Novel Mutations in PRPF31 Causing Retinitis Pigmentosa Identified Using Whole-Exome Sequencing

Xiaoqiang Xiao,1 Yingjie Cao,1 Zhun Zhang,1 Yanxuan Xu,1 Yuqian Zheng,1 Li Jia Chen,1,2 Chi Pui Pang,1,2 and Haoyu Chen1

1Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, Shantou, China
2Department of Ophthalmology and Visual Sciences, the Chinese University of Hong Kong, Hong Kong, China

Correspondence: Haoyu Chen, Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, North Dongxia Road, Shantou 515041, China; drchenhaoyu@gmail.com.

XX, YC, and ZZ contributed equally to the work presented here and should therefore be regarded as equivalent authors.

Submitted: September 11, 2017
Accepted: November 8, 2017

PURPOSE. The purpose of this study was to investigate the disease-causing mutations for retinitis pigmentosa (RP) patients and function of mutations.

METHODS. We recruited RP families and sporadic RP patients, and performed whole-exome sequencing (WES) to screen for sequence variations. Subsequently, we investigated the expression of green fluorescent protein (GFP) merged expression vectors containing PRPF31 wild type (WT) and its variants. We determined protein stability by cycloheximide (CHX) treatment.

RESULTS. Two frameshift variants, c.547delG (p.E183fs) and c.804delG (p.L268fs), and one stopgain variant, c.1060C>T (p.R354X), in the pre-mRNA processing factor 31 gene (PRPF31) were identified in three RP families. In comparison with WT, the expressions of GFP-fused PRPF31 (GFP-PRPF31) protein with the mutation c.547delG or c.804delG in HEK293 cells were significantly reduced. However, the expression of GFP-PRPF31 containing the stopgain mutation (GFP-PRPF31sg) was increased. CHX treatment of HEK293 showed the GFP-PRPF31sg protein was more stable than GFP-PRPF31 WT. The WT protein expression was localized in the nuclei, and the mutants in both nuclei and cytoplasm. We screened for PRPF31 mutations in 131 sporadic RP patients by WES and successfully identified three novel mutations: c.G781C (p.G261R), c.A1373T (p.Q458L), and c.C1222T (p.R408W).

CONCLUSIONS. Our study revealed novel mutations of PRPF31 in RP. Our results also showed that the two mutations (c.547delG or c.804delG) affect gene expression and GFP-PRPF31sg has increased protein stability.

Keywords: retinitis pigmentosa, PRPF31, whole-exome sequencing, mutation, localization

Retinitis pigmentosa (RP) is a group of inherited diseases that usually start with night blindness and progress to visual field loss and even to irreversible blindness.1 Although RP exhibits clinical and genetic heterogeneities, the dysfunction and death of photoreceptors in retina is a common step in the disease progression that leads to vision dysfunction.2 Genetically, RP is classified into autosomal dominant, autosomal recessive, and X-linked. There are also a small number of digenic and mitochondrial patients. RP patients without clear family history are usually referred to as simplex RP.3 More and more RP mutations that directly cause the disease are identified over the years, and in recent years with the advancement of technologies, including next-generation sequencing, more than 80 RP genes have been identified.4 However, there are still many RP cases in whom the disease-causing mutations have not been identified. Novel genes and mutations are still to be discovered and characterized to help our understanding of the genetic architecture and pathogenesis of RP.

As a transformative technology, exome sequencing is a strong tool in generating unbiased genetic data on a whole-genome scale. Identification and functional studies of disease-causing mutations, especially in the coding exons and splicing sites, can lead to important information on the molecular pathogenesis of the disease and help in designing effective therapies. To identify potential genetic etiology of RP, we adopted whole-exome sequencing (WES) and bioinformatics to recruited patients with RP, including seven affected members from three RP families and 131 sporadic RP patients. We identified novel mutations in the PRPF31 gene. We then analyzed the effects of the PRPF31 mutations on protein expression, stability and cellular localization in the HEK293 or A549 cells.

MATERIALS AND METHODS

Samples Collection, Clinical Examination

The study was approved by the Ethics Committee of the Joint Shantou International Eye Center (JSIEC), Shantou University and the Chinese University of Hong Kong, Shantou, Guangdong province, China, and conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from every study subject before recruitment. Peripheral blood samples from three Chinese RP families, 131 sporadic RP patients, and 200 healthy control individuals were collected in JSIEC. Clinical examinations for all study subjects included visual acuity testing, visual field test, slit-lamp examination, fundus photography, optical coherence tomography (OCT),
Novel Mutations in PRPF31 Causing RP

WHOLE-EXOME SEQUENCING

Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA of seven affected individuals from the three RP families, and 131 sporadic RP patients were subjected to WES using the TIANGEN Blood DNA Kit DP318 (TIANGEN, Beijing, China) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manufacturer’s instructions. Genomic DNA of all subjects was extracted from whole blood using the Promega Blood DNA Kit DP318 (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Genomic DNA was fragmented and library construction was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina with a fragment size of approximately 300 bp. Library quality control was performed using the Agilent Bioanalyzer 2100 and BioSem V Base Station. The libraries were sequenced on the HiSeq 2500 PE100 (Illumina, San Diego, CA, USA) according to the manuf
Chinese sporadic RP patients also present the RP phenotype (Supplementary Table S3).

**Mutation Analysis**

To identify potential causative mutations, we performed WES in seven members of the three RP families. A total of 70 GB of sequence data were generated and 95.5% of bases originated from the exon region, resulting in a mean coverage of 100-fold. The identified total SNPs (missense, nonsense, and splice site mutations) and insertions/deletions (short coding insertions or deletions) were 24135 for IV2, 24098 for IV5, and 24599 for VI1 in RP-F1; 23857 for II5 in RP-F2 and 25114 for V6, 24414 for IV6, 25092 for IV15 in RP-F3, respectively (Fig. 3). After the removal of common variants, synonymous single nucleotide variants (SNVs), the candidate SNPs/insertions/
deletions of known RP genes were reduced to two for RP-F1, seven for RP-F2, and one for RP-F3, respectively (Fig. 3). We further excluded the nonpathogenic variants based on SIFT/PolyPhen-2/MutationTaster analysis. Finally, we identified two for RP-F1, three for RP-F2, and one for RP-F3, respectively (Fig. 3).

After inheritance analysis, we detected several heterozygous variants that can be candidate genes for (1) family RP-F1: PRPF31 (NM_015629) c.547delG, and PROM1 (NM_001145851) c.C2288T; (2) family RP-F2: PROM1 (NM_001145851) c.C2431G, EYS (NM_001142800) c.C7679T, and PRPF31 (NM_015629) c.804delG; and (3) family
Novel Mutations in PRPF31 Causing RP

RP-3: PRPF31 (NM_015629.3) c.1060C>T (Table 1). After Sanger sequencing the individuals from the families, we found the mutations from PROM1, EYS did not cosegregate with the RP phenotype. Only variants c.547delG (p.E183fs), c.804delG (p.L268fs), and c.1060C>T (p.R354X), located at the seventh exon, and the c.804delG (p.L268fs), at the eighth exon, which both encode an NOP domain of PRPF31 protein (Fig. 5A). The NOP domain is responsible for the RNA binding (Fig. 5B). The stopgain mutant c.1060C>T (p.R354X), located at the 10th exon, encodes the nuclear localization sequence of PRPF31 (Figs. 5A, 5B). To determine the expression of PRPF31 in vivo, we extracted expression data from the public accessible database GenePaint (http://www.genepaint.org, in the public domain). The results showed PRPF31 expresses in all parts of the mouse embryo, including pretectum, with high concentration (Fig. 5C). Multiple orthologous protein sequence alignment revealed that all the three variants identified in three RP families are highly conserved from humans to fish (Fig. 5D). The prediction in silico showed the three mutations are damaging to its protein function (Table 2). The stopgain mutant located at the NOP domain, also showed very high conservation. Therefore, we consider those mutants as novel mutations. Mutants from the sporadic RP patients, of which Q568L and R408W both are located at the C-terminus of PRPF31 protein, are absent in both the existing database of disease-causing mutations and in the reported literature, we thereby considered those mutants as novel mutations. Mutants from the sporadic RP patients, of which Q568L and R408W both are located at the C-terminus of PRPF31 protein, are probably benign and are absent in both the reported literatures.

Through analyzing PRPF31 genomic and protein structure, we found the deletion mutant c.547delG (p.E183fs), located at the seventh exon, and the c.804delG (p.L268fs), at the eighth exon, which both encode an NOP domain of PRPF31 protein (Fig. 5A). The NOP domain is responsible for the RNA binding (Fig. 5B). The stopgain mutant c.1060C>T (p.R354X), located at the 10th exon, encodes the nuclear localization sequence of PRPF31 (Figs. 5A, 5B). To determine the expression of PRPF31 in vivo, we extracted expression data from the public accessible database GenePaint (http://www.genepaint.org, in the public domain). The results showed PRPF31 expresses in all parts of the mouse embryo, including pretectum, with high concentration (Fig. 5C). Multiple orthologous protein sequence alignment revealed that all the three variants identified in three RP families are highly conserved from humans to fish (Fig. 5D). The prediction in silico showed the three mutations are damaging to its protein function (Table 2). The stopgain mutant, which has been previously reported, is damaged in PRPF31 protein function as predicted by the above-stated protocols. The other two deletion mutations are absent in the database and published reports and are probably benign variations. Therefore, we consider these two deletion mutants as novel mutations.

To examine whether additional mutations in PRPF31 could be identified in unrelated Chinese subjects, we enrolled 131 sporadic RP patients and sequenced their genomic DNA by WES. Through the same filtering procedures, we identified another three heterozygous missense variants: c.G781C:p.G261R, c.A1373T:p.Q458L, c.C1222T:p.R408W (Supplementary Table S3). These missense mutations were not shown in the in-house Chinese exome database, and they were predicted to be harmful by the three programs, as mentioned above. After Sanger sequencing for 200 unrelated ethnically matched healthy controls, we did not find all three. These variants are also absent in the database of a probably benign variation, which is consistent with the results predicted by informational analysis. Because these mutants are absent in both the existing database of disease-causing mutations and in the reported literature, we thereby considered those mutants as novel mutations. Mutants from the sporadic RP patients, of which Q568L and R408W both are located at the C-terminus of PRPF31 protein, are probably benign and are absent in both the reported literatures.

Mutations Changed the Expression, Stability, and Localization of PRPF31

To examine whether the variants c.547delG, c.804delG, and c.1060C>T affect PRPF31 expression and localization in the cells, we constructed GFP-fused PRPF31 wild type (GFP-PRPF31) (WT). Using site-directed mutagenesis, we constructed GFP-fused PRPF31 mutant expression vectors. As shown in Figure 6A, when those constructs were transfected into HEK293 cells, the GFP-PRPF31 (WT) demonstrated high GFP expression. However, deletion mutant c.547delG of PRPF31 failed to express the GFP-fused PRPF31 protein, and the mutant c.804delG showed reduced GFP expression in comparison with the GFP-PRPF31 (WT). Interestingly, compared with the GFP-PRPF31 (WT), the GFP-PRPF31sg showed a higher level GFP expression (Fig. 6A). The expression from above constructs was further confirmed by immunoblotting (Fig. 6B). Because the transcriptions of GFP-PRPF31 (WT) or its mutants were controlled under the same promoter, transcription-mediated enhancement in their expressions was excluded. To explain the potential mechanism underlining this phenomenon that GFP-PRPF31sg has a higher level of protein than that of GFP-PRPF31 (WT), we treated HEK293 cells with 100 μM protein synthesis inhibitor CHX for designated time points (Fig. 6C). The immunoblotting was performed to detect change of protein. Our results showed the protein started to decrease at 36 hours for GFP-PRPF31sg and at 12 hours for GFP-PRPF31 (WT), after the inhibition of protein synthesis with CHX, suggesting the protein of GFP-PRPF31sg is more stable than that of WT (Fig. 6C).

Structurally, the stopgain mutant located at the 10th exon encoding a nuclear localization sequence (NLS) motif. Therefore, we compared the cellular localization of WT and stopgain mutant of PRPF31 in HEK293 and A549, a non–small-cell lung cancer cell line. We found, in HEK293, nuclear localization of GFP-PRPF31 (WT) while in A549 at both nuclear and plasma localization, although the concentration of protein in nuclei was much higher than that in plasma. The
TABLE 1. Identified Variants, Population Frequencies, and In Silico Predictions of Pathogenic Functions

<table>
<thead>
<tr>
<th>Exonic Variant</th>
<th>ExAC Frequency</th>
<th>Polyphen2 HDIV</th>
<th>Polyphen2 LRT</th>
<th>Polyphen2 MA</th>
<th>SIFT Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inheritance</strong></td>
<td><strong>Previously Reported</strong></td>
<td><strong>Exon Amino Acid Change</strong></td>
<td><strong>Variant</strong></td>
<td><strong>Effect</strong></td>
<td><strong>Score</strong></td>
</tr>
<tr>
<td><strong>Family</strong></td>
<td><strong>Gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AD, autosomal dominant; AR, autosomal recessive; D, damaging; ExAC, the Exome Aggregation Consortium; HDIV, likelihood ratio test; LRT, Mutation Taster; MA, Mutation Assessor; MA, medium damaging; NLS (RKKRGGRRYRKMKER) 18; MSR1, a minisatellite repeat element (MSR1) 8, which is adjacent to the core promoter of PRPF31, and CNOT3, which is a cis-acting transcriptional factor. 9 Play an important role in the regulation of incomplete penetrance of PRPF31 via modulating its mRNA transcription.

**DISCUSSION**

In this study, we found three variants of PRPF31, showing disease-causing capability, after sequencing seven family members from three RP families using WES. Families RP-F1 and RP-F2 show a classical autosomal dominant inheritance among generations. However, in family RP-F3, one of the mutation carriers (IV12) has no obvious RP pathologic change in retina except a history of 2-year night blindness. We therefore considered RP-F3 as an incomplete penetrance inheritance (Fig. 1). The percentage of penetrance is 100% in family 1 and family 2 and 87.5% in family 3. Previous reports have shown incomplete penetrance of PRPF31 in causing RP, 7 which is consistent with our RP-F3 family. It was reported that some transcriptional factors, such as the minisatellite repeat element (MSR1), 8 which is adjacent to the core promoter of PRPF31, and CNOT3, which is a cis-acting transcriptional factor, 9 play an important role in the regulation of incomplete penetrance of PRPF31 via modulating its mRNA transcription.

PRPF31 forms a bridge for U4/U6 and U5 by interacting with PRPF4, and thus benefits the U4/U6/U5 complex formation and stabilization. 10,11 PRPF31 is a ubiquitous protein; however, mutation in PRPF31 only links to RP and the molecular mechanism is largely unknown. 12 There are several models to explain this phenomenon. First, mutations in pre-mRNA splicing genes only influence the splicing of retina-specific genes, and second, the haploinsufficiency model, which mainly focuses on the small nuclear ribonucleoprotein (snRNP) concentration. Third, there is a model called the cell type-dependent snRNP assembly disorder caused by pre-mRNA splicing gene mutation. 3,13,14 The premature termination codons caused by frameshift mutation in PRPF31 frequently result in haploinsufficiency, as their corresponding mRNA is degraded by a mechanism called nonsense-mediated mRNA decay (NMD). 15 The two identified frameshift mutations, located at the small nucleolar RNA binding domain (NOP) of PRPF31, which is responsible for the binding of pre-mRNA 16,17 (Fig. 4), cause the protein frame shift at p.183 and p.268, respectively. These frameshifts result in early termination at p.195 after coding for 12 missense amino acids and p.319 for 51 amino acids, respectively. Based on the results of immunoblotting and immunofluorescence, we found the GFP-PRPF31 (WT) protein was expressed at a higher quantity than GFP-PRPF31 with c.547delG or c.804delG mutation (Fig. 6). These results indicate the mRNA transcripts from those two deletion mutants might be NMD-sensitive and therefore were degraded by this mechanism. The stopgain mutation (c.1060C>T[p.R354X]), located at the 10th exon, was reported before in an autosomal dominant family. 6 However, our current RP family is an incomplete penetrance family. Therefore, the penetrance rate of RP-F31 caused by this mutation might be family-dependent. Interestingly, the expression level of GFP-PRPF31sg (c.1060C>T[p.R354X]) was higher than that of GFP-PRPF31(WT). The enhancement of expression in protein is not due to the increase of transcriptional activity, but the protein stability (Fig. 6). Our current results are contradictory with the haploinsufficiency model (Fig. 6). Therefore, to further understand the potential mechanism, more investigations are required in protein function studies and pathway investigations in the tissue cells. PRPF31 localizes at nuclei through binding to the CTNNBL1 (catenin-like 1), with its NLS (RKKRGGRRYRKMKER). 18 We found, in HEK295 cells,
the GFP-PRPF31 (WT) localized in the nuclei, consistent with a previous report, whereas GFP-PRPF31sg (p.R354X) is localized both in nuclei and plasma.

In addition, we identified three PRPF31 missense variants in 131 sporadic RP cases. Although these variants are absent in our controls and available databases, and predicted to be possibly damaging based on in silico prediction, it is still questionable whether these missense variants are pathogenic or not. Further studies investigating the function of the above mutants are needed to understand their roles in the pathogenesis of RP. Furthermore, copy number loss of PRPF31 was also reported to cause RP. Therefore, we here cannot exclude this possibility in explaining our results. We will investigate the copy number variations of PRPF3 in our further studies.

In summary, we have identified six variants in PRPF31, three of them from the RP families and three from the sporadic RP patients. All the variants were predicted to be damaging to protein function. Our study has expanded the known mutation spectrum of PRPF31 in RP. We also determined their protein expression and localization of three mutants from RP families.

**FIGURE 4.** Sanger sequencing confirmation of three PRPF31 mutations identified in our pedigrees.

**FIGURE 5.** PRPF31 mutant, structure, and its expression in mouse embryo. (A) Variants, genomic structure, encoded protein with designated domain and motif of PRPF31; blue square, exon. (B) Three-dimensional structural model for NOP domain and their binding with RNA. (C) Sagittal in situ hybridization image in a mouse embryo at embryonic day 14.5 (GenePaint) shows PRPF31 expression in the neocortex, superior colliculus, and pretectum of the midbrain. (D) Multiple sequence alignment shows the evolutionary conservation of the indicated residues. The RP caused by PRPF31 mutations has been identified in the seven species listed. h, human; mo, *Mus musculus*; pt, *Pan troglodytes*; rn, *Rattus norvegicus*; xt, *Xenopus tropicalis*; fa, *Fopius arisanus*; dr, *Danio rerio*.
TABLE 2. Putative Pathogenic Variants Identified in PRPF31 Gene From Three RP Families

<table>
<thead>
<tr>
<th>ID</th>
<th>Ancestry, Age of Diagnosis, Sex</th>
<th>PRPF31 Variants</th>
<th>Protein Consequence</th>
<th>Frequency of Variant in Control and ExAC</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP-F1-II2</td>
<td>Chinese, 49, male</td>
<td>C.547delG</td>
<td>p.E183fs</td>
<td>NA</td>
<td>Damaging</td>
</tr>
<tr>
<td>RP-F1-II5</td>
<td>Chinese, 44, female</td>
<td>c.804delG</td>
<td>p.L268fs</td>
<td>NA</td>
<td>Damaging</td>
</tr>
<tr>
<td>RP-F1-VI1</td>
<td>Chinese, 8, female</td>
<td>c.1060C&gt;G</td>
<td>p.R354X</td>
<td>NA</td>
<td>Damaging</td>
</tr>
<tr>
<td>RP-F3-V6</td>
<td>Chinese, 17, male</td>
<td>c.1060C&gt;G</td>
<td>p.R354X</td>
<td>NA</td>
<td>Damaging</td>
</tr>
<tr>
<td>RP-F3-IV6</td>
<td>Chinese, 67, male</td>
<td>c.1060C&gt;G</td>
<td>p.R354X</td>
<td>NA</td>
<td>Damaging</td>
</tr>
</tbody>
</table>

FIGURE 6. Expression analysis of GFP-PRPF31 (WT) and its mutants. (A) Expression analysis of GFP-PRPF31 (WT) and its mutants in HEK293 cells by inverted fluorescence microscopy. (B) Expression analysis of GFP-PRPF31 (WT) and its mutants in HEK293 cells by Western blotting (WB). Left: 30 μg total protein, right: 80 μg protein without stopgain mutant. (C) Protein stability analysis after CHX treatment in designated time points by WB. The targeted plasmids were transfected into HEK293 for 24 hours, and then cells were used for direct observation with inverted fluorescence microscope, or lysate with 1×NP40 lysis buffer for WB, or continued treatment with 100 μM CHX for the designated time points. Then, the CHX-treated cells were used for WB analysis to detect amount of GFP-PRPF31 and GFP-PRPF31sg protein.
in A549 and HEK293 cells and confirmed these mutations could affect its protein stability and localization.

Acknowledgments

Supported by the National Nature Science Foundation of China (30901646 and 81170853), YangFan Program, and TeZhi Program of Guangdong Province.

Disclosure: X. Xiao, None; Y. Cao, None; Z. Zhang, None; Y. Xu, None; Y. Zheng, None; L.J. Chen, None; C.P. Pang, None; H. Chen, None

References


