A novel mutation in PRPF31, causative of autosomal dominant retinitis pigmentosa, using the BGISEQ-500 sequencer

Yu Zheng1,2,3, Hai-Lin Wang4, Jian-Kang Li2,3, Li Xu4, Laurent Tellier2,3, Xiao-Lin Li4, Xiao-Yan Huang1,2,3, Wei Li2,3, Tong-Tong Niu4, Huan-Ming Yang2,5, Jian-Guo Zhang2,3, Dong-Ning Liu4

1 BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China
2 BGI-Shenzhen, Shenzhen 518083, China
3 China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China
4 The Fourth People's Hospital of Shenyang, Shenyang 110031, Liaoning Province, China
5 James D. Watson Institute of Genome Sciences, Hangzhou 310058, China

Co-first authors: Yu Zheng and Hai-Lin Wang
Correspondence to: Dong-Ning Liu. The Fourth People's Hospital of Shenyang, Shenyang 110031, Liaoning Province, China. liudongning1976@126.com; Jian-Guo Zhang. China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China. zhangjg@genomics.cn

Received: 2017-05-31 Accepted: 2017-09-13

Abstract

• AIM: To study the genes responsible for retinitis pigmentosa.
• METHODS: A total of 15 Chinese families with retinitis pigmentosa, containing 94 sporadically afflicted cases, were recruited. The targeted sequences were captured using the Target_Eye_365_V3 chip and sequenced using the BGISEQ-500 sequencer, according to the manufacturer’s instructions. Data were aligned to UCSC Genome Browser build hg19, using the Burroughs Wheeler Aligner MEM algorithm. Local realignment was performed with the Genome Analysis Toolkit (GATK v.3.3.0) IndelRealigner, and variants were called with the Genome Analysis Toolkit Haplotypecaller, without any use of imputation. Variants were filtered against a panel derived from 1000 Genomes Project, 1000G ASN, ESP6500, ExAC and dbSNP138. In all members of Family ONE and Family TWO with available DNA samples, the genetic variant was validated using Sanger sequencing.
• RESULTS: A novel, pathogenic variant of retinitis pigmentosa, c.357_358delAA (p.Ser119SerfsX5) was identified in PRPF31 in 2 of 15 autosomal-dominant retinitis pigmentosa (ADRP) families, as well as in one, sporadic case. Sanger sequencing was performed upon probands, as well as upon other family members. This novel, pathogenic genotype co-segregated with retinitis pigmentosa phenotype in these two families.

• CONCLUSION: ADRP is a subtype of retinitis pigmentosa, defined by its genotype, which accounts for 20%-40% of the retinitis pigmentosa patients. Our study thus expands the spectrum of PRPF31 mutations known to occur in ADRP, and provides further demonstration of the applicability of the BGISEQ500 sequencer for genomics research.

• KEYWORDS: retinitis pigmentosa; PRPF31; BGISEQ-500

DOI:10.18240/ijo.2018.01.06


INTRODUCTION

Worldwide, nearly 1.5 million people suffered retinitis pigmentosa (RP). This hereditary, retinal-degenerative disease onsets with the death of photoreceptor cells, leading to stepwise visual damage[1-3], and has a frequency of 0.1% in China.

The inheritance pattern of RP includes three Mendelian subtypes: autosomal dominant RP (ADRP), autosomal recessive RP (ARRP), and X-linked RP (XLRP)[1]. ADRP accounts for nearly 40 percent of symptomatic RP, and more than 25 genes have been reported to be implicated in the ADRP subtype. Furthermore, a specific subtype of ADRP causative genes has been delineated in the literature: precursor mRNA processing genes[4]. Eight of these are either precursors of six ubiquitous core snRNP proteins (PRPF3, PRPF8, PRPF31, PRPF4, SNRNP200, and PRPF6), or splicing factors of two others (RP9 and DHX38). All eight of these have been implicated in RP. These eight genes are widely expressed, and are crucial to both symptomatic and asymptomatic individuals. Individuals with retina-specific phenotypes are evidenced pathogenic mutations in these specific genes[5].
A novel mutation in PRPF31 caused RP

Table 1  List of mutations of PRPF31 correlated with RP

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Gene</th>
<th>Disease</th>
<th>Clinical significance (last reviewed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.400delG (p.Asp134Ilefs)</td>
<td>PRPF31</td>
<td>Retinal dystrophy</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td>c.527+1G&gt;T</td>
<td>PRPF31</td>
<td>Not provided</td>
<td>Pathogenic (Nov. 7, 2016)</td>
</tr>
<tr>
<td>c.527+3A&gt;G</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa 11</td>
<td>Pathogenic (Mar. 1, 2007)</td>
</tr>
<tr>
<td>c.562G&gt;T (p.Glu188Ter)</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td>c.581C&gt;A (p.Ala194Glu)</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa 11</td>
<td>Pathogenic (Apr. 6, 2007)</td>
</tr>
<tr>
<td>c.615C&gt;G (p.Tyr205Ter)</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td>c.615C&gt;A (p.Tyr205Ter)</td>
<td>PRPF31</td>
<td>Not provided</td>
<td>Pathogenic (Jul. 30, 2013)</td>
</tr>
<tr>
<td>c.646G&gt;C (p.Ala216Pro)</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa 11</td>
<td>Pathogenic (Jun. 1, 2009)</td>
</tr>
<tr>
<td>c.764A&gt;T (p.Gln255Leu)</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td>c.950delG (p.Gly317Alafs)</td>
<td>PRPF31</td>
<td>Not provided</td>
<td>Pathogenic (May 11, 2016)</td>
</tr>
<tr>
<td>c.967G&gt;T (p.Glu323Ter)</td>
<td>PRPF31</td>
<td>Retinal dystrophy</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td>c.994C&gt;T (p.Gln332Ter)</td>
<td>PRPF31</td>
<td>Retinal dystrophy</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td>c.1060C&gt;T (p.Arg354Ter)</td>
<td>PRPF31</td>
<td>Retinal dystrophy</td>
<td>Likely pathogenic</td>
</tr>
<tr>
<td>c.1073+1G&gt;A</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa 11</td>
<td>Pathogenic (Apr. 1, 2015)</td>
</tr>
<tr>
<td>c.1120C&gt;T (p.Gln374Ter)</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa 11</td>
<td>Pathogenic (Jan. 30, 2015)</td>
</tr>
<tr>
<td>c.1273C&gt;T (p.Gln425Ter)</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa 11</td>
<td>Pathogenic (Oct. 8, 2015)</td>
</tr>
<tr>
<td>c.1374+654C&gt;G</td>
<td>PRPF31</td>
<td>Retinitis pigmentosa 11</td>
<td>Pathogenic (Sep. 1, 2009)</td>
</tr>
</tbody>
</table>

Incomplete dominance and penetrance are ubiquitous in Mendelian disease, and the genetic architecture underlying many Mendelian diseases remains unclear. It is important to elucidate how polymorphisms are associated with the development of disease, and the relationship between genetic variants and clinical phenotypes. Imperfect understanding of disease severity and incomplete penetrance retards implementation of genetic counseling in the clinic. PRPF31-associated RP is a canonical example of incomplete penetrance, because it is almost universal to observe the presence of symptomatic and asymptomatic mutation carriers in affected families, while the causes of incomplete penetrance are yet to be discovered. Mutations in PRPF31 tend to occur in the major ADRP locus (Table 1), causing the disease subtype termed RP11 (defined by genetic variants of chromosome 19q13.4, OMIM#60138). Meanwhile, genetic variants of PRPF31 compose a large proportion of ADRP, and account for almost 10 percent of cases. There are nonsense, missense, frameshift and large deletion variants. Previous studies have reported that the wildtype PRPF31 allele, which was inherited from the unaffected parent, promoted the generation of asymptomatic offspring. Symptomatic and asymptomatic offspring inherited distinct variants of the cis-typical wild type PRPF31 alleles from the parents, despite a genetic predisposition. This suggested that disease phenotype may be prevented by inheritance of particular parental factors. Compared with symptomatic relatives, the expression of the wild type PRPF31 allele has been shown to twofold higher in asymptomatic ones. This suggests that disease-preventative allele may be a common one. The disease-preventative allele acted in cis to improve the expression of the wild type PRPF31. A treatment for the symptoms resulting from inheriting the pathogenic PRPF31 variant causative of RP may be contrived, if the genetic factor in 19q13.4/PRPF31 which prevents disease phenotype can be identified and understood. The MSR1 element has two forms, 3 repeats and 4 repeats, in human genome. The 4 repeats allele was along with higher expression of PRPF31. This is also inferred elsewhere to affect phenotypes.

SUBJECTS AND METHODS

Subjects and Clinical Testing  Shenyang the Fourth People's Hospital (SFPH) agreed with this case research. This research abided by the tenets of the Helsinki Declaration. Subjects who took part in this research had signed the informed consent. Totally 15 Chinese families with 94 sporadic RP cases participated in the study. All of the patients accepted carefully inspection which included regular eye examinations, such as the conditions of the fundus. Totally 200 normal subjects were recruited as controls. They have no relationship with the patients.

BGISEQ-500 Sequencer and DNA Sequence Analysis  In line with the manufacturer’s instructions, the extraction of genomic DNA depended on the DNA extraction kits (TianGen, Beijing). A total of 5 μg genomic DNA was sequenced by a BGISEQ-500 sequencer. The targeted sequences were captured using the Target_Eye_365_V3 chip, which contained 365 genes related to eye diseases according to OMIM. First, BGISEQ-500 libraries for each sample were prepared. Second, the target areas were concentrated. Third, hybrid-capture libraries were prepared for paired-end sequencing with read lengths of 50 bp. Finally, each sample had an average coverage depth of nearly 300× on the BGISEQ-500 sequencer. The
Burroughs Wheeler Aligner MEM algorithm was applied for aligning reads to hg19 from UCSC Genome Browser. The Genome Analysis Toolkit (GATK v.3.3.0) IndelRealigner was used for realignment. Sequence variants were called with the Genome Analysis Toolkit Haplotypecaller, without any use of imputation. Variants were filtered against a panel derived from One Thousand Genomes Project, ESP6500, ExAC and dbSNP138.

**Sanger Sequencing and Pathogenic Analysis** The genetic variant was validated using Sanger sequencing in the two families with available DNA samples. Three steps were used to estimate whether the variant was the cause of the two families. Initially, the variant was positive and co-segregation in the two families. Thereafter, two websites (https://scholar.google.com.hk/; https://www.omim.org) were used to search previous reports about whether the variant existed in unaffected controls. Afterwards, the consistency of two families’ phenotypes and the gene with genetic variant should be taken into consideration.

**RESULTS**

**Clinical Evaluations** The pedigree and disease inheritance of Family ONE and Family TWO over three generations indicated a dominant inheritance pattern in both families (Figure 1A, 1B). The proband of Family ONE characterized by bilateral corrected visual acuity 0.6 with childhood onset of night blindness at 6 years old, the vision field concentric contraction less than 15 degrees. The proband of Family TWO characterized by bilateral corrected visual acuity 1.0 with childhood onset of night blindness at 10 years old, the vision field concentric contraction less than 20 degrees. The fundus findings of the patients in both families were similar, and both showed typical RP. The optic disc was waxy, the retinal vessel attenuating, and the osteocyte-like pigment deposited in peripheral retina. Scotopic rod responses of electroretinogram (ERG) in majority eyes was non-detectable, similarly photopic cone responses of ERG were non-detectable from all eyes. Moreover, the ellipsoid bands around the macular area, which could easily be detected with the optical coherence tomography (OCT), disappeared in both families with RP. Other families and sporadic cases were diagnosed with RP by clinicians, according to clinical descriptions of patients and relatives.

**Mutation Analysis** We performed Target_Eye_365_V3 chip sequencing on two families and 200 unaffected controls. The average of raw data for each sample was 3.2 Gb. The 99.94% of raw data could align to the targeted sequences. The mean sequencing depth was 300×. Sequencing coverage of the object region, sequencing depth more than 10×, was 98.94%. Three steps were used as the method of screening variants. Firstly, variants of which frequency is less than 1% in ExAC, ESP6500, One Thousand Genomes Project and dbSNP138, were preserved. Secondly, variants which located in coding region or around splicing site were kept. Thirdly, variants which were loss of function or predicted to be damage were preferential for analysis. Finally, a heterozygous variant c.357_358delAA in *PRPF31* gene was discovered in Family ONE, Family TWO, and in one sporadic case. Sanger sequencing confirmed the presence of this genetic variant (Figure 1C, 1D).
The heterozygous variant c.357_358delAA in PRPF31 gene was co-segregated in Family ONE and Family TWO. It signified that this heterozygous variant was found in all RP patients and was absent in all unaffected relatives in the two families (Figure 1A, 1B). This variant was a candidate causative variant for the two families. After analysis of other patients and their family members, we also identified another, previously reported[18], RP associated mutation, in a single RP patient: c.421-1G>A in PRPF31.

This primary variant, c.357_358delAA in PRPF31, co-segregated to the phenotype in Family ONE and Family TWO. The sequencing result of 200 normal subjects didn’t find this variant. Furthermore, this variant didn’t exist in the databases of implicitly benign variations. The genetic variant was absent in both the existing databases of genetic polymorphisms, and in the reported literatures. This primary variant appears to be discovered for the first time by our research, and can be considered to be novel.

**DISCUSSION**

ARRP accounts for nearly 55% of all RP patients. The ratio of XLRP is just about 10%. The proportion of ADRP is almost 35% of all RP patients[19]. Particular, there have been more than 25 genes reported to correlate with ADRP. The 10% of ADRP cases were related to mutations in PRPF31 gene[8-11].

Incomplete penetrance is an obvious and intriguing characteristic of RP11, which is related to PRPF31 gene and a subtype of RP. Incomplete penetrance has been discovered in numerous RP11 pedigrees which carry polymorphisms in PRPF31 gene. When we find a novel, likely pathogenic variant in PRPF31 in a single RP family with an autosomal, dominant mode of inheritance, the fact that the mutation co-segregates with the disease in the family can interpreted as evidence of pathogenicity of that particular variant. However, when it proves to not be perfectly pathogenic, we should find clues as to why this is the case by paying attention to the character of partial penetrance in other PRPF31 polymorphism carrying families. Although there already have been more than two theories concerning the cause of incomplete penetrance of PRPF31 variants. One is that the core promoter activity of PRPF31 gene is related to the forms of MSR1 repeat[17]. Another is that CNOT3 could bind to the core promoter resulting in the decrease expression of PRPF31 gene[19]. More studies are needed to clarify the exact mechanism of incomplete penetrance.

It is known that mutations in six, widely expressed splicing factors (PRPF3, PRPF6, PRPF8, PRPF31, snRNP200 and RP9) are responsible for a significant fraction of ADRP cases.

**Figure 2** The SWISS MODEL of the wild type protein (A), and the variant protein translated by the mutation c.357_358delAA (B) Locations of p.Ser119 polymorphism-derived protein variation are pointed out by an arrow.

The causative mechanism underlying why polymorphisms in such “housekeeping” genes bring about retina-specific phenotype remains unclear. It may be that mutations in PRPF31 disturb pre-mRNA splicing of RP9 gene and FSCN2 gene transcripts which are more directly causative of ADRP[20,21], due to interactions between proteins and variants in the genes RP9 and FSCN2 (http://string-db.org/cgi/input.pl?input_page_show_search=on&UserId=fiYXDJqzw89n&sessionId=bf0VWE8QypBI).

Deletions in two residues such as c.357_358delAA may impair the structure and function of the protein resulting from PRPF31 transcription. PRPF31 gene encodes a 499-amino acid protein. The mutation c.357_358delAA in PRPF31 may cause a 123-amino truncated protein product. According to Table 1, there are already 18 pathogenic mutations found in PRPF31 which are all behind of the location of c.357_358delAA (p.Ser119SerfsX5). The models which emerge from the SWISS MODEL (https://swissmodel.expasy.org/) of the wild type protein and the protein transcribed by the mutation c.357_358delAA are shown in Figure 2A and 2B. In summary, one novel mutation, NM_015629 c.357_358delAA (p.Ser119SerfsX5), was identified in our study, expanding the spectrum of PRPF31 mutations known to be implicated in RP.

In addition, the BGISEQ500 sequencer proves adequate for all genome sequencing by the mutation c.357_358delAA are shown in Figure 2A and 2B. In summary, one novel mutation, NM_015629 c.357_358delAA (p.Ser119SerfsX5), was identified in our study, expanding the spectrum of PRPF31 mutations known to be implicated in RP.
For clean data where the leading sequences have been trimmed off each end of the individual reads, a Q score of 20 is obtained for almost 95% of sequence, and a Q score of 30 for almost 84% of sequence. Of the variants identified, 99.62% are represented in dbSNP, and 97.71% are annotated in the One Thousand Genomes Project database. In this study, more than 50% of sequenced genotypes proved to contain pathogenic polymorphisms in accord with diseased phenotype. Finally, the presence of those polymorphisms was verified by Sanger sequencing in all cases. This result provides further affirmation of the already high standard demonstrated elsewhere in the BGISEQ-500 sequencer for this type of genomic research.

ACKNOWLEDGEMENTS

Conflicts of Interest: Zheng Y, None; Wang HL, None; Li JK, None; Xu L, None; Tellier L, None; Li XL, None; Huang XY, None; Li W, None; Niu TT, None; Yang HM, None; Zhang JG, None; Liu DN, None.

REFERENCES


3 Glöckle N, Kohl S, Mohr J, Scheurenbrand T, Sprecher A, Weisschuh N, Bernd A, Rudolph G, Schubach M, Poloschek C, Zrenner E, Biskup S, Berger W, Wissinger B, Neidhardt J. Panel-based next generation sequencing in all cases. This result provides further affirmation of the presence of those polymorphisms was verified by Sanger sequencing in all cases. This result provides further affirmation of the already high standard demonstrated elsewhere in the BGISEQ-500 sequencer for this type of genomic research.


