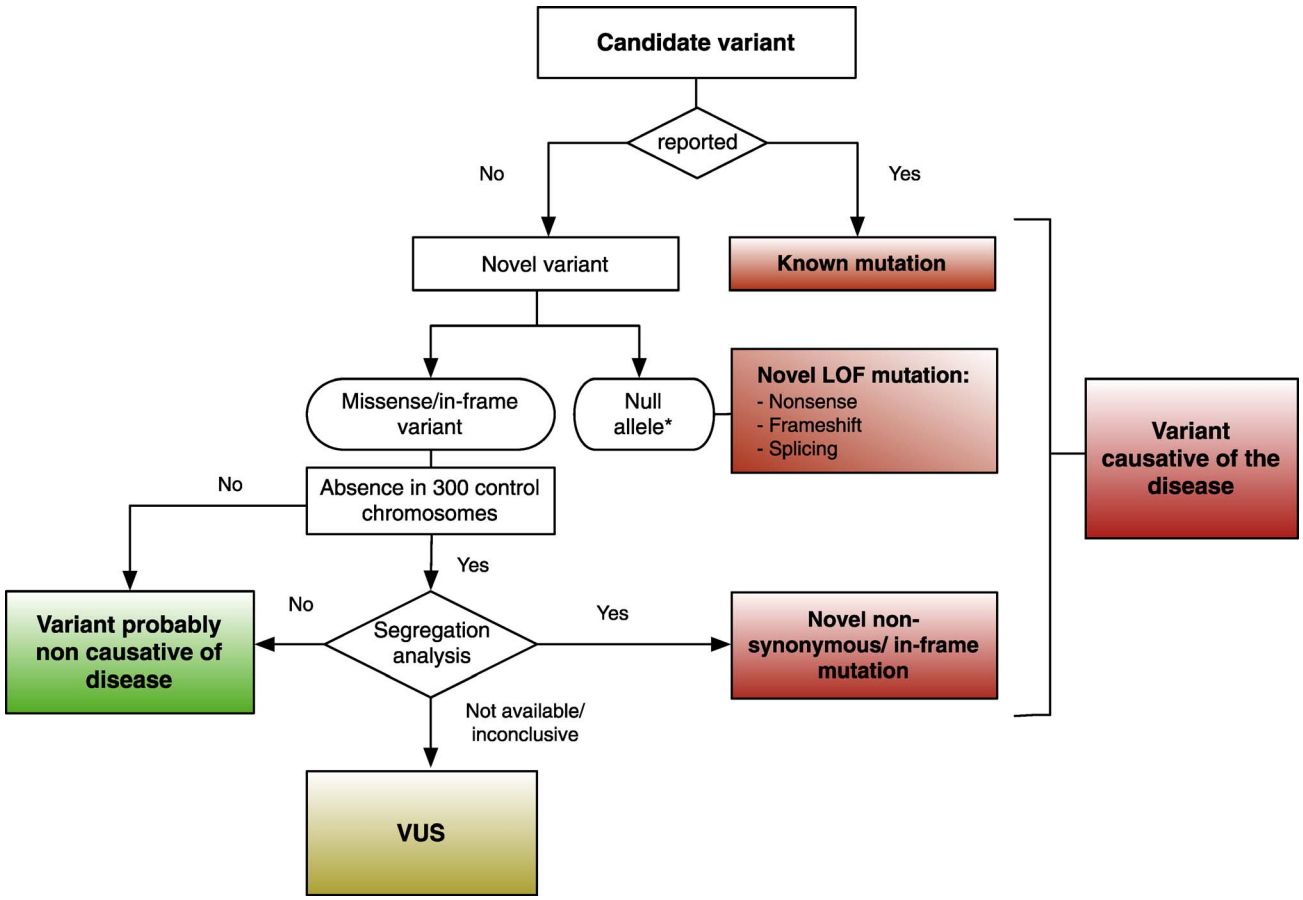

FIGURE 2. Algorithm applied for classification of candidate variants. *Cosegregation analysis was performed, when possible.

(LOF), nonsynonymous and in-frame mutations, (2) variants with uncertain significance (VUS): nonsynonymous and in-frame rare variants predicted as pathogenic with in silico tools, absent in 300 Spanish control population alleles and with unavailable or inconclusive cosegregation testing, and (3) variants probably not causative of disease (rare variants excluded as responsible of the disease through cosegregation studies and/or presence in at least one of 300 control population alleles (MAF: 0.003).

RESULTS

Analysis of Target NGS Approach

A custom RD_NGS_Panel for 73 genes was applied in 59 adRP index patients unsuccessfully studied by classical genetic tools. A total of 1236 regions were covered.

A mean of 2,915,152 reads in binary alignment/map (BAM) files was generated per sample with an average read length of 130 bp. From them, 2,217,542 reads (76.07%) were in analyzable target regions with an average and median read depth of 606 and 512, respectively. The 0.5% of target regions were covered less than 20 reads, and the 98.1% and 84.7% were covered by 100 and 500 reads, respectively.

Additionally, in order to check the reliability of our approach, 29 known variants in six different genes, previously detected by Sanger sequencing in 16 patients, were used as control samples and all of them were redetected by this technology (Supplementary Table S4).

A mean of 792 (±32) SNVs and indels were identified per sample. Removing off-target variants as intergenic or intronic resulted in a mean of 264 (±15) variants. Filtering by frequency data, a mean of 53 (±5) rare variants were selected. Removing synonymous and noncoding variants, a mean of 52 (±5) variants remained per sample (Fig. 1). Finally, considering an autosomal dominant model of inheritance and in silico prediction analysis, we identified a total of 31 candidate variants in 28 families. Of them, 68% (21/31) were missense variants, 10% (3/31) were frameshift changes, 10% (3/31) were stop codon mutations, 6% (2/31) were SNVs located in splice sites, and 6% (2/31) were in-frame deletions. All these variants were confirmed by Sanger sequencing. Additionally cosegregation analysis and population studies were performed in order to identify potential pathogenic mutations (Supplementary Figs. S1, S2).

Identification of Mutations and Spectrum of Variants

Known mutations previously reported related to adRP were identified in four families: *CRX* (p.Ala198Glyfs*38), *PRPF31* (c.1146+2T>C), *RHO* (p.Asn15Ser), and *SNRNP200* (p.Ser1087Leu), as shown in Table 1. Additionally, five families displayed novel putative LOF protein mutations: two nonsense mutations in *RP1* (p.Glu661* and p.Tyr915*); two frameshift variants: in *PRPF31* (p.Gly314Argfs*10) and *RP1* (p.Asn763Ilefs*12), and last, one splicing mutation in *PRPH2* (c.582-1G>A).

Moreover, another four families carried novel nonsynonymous mutations in *GUCA1B* (p.Arg44His), *IMPDH1* (p.Ala321V-

Table 1. Spectrum of Variants Causative of the Disease Identified With the RD_NGS_Panel in Autosomal Dominant Retinitis Pigmentosa Families

Family	Gene	NM	HGVS-cdna	HGVS-prot	Effect	Zyg	dbSNP (MAF)	SIFT	Polyphen-2	Mutation Taster	Human Splicing Finder	Cosegregation (Affected/ Unaffected/ Asymptomatic)	Frequency in Spanish Control Alleles	Reference
Known mutations														
RP-1875	*ABCA4**	NM_000350.2	c.3386G>T	p.Arg1129Leu	Missense	Het	rs1801269 (<0.01)	D (0)	Pr-D (0.992)	DC (1.000)	-	Yes (2/11)	0/150	Allikmets et al.,[28] 1997
			c.6148G>C	p.Val2050Leu	Missense	Het	rs41292677 (0.04)	D (0.01)	Pr-D (0.950)	DC (0.999)	-		0/150	Allikmets et al.,[29] 1997
RP-1688	*CRX*	NM_000554.4	c.586_587insC	p.Ala198Glyfs*38	Frameshift	Het	-	-	-	DC (1.000)	-	NA	0/150	Sohocki et al.,[30] 1998
RP-1970	*PRPF31*	NM_015629.3	c.1146+2T>C	-	Splicing	Het	-	-	-	DC (1.000)	Decrease 5′donor site of exon 11 (90.88->72.59)	Yes (2)	0/150	Waseem et al.,[31] 2007
RP-0642	*RHO*	NM_000539.3	c.44A>G	p.Asn15Ser	Missense	Het	rs104893786 (nfd)	D (0)	Pr-D (0.998)	DC (1.000)	-	Yes (1/2)	0/150	Kranich et al.,[32] 1993
RP-1480	*SNRNP200*	NM_014014.4	c.3260C>T	p.Ser1087Leu	Missense	Het	-	D (0)	Pr-D (0.995)	DC (1.000)	-	Yes (4/4/4)	0/150	Zhao et al.,[33] 2009
Novel LOF protein mutations														
RP-1541	*PRPF31*	NM_015629.3	c.937_938insA	p.Gly314Argfs*10	Frameshift	Het	-	-	-	DC (1.000)	-	Yes (2/2)	0/150	This study
RP-1176	*PRPH2*	NM_000322.4	c.582-1G>A	-	Splicing	Het	-	-	-	-	Decrease 3′acceptor site of exon 2 (82.97->54.03)	Yes (2)	0/150	This study
RP-2072	*RP1*	NM_006269.1	c.1981G>T	p.Glu661*	Nonsense	Het	-	-	-	DC (1.000)	-	NA	0/150	This study
RP-1890	*RP1*	NM_006269.1	c.2286delA	p.Asn763Ilefs*12	Frameshift	Het	-	-	-	DC (1.000)	-	Yes (2)	0/150	This study
RP-1387	*RP1*	NM_006269.1	c.2745_2749del	p.Tyr915*	Nonsense	Het	-	-	-	DC (1.000)	-	Yes (4/2)	0/150	This study
RP-0631	*RPGR**	NM_000328.2	c.1234C>T	p.Arg412*	Nonsense	Het	-	-	-	DC (1.000)	-	Yes (2)	0/150	This study
Novel nonsynonymous/in-frame mutations														
RP-1728	*GUCA1B*	NM_002098.5	c.131G>A	p.Arg44His	Missense	Het	-	D (0.004)	B (0.065)	DC (0.997)	-	Yes (2/1)	0/300	This study
RP-0422	*IMPDH1*	NM_000883.3	c.962C>T	p.Ala321Val	Missense	Het	-	D (0)	Pr-D (0.926)	DC (1.000)	-	Yes (3/11)	0/300	This study
RP-0652	*PRPH2*	NM_000322.4	c.536G>T	p.Trp179Leu	Missense	Het	-	D (0.010)	Pr-D (1.000)	DC (1.000)	-	Yes (2)	0/300	This study
RP-0948	*RP1*	NM_006269.1	c.4328G>A	p.Arg1443Gln	Missense	Het	-	D (0)	Pr-D (0.974)	P (0.993)	-	Yes (2/1)	0/300	This study
RP-1682	*RP2**	NM_006915.2	c.9_11del	p.Phe4del	In-frame	Het	-	-	-	DC (0.986)	-	Yes (1/3)	0/300	This study

LOF, loss of function; Zyg, zygosity; Het, heterozygosis; nfd, no frequency data; SIFT, deleterious (D); Polyphen2: probably damaging (Pr-D) and benign (B); Mutation Taster: disease causing (DC) and polymorphism (P); NA, not available.

* Gene associated with other inherited model different to dominant pattern.

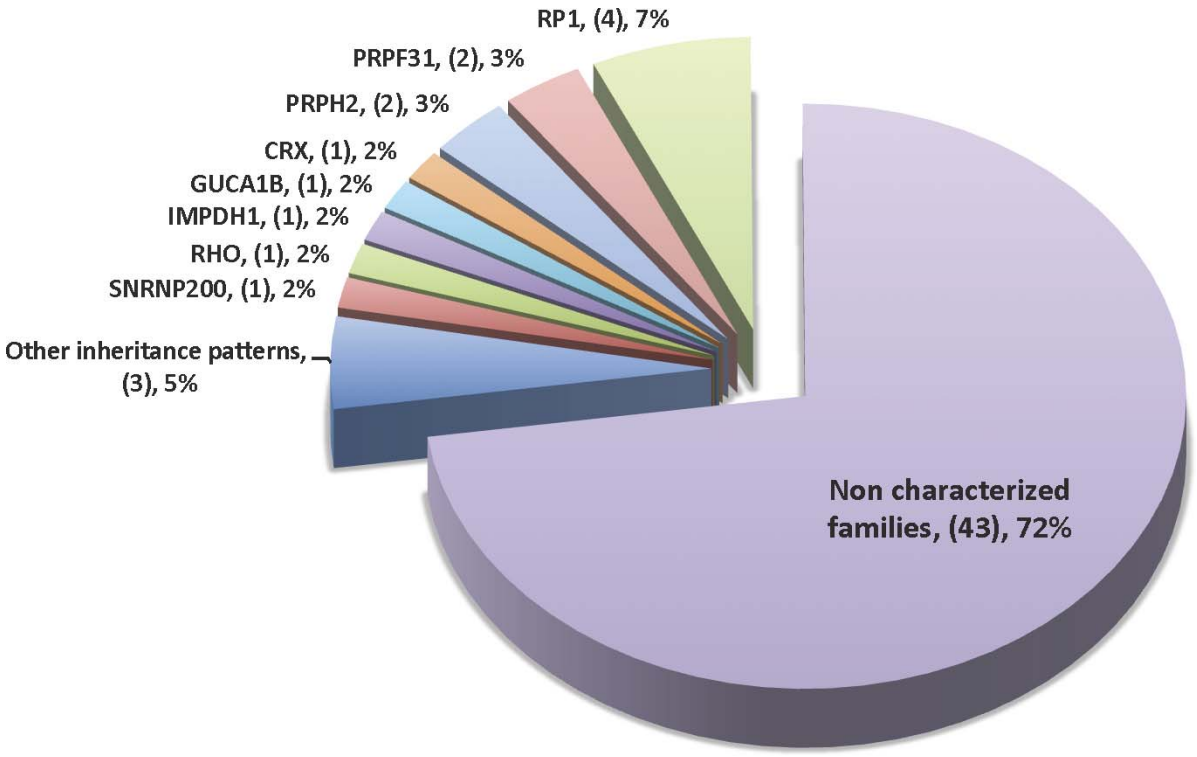

FIGURE 3. Percentage of families characterized with RD_NGS_Panel and spectrum of genes implicated in a cohort of 59 adRP families.

al), *PRPH2* (p.Trp179Leu), and *RP1* (p.Arg1443Gln). All of them were potentially pathogenic variants not previously reported in mutation or polymorphism databases, cosegregating with the disease in the family, predicted as pathogenic with in silico tools and excluded in at least 300 Spanish control chromosomes.

In addition, we found potentially deleterious mutations in other genes with a different model of inheritance. Two known compound heterozygous mutations (p.Arg1129Leu) and (p.Val2050Leu) in the *ABCA4* gene were detected in one family, and two novel mutations in the chromosome X genes, *RPGR* (p.Arg412*) and *RP2* (p.Phe4del). Those were also detected allowing us a genetic and/or even clinical reclassification of three families.

In brief, 16 families presented 17 different variants causative of the disease giving us a mutation detection rate of 27% (16/59) with our strategy. Additionally we could reclassify 5% (3/59) of cases. Those variants were known mutations in 36% (5/17) and the remaining 64% (12/17) were novel mutations.

We have found mutations in 9 of 23 adRP genes included in the RD_NGS_Panel. The frequency of RD causative genes detected in our cohort is represented in Figure 3. The most frequently mutated genes among our cohort were *RP1* found in 7% of families and *PRPH2* and *PRPF31* genes representing 3% of families each.

VUS and Variants Probably not Causative of Disease

Remarkably, 12% (7/59) of cases carried seven unreported nonsynonymous variants in *BEST1*, *FSCN2*, *GUCA1B*, *IMPDH1*, *PRPF8*, and two in *SNRNP200* (Table 2). Although all of them were predicted as very likely pathogenic and they were not present in our Spanish control population, unfortunately cosegregation analysis could not be performed or it was inconclusive to establish the causal relation in the family. Accordingly, they were classified as VUS (Supplementary Fig. S2).

Finally, seven nonsynonymous variants, detected in 10% of studied families (6/59), were classified as variants probably not causative of disease in *IMPG1*, *PRPF3*, *PRPF8*, *PRPH2*, *TOPORS*, and *USH2A* (two variants). Although all changes were predicted as likely pathogenic, variants found in four of these genes were finally excluded as the genetic cause of the disease through cosegregation studies (Supplementary Fig. S2).

In the other two cases, missense changes in*PRPF3* (p.Arg641Gln) and *PRPF8* (p.Met1140Thr) identified in two families, were found in one Spanish chromosome control (MAF: 0.003). Thus, they were considered as rare polymorphic variants in our population.

DISCUSSION

In this study we applied a novel NGS-based strategy to study 59 unrelated index adRP Spanish patients. This is one of the largest cohorts that has been studied by exon capture targeted NGS to date. Our results show 17 mutations detected in 27% of families tested (16/59). From them, 64% (11/17) were first reported here and 36% (6/17) were previously known mutations.

The known mutations detected in this study were not screened before (Supplementary Table S1), except for the known mutations identified in *CRX* and *PRPF31* genes that represented false positives for the ADRP Chip. Additionally, in family RP-0642, SSCP for the *RHO* gene and the genotyping ADRP Chip were performed in the affected member IV:4 instead of the index patient due to its better DNA quality. A new DNA sample of the proband was obtained and NGS was then performed and the mutation (p. Asn15Ser) in the *RHO* gene was found (Supplementary Fig. S1). This known mutation was not detected in IV:4 patient and a subsequent ophthalmic

Table 2. Spectrum of VUS and Variants Probably not Causative of Disease Identified with the RD_NGS_Panel in Autosomal Dominant Retinitis Pigmentosa Families

Family	Gene	NM	HGVS-cdna	HGVS-prot	effect	Zyg	dbSNP (MAF)	SIFT	Polyphen-2	Mutation Taster	Human Splicing Finder	Cosegregation (Affected/ Unaffected/ Asymptomatic)	Frequency in Spanish Control Alleles	Reference
Variant uncertain significance														
RP-1524	*BEST1*	NM_004183.3	c.1063C>T	p.R355C	Missense	Het	rs139637557 (nfd)	T (0.090)	Pr-D (1.000)	DC (1.000)	-	NA	0/300	-
RP-1684	*FSCN2*	NM_001077182.2	c.445C>T	p.Arg149Trp	Missense	Het	-	D (0)	Pr-D (1.000)	DC (0.999)	-	NA	0/300	-
RP-0992	*GUCA1B*	NM_000322.4	c.119A>G	p.His40Arg	Missense	Het	-	D (0)	Pr-D (1.000)	DC (1.000)	-	In (2/0/1)	0/300	-
RP-1903	*IMPDH1*	NM_000883.3	c.595_597delTTC	p.Phe199del	In-frame	Het	-	-	-	DC (0.999)	-	NA	0/300	-
RP-1187	*PRPF8*	NM_006445.3	c.6835T>G	p.Trp2279Gly	missense	Het	-	D (0)	Pr-D (1.000)	DC (1.000)	-	In (2/0/1)	0/300	-
RP-1641	*SNRNP200*	NM_014014.4	c.1627C>T	p.Pro543Ser	Missense	Het	-	D (0)	Pr-D (1.000)	DC (1.000)	-	NA	0/300	Eisenberger et al.,[8] 2013
RP-0723	*SNRNP200*	NM_014014.4	c.2579A>G	p.Gln860Arg	Missense	Het	-	D (0.030)	Ps-D (0.870)	DC (1.000)	-	In (2)	0/300	-
Variants probably not causative of disease														
RP-0251	*IMPG1*	NM_001563.2	c.1891G>A	p.Gly631Arg	Missense	Het	-	D (0)	Pr-D (1.000)	DC (1.000)	-	No	0/300	-
RP-0992	*PRPH2*	NM_000322.4	c.299C>T	p.Pro100Leu	Missense	Het	-	D (0.460)	B (0.018)	DC (1.000)	-	No	0/300	-
RP-0258	*PRPF3*	NM_004698.2	c.1922G>A	p.Arg641Gln	Missense	Het	rs201590000 (nfd)	D (0.06)	Pr-D (0.995)	DC (1.000)	-	Yes (3/1)	1/300	-
RP-1649	*PRPF8*	NM_006445.3	c.3419T>C	p.Met1140Thr	Missense	Het	.	D (0)	Pr-D (1.000)	DC (0.998)	-	NA	1/300	-
RP-1608	TOPORS	NM_005802	c.1730C>A	p.Ser577Tyr	Missense	Het	rs79708790 (0.001)	D (0.010)	Ps-D (0.910)	DC (0.981)	-	No	0/300	-
RP-1402	*USH2A*	NM_206933.2	c.1000C>T	p.Arg334Trp	Missense	Het	-	D (0.010)	Pr-D (1.000)	DC (1.000)	-	No	0/150	Adato et al.,[34] 2000
			c.2276G>T	p.Cys759Phe	Missense	Het	rs80338902 (<0.001)	D (0)	Pr-D (1.000)	DC (1.000)	-	No	0/150	Rivolta et al.,[35] 2000

Zyg, zygosity; Het, heterozygosis; nfd, no frequency data; SIFT: deleterious (D) and tolerated (T); Polyphen2: probably damaging (Pr-D), possibly damaging (Ps-D), and benign (B); Mutation Taster: disease causing (DC); P, polymorphism; NA, not available; In, inconclusive.

re-evaluation showed that this patient suffered from congenital stationary night blindness.

All candidate variants detected by NGS were confirmed by Sanger sequencing, therefore no false positives were found in this study. Moreover, all known variants used as positive controls were redetected by this technology. Both results show a high specificity and sensitivity that supports the reliability of our approach.

Due to its novelty, few reports have used NGS approaches to study adRP patients in the last 3 years. These works show a high variability of mutation detection between 23% and 86%.[5–11] The comparison is not straightforward because of differences in cohort size, studies previously made in patients, number of target genes, coverage rates, and other technical issues. Considering similar reports using cohorts of patients previously screened for mutations in the most prevalent genes, the detection mutation rates were similar to our results, approximately 25%[5,10] (Supplementary Table S5). These results probably underestimate the detection rate that these strategies can reach if naïve adRP cases are studied. Moreover, the detection of mutations in less frequent adRP genes was higher than expected for a first screening procedure (Fig. 3).

The inclusion of 73 genes related to RP and LCA independent of the model of inheritance to our NGS approach has had the benefit of allowing us the reorientation of the initial autosomal dominant classification in several families. In two of them, index patients were affected females carrying novel mutations in chromosome X whose pedigrees did not present male-to-male transmission. These findings are not surprising because mutations in the two X-linked RP genes (*RPGR* and *RP2*) have been described previously in 8.5% of families with provisional diagnosis of adRP with affected carrier females.[36] The third family was finally diagnosed as cone–rod dystrophy after detecting known compound heterozygous mutations in *ABCA4*, showing more a pseudodominant pattern of inheritance than a true dominant transmission.

The application of high-throughput sequencing approaches in Mendelian disorders has displayed several limitations particularly in the heterozygous context of dominant forms, as previously described.[37] First, it is not easy to differentiate between true heterozygous variants and false positives of the sequencing. Our NGS approach was designed to obtain high coverage in order to minimize sequencing mistakes described for resequencing technologies. Also, all exons of genes related to adRP included in the panel showed good sequencing depth. Secondly, to distinguish among rare pathogenic and rare benign variants, we have applied a very restrictive pipeline adapted to dominant pathology filtering variants with a MAF less than or equal to 0.001, as recommended previously[37] because all reported adRP mutations are very rare. Additionally, all candidate variants were specifically searched in population-matched controls. We consider that this step is critical to identify rare polymorphic variants that could not be reflected in 1000 genomes and EVS projects. Hence, in our study, all candidate variants detected at least once in our Spanish control cohort were considered as variants probably not causative of disease based on a large MAF. Additionally, cosegregation testing to evaluate the implication of variants is an important tool for dominant Mendelian diseases and sometimes DNA and phenotype information on appropriate family members are unavailable or inconclusive. In those cases, we consider such variants as VUS.

This is the case of the novel missense change (p.Gln860Arg) in *SNRNP200* detected in family RP-0723 that could only be checked in two affected siblings. Moreover, the interpretation of cosegregation results can be also complicated in adRP families with suspected incomplete penetrance. Thus, for the family RP-0992 with the variant p.His40Arg found in *GUCA1B* gene, one mutated member is asymptomatic at 31 years of age, while another affected in the same family showed first symptoms at 12 years of age. Although ophthalmologic follow-up is needed, the occurrence of incomplete penetrance associated to *GUCA1B* mutations has been described previously,[38] therefore we cannot completely exclude that this mutation can be causative of RP in this family.

Cosegregation analysis assuming an autosomal dominant model with complete penetrance, zero recombination fraction, and no phenocopies employed in this work, is equivalent to linkage analysis. Using two-point linkage analysis with more complex trait models could help resolve the probability of causality of variants with possible incomplete penetrance, at least in moderately large families. Although this event has been previously associated with *PRPF8*,[39] the variant p.Trp2279Gly detected in RP-1187 family should be studied in other members of the family to establish the true implication of this change in the disease.

Lastly, the nonsynonymous variant in *PRPH2* (p.Pro100Leu) that putatively induced a critical structural change in the cytoplasmatic domain, was not detected in the affected mother of family RP-0992. However, it was not possible to discard a "de novo" event with two different molecular causes of the disease in this family.

Finally, the information of some variants and mutations in public databases could be incomplete or even misinterpreted.[40] For example the change p.Gln860Arg in *SNRNP200* reported in HGMD Professional database (available in the public domain at www.biobase-international.com/product/hgmd) as mutation, originally was published as VUS in heterozygosis in one case with autosomal recessive LCA.[8] Accordingly, we considered also this variant detected in the index patient of RP-1641 adRP family as VUS. Therefore, the pathogenicity of all variants described previously must be confirmed in original articles.

There are some restrictions in our NGS-based design that could explain the absence of detection of causal mutations in the remaining studied patients. First, novel deep intronic mutations would be missed out because these regions were not included in the panel with the aim to do a cost-effective approach. In addition, some families could carry mutations in *PRPF4* and *HK1*, two genes recently identified associated with adRP and not included in our design.[41,42] Further, up to 2.5% of adRP mutations could be due to genomic rearrangements or deletions in *PRPF31* that could also explain a minor percentage of adRP cases, as previously reported.[43] However, they were not detectable with our bioinformatic analysis. Finally, the presence of the locus RP63 that has been mapped by linkage analysis but for which the causal gene has not yet been identified, suggests that mutations in other genes remain to be discovered.[44]

This work shows that our NGS panel is an effective approach to detect mutations in families affected by adRP and is a feasible tool for application in clinical practice in order to offer a correct diagnosis and genetic counselling. Additionally, our results provide relevant information for mutations and variants found in a large cohort of adRP Spanish patients that increase the knowledge about molecular findings related to RD.

Acknowledgments

The authors thank the Services of Genetics and Ophthalmology IIS-Fundacion Jimenez Diaz (Madrid).

Supported by grants from FIS (PI: 13/00226; ISCIII, Madrid, Spain), the Centre for Biomedical Network Research on Rare Diseases - CIBERER (06/07/0036; ISCIII, Madrid, Spain), the Biobank of Fundacion Jimenez Diaz University Hospital (RD09/0076/00101;

ISCIII, Madrid, Spain), ONCE 2014 and Fundaluce (4019-002; Madrid, Spain), a Rio Hortega Grant (CM12/00013; PFS; Madrid, Spain), a Servet Grant (CP/03256; MC; ISCIII, Madrid, Spain), a Sara Borrell Grant (CD12/00676 [RSA] and CD13/00085 [ISN]; Madrid, Spain), all from Instituto de Salud Carlos (ISCIII), and the CIBERER (Madrid, Spain [AAF and OZ]).

Disclosure: **P. Fernandez-San Jose**, None; **M. Corton**, None; **F. Blanco-Kelly**, None; **A. Avila-Fernandez**, None; **M.A. Lopez-Martinez**, None; **I. Sanchez-Navarro**, None; **R. Sanchez-Alcudia**, None; **R. Perez-Carro**, None; **O. Zurita**, None; **N. Sanchez-Bolivar**, None; **M.I. Lopez-Molina**, None; **B. Garcia-Sandoval**, None; **R. Riveiro-Alvarez**, None; **C. Ayuso**, None

References

1. Hamel C. Retinitis pigmentosa. *Orphanet J Rare Dis*. 2006;1:40.
2. Hartong DT, Berson EL, Dryja TP. Retinitis pigmentosa. *Lancet*. 2006;368:1795–1809.
3. Ayuso C, Garcia-Sandoval B, Najera C, Valverde D, Carballo M, Antiñolo G. Retinitis pigmentosa in Spain. The Spanish Multicentric and Multidisciplinary Group for Research into Retinitis Pigmentosa. *Clin Genet*. 1995;48:120–122.
4. Daiger SP, Bowne SJ, Sullivan LS. Perspective on genes and mutations causing retinitis pigmentosa. *Arch Ophthalmol*. 2007;125:151–158.
5. Bowne SJ, Sullivan LS, Koboldt DC, et al. Identification of disease-causing mutations in autosomal dominant retinitis pigmentosa (adRP) using next-generation DNA sequencing. *Invest Ophthalmol Vis Sci*. 2011;52:494–503.
6. Audo I, Bujakowska KM, Léveillard T, et al. Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J Rare Dis*. 2012;1–17.
7. Sullivan JO, Mullaney BG, Bhaskar SS, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J Med Genet*. 2012;49:322–327.
8. 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.
9. Glöckle N, Kohl S, Mohr J, et al. Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *Eur J Hum Genet*. 2014;22:99–104.
10. Daiger SP, Bowne SJ, Sullivan LS, et al. Application of next-generation sequencing to identify genes and mutations causing autosomal dominant retinitis pigmentosa (adRP). *Adv Exp Med Biol*. 2014;801:123–129.
11. Oishi M, Oishi A, Gotoh N, et al. Comprehensive molecular diagnosis of a large cohort of Japanese retinitis pigmentosa and usher syndrome patients by next-generation sequencing. *Invest Ophthalmol Vis Sci*. 2014;55:7369–7375.
12. Hamel CP. Gene discovery and prevalence in inherited retinal dystrophies. *C R Biol*. 2014;337:160–166.
13. Blanco-Kelly F, García-Hoyos M, Cortón M, et al. Genotyping microarray: mutation screening in Spanish families with autosomal dominant retinitis pigmentosa. *Mol Vis*. 2012;18:1478–1483.
14. Benaglio P, Fernandez-San Jose P, Avila-Fernandez A, et al. Mutational screening of splicing factor genes in cases with autosomal dominant retinitis pigmentosa. *Mol Vis*. 2014;20:843–851.
15. Manes G, Meunier I, Avila-Fernández A, et al. Mutations in IMPG1 cause vitelliform macular dystrophies. *Am J Hum Genet*. 2013;93:571–578.
16. Fernandez-San Jose 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*. 2014;93:e38–e44.
17. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. 2010;11:R86.
18. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589–595.
19. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303.
20. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–498.
21. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27:863–864.
22. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–842.
23. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010;26:2069–2070.
24. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073–1081.
25. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7:575–576.
26. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–249.
27. Corton M, Tatu SD, Avila-Fernandez A, et al. High frequency of CRB1 mutations as cause of early-onset retinal dystrophies in the Spanish population. *Orphanet J Rare Dis*. 2013;8:20.
28. Allikmets R, Shroyer NF, Singh N, et al. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. *Science*. 1997;277:1805–1807.
29. Allikmets R. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet*. 1997;17:122.
30. Sohocki MM, Sullivan LS, Mintz-Hittner HA, et al. A range of clinical phenotypes associated with mutations in CRX, a photoreceptor transcription-factor gene. *Am J Hum Genet*. 1998;63:1307–1315.
31. 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.
32. Kranich H, Bartkowski S, Denton MJ, et al. Autosomal dominant "sector" retinitis pigmentosa due to a point mutation predicting an Asn-15-Ser substitution of rhodopsin. *Hum Mol Genet*. 1993;2:813–814.
33. Zhao C, Bellur DL, Lu S, et al. Autosomal-dominant retinitis pigmentosa caused by a mutation in SNRNP200, a gene required for unwinding of U4/U6 snRNAs. *Am J Hum Genet*. 2009;85:617–627.
34. Adato A, Weston MD, Berry A, Kimberling WJ, Bonne-Tamir A. Three novel mutations and twelve polymorphisms identified in the USH2A gene in Israeli USH2 families. *Hum Mutat*. 2000;15:388.
35. Rivolta C, Sweklo EA, Berson EL, Dryja TP. Missense mutation in the USH2A gene: association with recessive retinitis

pigmentosa without hearing loss. *Am J Hum Genet*. 2000;66: 1975–1978.

36. Churchill JD, Bowne SJ, Sullivan LS, et al. Mutations in the X-linked retinitis pigmentosa genes RPGR and RP2 found in 8.5% of families with a provisional diagnosis of autosomal dominant retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2013;54: 1411–1416.
37. Daiger SP, Bowne SJ, Sullivan LS. Genes and mutations causing autosomal dominant retinitis pigmentosa [published online ahead of print October 10, 2014]. *Cold Spring Harb Perspect Med*. doi: 10.1101/cshperspect.a017129.
38. Sato M, Nakazawa M, Usui T, Tanimoto N, Abe H, Ohguro H. Mutations in the gene coding for guanylate cyclase-activating protein 2 (GUCA1B gene) in patients with autosomal dominant retinal dystrophies. *Graefes Arch Clin Exp Ophthalmol*. 2005;243:235–242.
39. Maubaret CG, Vaclavik V, Mukhopadhyay R, et al. Autosomal dominant retinitis pigmentosa with intrafamilial variability and incomplete penetrance in two families carrying mutations in PRPF8. *Invest Ophthalmol Vis Sci*. 2011;52:9304–9309.
40. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014;508:469–476.
41. Chen X, Liu Y, Sheng X, et al. PRPF4 mutations cause autosomal dominant retinitis pigmentosa. *Hum Mol Genet*. 2014;23:2926–2939.
42. Wang F, Wang Y, Zhang B, et al. A missense mutation in HK1 leads to autosomal dominant retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2014;55:7159–7164.
43. 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.
44. Kannabiran C, Pal H. Mapping of locus for autosomal dominant retinitis pigmentosa on chromosome 6q23. *Hum Genet*. 2012;131:2012;717–723.