Author Manuscript
Faculty of Biology and Medicine Publication

This paper has been peer-reviewed but dos not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

serval
serveur académique lausannois

UNIL | Université de Lausanne
Faculté de biologie et de médecine

A single-base substitution within an intronic repetitive element causes dominant retinitis pigmentosa with reduced penetrance

Thomas Rio Frio[1], **Terri L. McGee**[2], **Nicholas M. Wade**[1], **Christian Iseli**[3], **Jacques S. Beckmann**[1,4], **Eliot L. Berson**[2], and **Carlo Rivolta**[1]

[1]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland [2]The Berman-Gund Laboratory for the Study of Retinal Degenerations, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts, USA [3]Ludwig Institute for Cancer Research and Swiss Institute of Bioinformatics, Lausanne, Switzerland [4]Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

Abstract

We report the study of a large American family displaying autosomal dominant retinitis pigmentosa with reduced penetrance, a form of progressive and hereditary retinal degeneration. Although the particular inheritance pattern and previous linkage mapping incriminated the *PRPF31* gene, extensive screening of all its exons and their boundaries failed in the past to reveal any mutation. In this work, we sequenced the entire *PRPF31* genomic region by both the classical Sanger method and ultra-high throughput (UHT) sequencing. Among the many variants identified, a single-base substitution (c.1374+654C>G) located deep within intron 13 and inside a repetitive DNA element was common to all patients and obligate asymptomatic carriers. This change created a new splice donor site leading to the synthesis of two mutant *PRPF31* isoforms, degraded by nonsense-mediated mRNA decay. As a consequence, we observed an overall reduction in full-length *PRPF31* mRNA and protein levels, with no evidence of mutant proteins being synthesized. Our results indicate that c.1374+654C>G causes retinitis pigmentosa via haploinsufficiency, similar to the vast majority of *PRPF31* mutations described so far. We discuss the potential of UHT sequencing technologies in mutation screening and the continued identification of pathogenic splicing mutations buried deep within intronic regions.

Keywords

PRPF31; RP11; splicing; UHT sequencing

INTRODUCTION

Retinitis pigmentosa (RP) is the name given to a group of inherited diseases that causes progressive degeneration of the retina. It has a cumulative prevalence of 1/4000, affecting over 1 million individuals worldwide [Hartong et al., 2006; Berson, 2007]. In early stages, it is characterized by night blindness, followed by a decrease in mid-peripheral vision and ultimately loss of central vision as well, due to the death of both rod and cone photoreceptor cells [Berson, 1993].

RP is genetically highly heterogeneous; it can be transmitted as a Mendelian autosomal dominant (adRP), autosomal recessive, or X-linked trait and also in small proportions as a digenic and non-Mendelian trait [Rivolta et al., 2002]. It can occur alone or as part of more complex syndromes. So far, 50 genes or loci have been shown to be involved in non-syndromic RP (http://www.sph.uth.tmc.edu/retnet/), but together they account for only about half of all reported cases and thus many genes associated with RP remain unknown [Hartong et al., 2006]. The *RP11* locus, located on 19q13.4 (MIM# 600138), was first linked to adRP from the analysis of a large British family [Al Maghtheh et al., 1994] and later confirmed to be responsible for the disease in several other pedigrees [Xu et al., 1995; Al Maghtheh et al., 1996; McGee et al., 1997; Vithana et al., 1998]. Mutations in the pre-mRNA splicing factor gene *PRPF31* (MIM# 606419) were subsequently identified as the cause of adRP in *RP11* families [Vithana et al., 2001]. So far, more than 40 *PRPF31* mutations have been detected, some of which are characterized by large deletions [Abu-Safieh et al., 2006; Sullivan et al., 2006; Kohn et al., 2008]. However, the vast majority of *PRPF31* mutations create premature termination codon (PTC) in the mRNA transcribed from the mutant allele, which in turn is degraded by nonsense-mediated mRNA decay (NMD). In such cases, the absence of *PRPF31* mRNA and protein derived from the mutant allele has been shown to be the likely cause of the disease [Rio Frio et al., 2008b].

Analysis of several families with RP indicates that *PRPF31* mutations may display incomplete penetrance, i.e. there are obligate carriers of mutations who do not manifest the disease (e.g. [Berson et al., 1969; Al Maghtheh et al., 1996]). Penetrance of the disorder in heterozygous carriers of *PRPF31* mutations has been previously associated with the level of expression of the wild-type *PRPF31* allele [Vithana et al., 2003; Rivolta et al., 2006; Liu et al., 2008]. Specifically, high expression levels of wild-type *PRPF31* mRNA appear to offer protection from RP, possibly because they compensate for haploinsufficiency resulting from the degradation of mutant transcripts. By linkage studies, *PRPF31* mRNA expression and therefore the penetrance of *PRPF31* mutations has been linked to modifiers present within the *PRPF31* locus itself and on chromosome 14q21-23 [McGee et al., 1997; Vithana et al., 2001; Rio Frio et al., 2008a].

Here we report the study of a large family (ID #1562, Fig. 1) which was historically described as the first example of retinitis pigmentosa segregating as an autosomal dominant form with reduced penetrance [Berson et al., 1969;Berson and Simonoff, 1979]. Subsequent genotyping of selected members, using markers spanning the *RP11* locus, strongly suggested that the *PRPF31* gene was responsible for the disease [McGee et al., 1997]. Furthermore, *PRPF31* expression analyses in lymphoblastoid cell lines (LCLs) from relevant members of this pedigree showed amounts of transcripts that were reduced with respect to controls and comparable to those present in individuals with known *PRPF31* mutations [Rivolta et al., 2006]. Although all these data indicated that a dominant mutation in *PRPF31* leading to haploinsufficiency was the likely cause of RP in family #1562, prior screens revealed no pathogenic changes [McGee et al., 2002] and no deletions [Rivolta et al., 2006] in any of the 14 *PRPF31* exons.

MATERIALS AND METHODS

Ascertainment of patients

Clinical evaluation of members of this family was originally described in 1969 and again in 1979 [Berson et al., 1969; Berson and Simonoff, 1979]. Following informed consent, DNA was extracted from lymphocytes collected from blood samples from family members as described previously [McGee et al., 1997].

Cell culture and drug treatments

Lymphoblastoid cell lines (LCLs) from individuals belonging to family #1562, from 10 members of the CEPH collection (normal controls), from 6 patients with known mutations in *PRPF31* and one unaffected relative (cell line #14523, 14686, 14266, 12688, 14284, AG0293, and 14251) were obtained and cultured as previously described [Rivolta et al., 2006; Rio Frio et al., 2008b]. Inhibition of NMD was achieved by treating 5 million cells with 100 μM wortmannin or 300 μg/ml emetine (Sigma-Aldrich, St Louis, MO), dissolved in water, for a 6-hour period. Negative controls were performed by supplementing the cells with equivalent amounts of water and by incubating them in the same conditions.

Long-range PCR and DNA sequencing

Using previously-defined methods for long-range PCR of large genomic regions [Hinds et al., 2005], we selected 4 fragments containing the *PRPF31* gene to amplify the genomic DNA from selected individuals (III-9, III-11, and IV-16, Fig.1). These fragments were 6893 bp, 10542 bp, 8124 bp, and 5690 bp in length, respectively, and encompassed a ~31 kb genomic region, from position 59′297′569 to 59′328′826 on chromosome 19 (NC_000019.8). The primers used in these 4 PCRs were in part previously published [Hinds et al., 2005] and in part newly-designed oligos (Supp. Table S1). PCR reactions were each performed in a final reaction volume of 50 μl, containing 1x GC Buffer I (TaKaRa, Otsu, Shiga, Japan), 0.4 mM of each dNTP, 0.2 μM of each primer, 0.5 U of TaKaRa LA Taq (TaKaRa) and 500 ng of DNA. Reactions were incubated at 94°C for 1 minute, followed by 30 cycles of 98°C for 5 seconds and 68°C for 15 minutes and a final elongation step of 70°C for 10 minutes. PCR products derived from each patient's DNA were pooled in an equimolar ratio.

Ultra-high throughput (UHT) sequencing reactions were performed on these long-range PCRs using an Illumina Genome Analyzer (Illumina, San Diego, CA). Small gaps between the first and second, and between the second and third regions (100 bp and 72 bp, respectively) were amplified and sequenced using the Sanger (dideoxy) method. The reads intensity files produced by the Illumina GA were processed through the Rolexa software package [Rougemont et al., 2008], which performs base calling to produce a maximal number of sequence reads with well defined quality information for each nucleotide. A total of 2′089′433 reads with sufficient length and quality were obtained. These reads were aligned to the human genome reference assembly 36 to find exact matches using the fetchGWI software [Iseli et al., 2007]: 718′883 reads found a single exact match (group U of unique reads), 235′273 reads found multiple exact matches (group R of repeated reads), and 1′135′277 reads remained with no exact match (group M of missed reads). A total of 689′957 among all U reads (96%) were mapped to the region of interest on chromosome 19. In order to discover polymorphisms in the sequenced region, the reads from group M were aligned using the global alignment software align0 [Myers and Miller, 1988] to the reference genomic sequence using the following method. First, we selected from all groups all the reads that shared a common 12-mer with any other 12-mer from the chromosomal region of interest, generating subsets of sequences termed U_{sel}, M_{sel}, and R_{sel}. Then, we performed a global alignment of all M_{sel} sequences against the reference genome sequence, and discarded all alignments with more than 3 mismatches. Subsequently, we identified from U_{sel} the reads mapping outside of the region of interest (U_{out}), performed a global alignment of M_{sel} reads against the reads from R_{sel} and U_{out}, and discarded the reads from M_{sel} which received a better alignment score in this latter global alignment. We predicted a SNP (or a variant) when an observed difference between the reads and the reference sequence occurred at a sufficient rate among all the reads that cover the position. In this case, we required at least 4 reads exhibiting the change and 20% of the total coverage.

Sequencing of the *PRPF31* gene was repeated on the same samples as well as long-range PCRs from 2 normal CEPH controls [Dausset et al., 1990] using the Sanger method, by amplifying its full genomic region (exon 1 to exon 14, corresponding to bases 5001 to 21361 from *PRPF31* GenBank reference genomic sequence NG_009759.1) with specific primers (Supp. Table S1) and TaKaRa LA Taq (Takara), as described above. PCR products were purified using the Qiaquick PCR Purification kit (Qiagen, Venlo, The Netherlands) and sequenced using internal primers (Supp. Table S2). Sequencing reactions were performed by mixing 4.5 μl of a 1/10 dilution of purified PCR product, 0.75 μM of internal primer, 0.75 μl of BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) and run on an ABI-3130XL (Applied Biosystems). Variants detected were named using cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the *PRPF31* GenBank reference sequence NM_015629.3, according to the Human Genome Variation Society guidelines (www.hgvs.org/mutnomen).

RNA extraction and cDNA synthesis

Approximately 5 million cells from each cell line were homogenized using QIAshredder columns and total RNA was isolated using the RNeasy kit (Qiagen) according to manufacturer's recommendations, with an on-column DNase treatment. To completely remove DNA, isolated RNA was incubated again with RNase-free DNase I (Roche, Indianapolis, IN) at 37°C for 30 minutes and 75°C for 10 minutes. RNA was retro-transcribed using anchored poly-dT oligos ($dT_{20}VN$) and Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA). Negative control samples for PCR reactions were prepared by this same procedure, but reverse transcriptase was not added.

Mutation analysis by PCR and semi-quantitative PCR

Specific primers (Supp. Table S1) were designed to amplify the exon 13-exon 14 region of the cDNA derived from *PRPF31* mRNA (NM_015629.3), except for LCLs from the patient with the *PRPF31* c.323-2A>G change, for which previously developed mutation-specific primers were used [Rio Frio et al., 2008b]. PCR reactions were performed in a final volume of 10 μl, containing 1x GC Buffer I, 0.4 mM of each dNTP, 0.2 μM of each primer, 0.5 U of TaKaRa LA Taq (Takara) and 2 μl of a 1/10 dilution of retrotranscribed mRNA. Reactions were incubated at 94°C for 1 minute followed by 30 cycles of 94°C for 30 seconds, 60°C for 30 seconds, 68°C for 2 minutes, and a final elongation step of 70°C for 10 minutes. A final stage of 95°C for 5 minutes followed by a slow cooling down to 4°C (0.01°C per second) was added to prevent the formation of heteroduplexes between PCR products containing similar sequences. PCR products were resolved on 1.5% agarose gels and, for semi-quantitative PCR, by capillary electrophoresis with the eGene HDA-GT12 Multi-Channel Genetic Analyzer (eGene Inc., Irvine, CA). Quantification was performed by 2 independent methods: the ImageJ software [Abramoff et al., 2004] to analyze the agarose gel image, and the Biocalculator software (eGene) for runs performed with the HDA-GT12 instrument. Data obtained from these 2 methods were combined to calculate the mean values of the ratio between mutant and wild-type band intensities, and the error propagation relative to this operation calculated accordingly.

To identify cDNA isoforms, PCR products were sub-cloned into TOPO TA vectors (Invitrogen) and propagated in the *E. coli* cells. PCR reactions to identify positive clones were performed directly on colonies and products were sequenced by the Sanger method.

Quantification of wild-type mRNA derived from the mutant allele

cDNA from individuals carrying the c.1374+654C>G intronic mutation, who were also heterozygotes for SNP rs1058572:G>A alleles in exon 7, was amplified from exon 7 to exon 14 using specific oligos (Supp. Table S1). The A allele of this SNP was demonstrated to be

in linkage disequilibrium with the mutation and was thus used as a marker to quantify allele-specific expression. The oligo annealing to exon 7 was designed with one mismatch (base 14, T instead of G) to prevent digestion of cDNA at a site located 18 bp upstream of the site of interest by *Bse*RI, used for quantification of the wild-type mRNA derived from the mutant allele. PCR reactions were performed in a total volume of 25 μl, containing 1x Expand High Fidelity PCR System (Roche) buffer with $MgCl_2$, 100 μM dNTP mix, 200 nM of each primer, 1 U of Expand High Fidelity PCR System (Roche) and 500 ng of retrotranscribed mRNA. The thermal profile used was optimized to prevent the amplification of other cDNAs than those derived from wild-type mRNA. Specifically, PCR reactions were incubated at 94°C for 15 minutes, followed by 40 cycles of 94°C for 15 seconds, 63°C for 30 seconds and 68°C for 1 minute.

Quantification of wild-type mRNA allelic expression using the SNP rs1058572:G>A was performed by interpolation, via the use of calibration standards. Two different cDNA templates from CEPH LCLs that were homozygous for either the major or minor SNP alleles were mixed to obtain quantification standards in the following ratios (G/A): 90/10, 80/20, 70/30, 60/40 and 50/50 at concentrations comparable with cDNA templates (to avoid differences in signal strength). cDNA from samples, quantification standards and controls (cDNA derived from homozygotes for either SNP alleles) were amplified on the same plate.

PCR products were first assessed to be unique products on a 2% agarose gel, and then digested using two restriction enzymes. *Bss*SI (New England Biolabs, Ipswich, MA), which cuts all PCR products regardless of their allelic origin, was used to reduce the size in order to increase resolution for quantification; *Bse*RI (New England Biolabs), which specifically recognizes the G allele of SNP rs1058572:G>A and cuts 14 bp away from it (towards the internal part of the PCR product), was used to distinguish PCR products derived from each allele. Digested products were run again on a 2% agarose gel and quantified by the ImageJ software. PCR products from transcripts derived from homozygous genotypes were used as digestion controls.

Antibodies and western blotting

Total protein was extracted from LCLs in RIPA buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.4, 1 mM EDTA, 1% Tx-100, 0.1% SDS, 0.5% sodium deoxycholate) with Complete, Mini EDTA-free Protease Inhibitor Tablets (Roche), and protein content was estimated using the BCA Protein Assay Kit (Thermo Fisher Scientific, Rockford, IL). Detection of PRPF31 protein was performed via western blotting using a specific N-terminal PRPF31 antibody and analyzed with the Odyssey Infra-Red Detector (LI-COR Biosciences, Lincoln, NE) as previously described [Rio Frio et al., 2008b]. Individual extractions were performed on 3 independent cell cultures of LCLs derived from 1 affected (IV-16) and 2 asymptomatic (III-9, III-11) carriers of the mutation from family #1562 and from the normal control cell line #14251. Equal amounts were run on denaturing SDS-PAGE gels with the Precision Prestained Protein Standards (Bio-Rad, Hercules, CA). For each extraction, 4 gels were run independently.

RESULTS

Identification of a novel intronic mutation in the *PRPF31* gene

We amplified by long-range PCR a region of approximately 31 kb that encompasses the *PRPF31* gene (including flanking regions 13 kb upstream and 2 kb downstream) by using as template material the genomic DNA of 1 affected patient (IV-16) and 2 asymptomatic carriers of the disease (III-9, III-11) from family #1562 (Fig. 1). Ultra-high throughput (UHT) sequencing of these amplicons, performed independently for each individual,

revealed the presence of 102 SNPs, 56 of which already recorded in dbSNP. The median coverage of the chromosomal region of interest was 620 X, with an interquartile range of 324 to 1185 X. None of the variants specifically present within the 16 kb genomic region spanning *PRPF31* exons 1 to 14 were deemed pathogenic as they were referenced in SNP databases, were found in DNA from control CEPH LCLs, or did not cosegregate with the disease haplotype.

We therefore resequenced this 16 kb DNA fragment in the same individuals by the Sanger method. In IV-16, chosen as reference, we identified 17 heterozygous variants that were also detected by UHT processing, and an additional change in intron 13 (Table 1). This novel change was a C>G transition at position 654 of the intron, c.1374+654C>G, located within the first of 7 consecutive and almost identical sequences of 56 bp composing a VNTR (variable number of tandem repeats) element (Fig. 2). The new variant was in fact present in the list of potential changes detected by UHT sequencing, but the coverage was much lower than the median in this region, and hence it was discarded as noise. Specifically, UHT reads could not be reliably aligned on this VNTR since they were in average 30-35 bp-long, i.e. shorter than the size of the repeated module itself.

The c.1374+654C>G change represented a good pathogenic candidate since it was the only one detected in all 3 individuals who had the disease haplotype, was not detected in DNA from CEPH cell lines, and was not referenced in dbSNP (www.ncbi.nlm.nih.gov/projects/SNP) or the Ensembl SNP (www.ensembl.org) database. Sequencing of DNA extracted from blood samples from remaining members of family #1562 showed that c.1374+654C>G was present in all affected members of the family, as well as all obligate asymptomatic carriers from previous haplotype analyses ([McGee et al., 1997], Fig. 1). Furthermore, this variant was not detected in 300 control chromosomes, suggesting that it was likely to be the causative mutation in this family. Previously clinically uncharacterized family members (IV-2, 8, 10, 11, 23 and 33) were found to carry the mutation and likely represent asymptomatic individuals.

Effect of the c.1374+654C>G mutation on *PRPF31* pre-mRNA splicing

In silico analyses of the *PRPF31* allele carrying the c.1374+654C>G mutation, using the NNSPLICE software [Pedersen and Nielsen, 1997], predicted that this variant would very likely create a new splice site. Specifically, the likelihood score for splicing donor sites at this position increased from <0.1 in the wild-type sequence to 0.96 in the mutant sequence, out of a maximum of 1.00. PCR amplification of cDNA from LCLs from individuals carrying the c.1374+654C>G variant and a normal control using primers in exons 13 and 14 yielded a 158 bp product in all of the samples, as well as 2 supplementary PCR products of 333 bp (MUT1) and 811 bp (MUT2), present in lower amounts, only in LCLs from the 3 carriers of c.1374+654C>G (Fig. 3A). Cloning and sequencing of these PCR products revealed that the common 158 bp PCR product contained the wild-type exon 13-exon 14 junction. The MUT1 mRNA isoform (r.1770_1771ins1770+479_1770+653) was comprised of exon 13 spliced with bases 479 to 653 of intron 13, followed by exon 14, indicating that a cryptic acceptor site at position 478 was activated and an additional exon of 175 bp was inserted (Fig. 3B). The MUT2 mRNA (r.1770_1771ins1770+1_1770+653) did not use the native intron 13 splice donor site and instead retained bases 1 to 653 of intron 13 as part of exon 13, which was then joined with exon 14 (Fig. 3B). These aberrant isoforms were not observed in transcripts present in LCLs from 26 control individuals and 6 RP patients with different mutations in *PRPF31* (data not shown). Altogether, these data indicate that splicing of pre-mRNAs derived from the *PRPF31* allele containing the c.1374+654C>G mutation is specifically affected by the presence of a newly-created strong donor splice site and results in partial retention of intron 13.

Analysis of mRNA derived from the mutant allele

By *in silico* analysis, MUT1 and MUT2 mRNAs were predicted to produce transcripts with 37 and 11 mutant codons after codon 458, respectively, followed by a premature termination codon (PTC) (Fig. 3B). Since these PTCs are located before the last exon, both mutant *PRPF31* transcripts are potential target for NMD-driven degradation, a phenomenon that could explain their presence in reduced amounts with respect to the wild-type isoform (Fig. 3A).

To test this hypothesis, we analyzed LCLs derived from the 3 carriers of the c. 1374+654C>G mutation, one control CEPH individual, and one patient who is a carrier of a different *PRPF31* mutation (c.323-2A>G) known to result in an NMD-sensitive transcript [Rio Frio et al., 2008b]. Treatment of these LCLs with 2 potent NMD inhibitors, emetine and wortmannin, rescued mutant mRNA from LCLs with the c.323-2A>G mutation, indicating that inhibition of NMD was effective (data not shown). Similarly, amounts of both mutant mRNA isoforms MUT1 and MUT2 significantly increased relative to wild-type transcripts in LCLs carrying the c.1374+654C>G mutation. Meanwhile, only wild-type mRNA was amplified from control LCLs (CEPH), even when NMD was blocked. This indicates that the MUT1 or MUT2 isoforms are not naturally present or, if present, would be produced in extremely low amounts in absence of the mutation (Fig. 4A). Although precise quantification was not possible, since PCR efficiencies were product-specific, MUT1 and MUT2 mRNA isoforms represented on average 5% and 10%, respectively, of the wild-type mRNA in untreated LCLs, which increased to 38% and 83%, respectively, in LCLs grown in the presence of wortmannin and to 15% and 51%, respectively, in emetine-treated LCLs (Fig. 4B).

Unlike other *PRPF31* mutations that affect canonical splicing sites, the c.1374+654C>G mutation creates a new donor splice site but does not abolish the natural donor site of exon 13, which in fact is used to create the MUT1 isoform (Fig. 3). Consequently, it is conceivable that some wild-type mRNA could be derived from mutant alleles. To test this hypothesis, we searched for a heterozygous SNP in the *PRPF31* coding sequence that could be used to tag the mutant allele. Individuals III-9 and IV-16 were found to be heterozygotes (G/A) for rs1058572:G>A, located in exon 7, whereas III-11 was a homozygote for the A allele, indicating that the A allele was in linkage disequilibrium with the mutation. We amplified *PRPF31* cDNA derived from the 2 heterozygotes from exon 7 to exon 14, using PCR conditions optimized to prevent the amplification of the MUT1 and MUT2 isoforms. Analysis of *Bse*RI-digested PCR products, cleaved when the rs1058572:G>A G allele is present, revealed that approximately 10% of the wild-type mRNA was derived from the mutant *PRPF31* allele (Supp. Figure S1). Overall, the c.1374+654C>G mutation leads then to the formation of 3 different mRNA forms r.[=, 1770_1771ins1770+479_1770+653, 1770_1771ins1770+1_1770+653].

Protein analysis in LCLs from carriers of the c.1374+654C>G mutation

Western blots in LCLs from individuals with the c.1374+654C>G mutation, by using an antibody recognizing the N-terminal PRPF31 moiety, revealed a single 61-kDa band corresponding to the expected size of wild-type PRPF31. Importantly, no truncated proteins resulting from the possible translation of the MUT1 or MUT2 isoforms were detected (Fig. 5).

DISCUSSION

In this study we report the identification of a mutation in the *PRPF31* gene, which has been elusive for almost 40 years, in a large family with autosomal dominant RP with reduced

penetrance. The c.1374+654C>G change is the first reported *PRPF31* mutation located deep within an intron, confirming the suspicion that the genetic defect in this family was nonconventional. This sequence variant creates a very strong new donor splice site at position 654 of intron 13, resulting in the production of 2 novel mRNA isoforms, MUT1 and MUT2, that retain different parts of intron 13 during mRNA splicing and maturation. This new donor splice site was never used in control cell lines and, consequently, we conclude that such isoforms are the result of aberrant splicing, rather than natural alternative splicing.

Both the MUT1 and MUT2 mRNA isoforms harbour premature termination codons before the last exon and, similar to the majority of *PRPF31* mutations described so far, have been shown to be present in reduced amounts compared to wild-type *PRPF31* mRNA. This can be attributed to the action of NMD, since NMD inhibitors significantly rescued these transcripts. Analyses of PRPF31 proteins in carriers of the c.1374+654C>G mutation revealed no detectable amounts of mutant proteins derived from MUT1 and MUT2 mRNA. Considering the demonstrated high sensitivity of the methods used (3-4% of the full-length PRPF31 were detectable [Rio Frio et al., 2008b]), it is highly unlikely that significant quantities of mutant PRPF31 protein were indeed present in these patient cell lines. These results are consistent with previous findings on 6 other *PRPF31* mutations that lead to PTC-containing and NMD-sensitive transcripts [Rio Frio et al., 2008b], reinforcing the notion that haploinsufficiency is likely the primary cause of *PRPF31*-linked RP.

The canonical donor splice site in exon 13 was not abolished by c.1374+654C>G and retained some functional activity, in conjunction with the natural acceptor site for exon 14 or with a cryptic site at position 478 in intron 13. Specifically, approximately 1/10 of the pre-mRNA derived from the mutant allele was correctly spliced, generating wild-type transcripts. This represents the first demonstrated example of some residual wild-type mRNA being derived from an allele carrying a *PRPF31* mutation and may explain the slightly higher mRNA levels found in patients from family #1562, compared with all other affected and asymptomatic carriers from other families with different *PRPF31* mutations [Rivolta et al., 2006]. Interestingly, many asymptomatic individuals in this family (IV-2, IV-8, IV-10, IV-11, IV-23 and IV-33) were only identified by molecular genetic evidence, reinforcing the notion that asymptomatic carriers of *PRPF31* mutations are truly clinically unaffected. Furthermore, the large number of asymptomatic and affected individuals in this family would be a great asset for the future characterization of the penetrance factors influencing clinical manifestations of *PRPF31*-linked RP if additional lymphoblast cell lines were to be created and analyzed from these family members.

Splicing mutations, most of which lead to exon skipping and intron retention [Cooper et al., 2009], account for approximately 10% of all diseases caused by point mutations [Wang and Cooper, 2007] and the vast majority have been shown to affect natural splice sites and their surrounding canonical sequences [Teraoka et al., 1999]. The creation of additional splice sites in the middle of an intron which leads to the formation of pseudo exons is relatively rare [Kralovicova et al., 2005], with only a small number of genes having been reported so far (including *CEP290*, associated with Leber congenital amaurosis [den Hollander et al., 2006], β-globin [Treisman et al., 1983], factor VIII [Bagnall et al., 1999], *CFTR* [Highsmith et al., 1994; Chillon et al., 1995], and a few others). This is perhaps because screening for deep intronic mutations is not routinely conducted, despite, as demonstrated here, such mutations can have dramatic effects on splicing and consequently be harmful for the cell. Our data support therefore the relatively recent concept that mutational screens should not be limited to coding regions, but extended to intronic sequences as well.

Since intronic sequencing requires large processing capacity, we primarily utilized UHT sequencing and, as a complementary method, conventional Sanger sequencing. Although the

genomic variations identified using these two methods largely overlapped, the pathogenic mutation was surprisingly not identified using the UHT method. Alignments of the ~30 bp sequences that were generated by UHT sequencing could not identify the c.1374+654C>G change, since it was present in the first of 7 nearly-identical repetitive sequences of 56 base pairs and therefore could not be correctly matched. Specifically, the mutation was only present in 1/14 of the sequences aligned with the repeat consensus, and this low signal was undistinguishable from noise. Therefore, despite the numerous advantages of the new UHT system, small read-length and complex repetitive genomic elements remain an issue for UHT sequencing methods [Hardiman, 2008; Pop and Salzberg, 2008]. However, projected improvements in sequence length and alignment assembly for UHT technologies are likely to overcome these problems in the near future, allowing the possibility to routinely screen entire candidate genes to identify disease-causing mutations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

Abramoff MD, Magelhaes PJ, Ram SJ. Image Processing with ImageJ. Biophotonics International. 2004; 11:36–42.

Abu-Safieh L, Vithana EN, Mantel I, Holder GE, Pelosini L, Bird AC, Bhattacharya SS. A large deletion in the adRP gene PRPF31: evidence that haploinsufficiency is the cause of disease. Mol Vis. 2006; 12:384–388. [PubMed: 16636657]

Al Maghtheh M, Inglehearn CF, Keen TJ, Evans K, Moore AT, Jay M, Bird AC, Bhattacharya SS. Identification of a sixth locus for autosomal dominant retinitis pigmentosa on chromosome 19. Hum Mol Genet. 1994; 3:351–354. [PubMed: 8004108]

Al Maghtheh M, Vithana E, Tarttelin E, Jay M, Evans K, Moore T, Bhattacharya S, Inglehearn CF. Evidence for a major retinitis pigmentosa locus on 19q13.4 (RP11) and association with a unique bimodal expressivity phenotype. Am J Hum Genet. 1996; 59:864–871. [PubMed: 8808602]

Bagnall RD, Waseem NH, Green PM, Colvin B, Lee C, Giannelli F. Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two haemophilia A patients. Br J Haematol. 1999; 107:766–771. [PubMed: 10606882]

Berson EL. Retinitis pigmentosa. The Friedenwald Lecture. Invest Ophthalmol Vis Sci. 1993; 34:1659–1676. [PubMed: 8473105]

Berson EL. Long-term visual prognoses in patients with retinitis pigmentosa: the Ludwig von Sallmann lecture. Exp Eye Res. 2007; 85:7–14. [PubMed: 17531222]

Berson EL, Gouras P, Gunkel RD, Myrianthopoulos NC. Dominant retinitis pigmentosa with reduced penetrance. Arch Ophthalmol. 1969; 81:226–234. [PubMed: 5764686]

Berson EL, Simonoff EA. Dominant retinitis pigmentosa with reduced penetrance. Further studies of the electroretinogram. Arch Ophthalmol. 1979; 97:1286–1291. [PubMed: 454264]

Chillon M, Dork T, Casals T, Gimenez J, Fonknechten N, Will K, Ramos D, Nunes V, Estivill X. A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+1.6kbA-->G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. Am J Hum Genet. 1995; 56:623–629. [PubMed: 7534040]

Cooper TA, Wan L, Dreyfuss G. RNA and disease. Cell. 2009; 136:777–793. [PubMed: 19239895]

Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R. Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. Genomics. 1990; 6:575–577. [PubMed: 2184120]

den Hollander AI, Koenekoop RK, Yzer S, Lopez I, Arends ML, Voesenek KE, Zonneveld MN, Strom TM, Meitinger T, Brunner HG. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. Am J Hum Genet. 2006; 79:556–561. and others. [PubMed: 16909394]

Hardiman G. Ultra-high-throughput sequencing, microarray-based genomic selection and pharmacogenomics. Pharmacogenomics. 2008; 9:5–9. [PubMed: 18154443]

Hartong DT, Berson EL, Dryja TP. Retinitis pigmentosa. Lancet. 2006; 368:1795–1809. [PubMed: 17113430]

Highsmith WE, Burch LH, Zhou Z, Olsen JC, Boat TE, Spock A, Gorvoy JD, Quittel L, Friedman KJ, Silverman LM. A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. N Engl J Med. 1994; 331:974–980. and others. [PubMed: 7521937]

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. Whole-genome patterns of common DNA variation in three human populations. Science. 2005; 307:1072–1079. [PubMed: 15718463]

Iseli C, Ambrosini G, Bucher P, Jongeneel CV. Indexing strategies for rapid searches of short words in genome sequences. PLoS ONE. 2007; 2:e579. [PubMed: 17593978]

Kohn L, Bowne SJ, L SS, Daiger SP, Burstedt MS, Kadzhaev K, Sandgren O, Golovleva I. 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. 2008

Kralovicova J, Christensen MB, Vorechovsky I. Biased exon/intron distribution of cryptic and de novo 3′ splice sites. Nucleic Acids Res. 2005; 33:4882–4898. [PubMed: 16141195]

Liu J, Dai X, Sheng J, Cui X, Wang X, Jiang X, Tu X, Tang Z, Bai Y, Liu M. Identification and functional characterization of a novel splicing mutation in RP gene PRPF31. Biochem Biophys Res Commun. 2008; 367:420–426. and others. [PubMed: 18177735]

McGee TL, Berson EL, Dryja TP. Identification Of Novel Mutations In The PRPF31 Gene (RP11) In Patients With Autosomal Dominant Retinitis Pigmentosa With Reduced Penetrance. Investigative Ophthalmology and Visual Science. 2002; 43:792.

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. [PubMed: 9345108]

Myers EW, Miller W. Optimal alignments in linear space. Comput Appl Biosci. 1988; 4:11–17. [PubMed: 3382986]

Pedersen AG, Nielsen H. Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. Proc Int Conf Intell Syst Mol Biol. 1997; 5:226–233. [PubMed: 9322041]

Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology. Trends Genet. 2008; 24:142–149. [PubMed: 18262676]

Rio Frio T, Civic N, Ransijn A, Beckmann JS, Rivolta C. Two trans-acting eQTLs modulate the penetrance of PRPF31 mutations. Hum Mol Genet. 2008a; 17:3154–3165. [PubMed: 18640990]

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. 2008b; 118:1519–1531. [PubMed: 18317597]

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. [PubMed: 16708387]

Rivolta C, Sharon D, DeAngelis MM, Dryja TP. Retinitis pigmentosa and allied diseases: numerous diseases, genes, and inheritance patterns. Hum Mol Genet. 2002; 11:1219–1227. [PubMed: 12015282]

Rougemont J, Amzallag A, Iseli C, Farinelli L, Xenarios I, Naef F. Probabilistic base calling of Solexa sequencing data. BMC Bioinformatics. 2008; 9:431. [PubMed: 18851737]

Sullivan LS, Bowne SJ, Seaman CR, Blanton SH, Lewis RA, Heckenlively JR, Birch DG, Hughbanks-Wheaton D, Daiger SP. Genomic rearrangements of the PRPF31 gene account for 2.5% of autosomal dominant retinitis pigmentosa. Invest Ophthalmol Vis Sci. 2006; 47:4579–4588. [PubMed: 17003455]

Teraoka SN, Telatar M, Becker-Catania S, Liang T, Onengut S, Tolun A, Chessa L, Sanal O, Bernatowska E, Gatti RA. Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. Am J Hum Genet. 1999; 64:1617–1631. and others. [PubMed: 10330348]

Treisman R, Orkin SH, Maniatis T. Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. Nature. 1983; 302:591–596. [PubMed: 6188062]

Vithana E, Al-Maghtheh M, Bhattacharya SS, Inglehearn CF. RP11 is the second most common locus for dominant retinitis pigmentosa. J Med Genet. 1998; 35:174–175. [PubMed: 9556378]

Vithana EN, Abu-Safieh L, Allen MJ, Carey A, Papaioannou M, Chakarova C, Al Maghtheh M, Ebenezer ND, Willis C, Moore AT. 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. and others. [PubMed: 11545739]

Vithana EN, Abu-Safieh L, Pelosini L, Winchester E, Hornan D, Bird AC, Hunt DM, Bustin SA, Bhattacharya SS. 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. [PubMed: 14507862]

Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet. 2007; 8:749–761. [PubMed: 17726481]

Xu S, Nakazawa M, Tamai M, Gal A. Autosomal dominant retinitis pigmentosa locus on chromosome 19q in a Japanese family. J Med Genet. 1995; 32:915–916. [PubMed: 8592343]
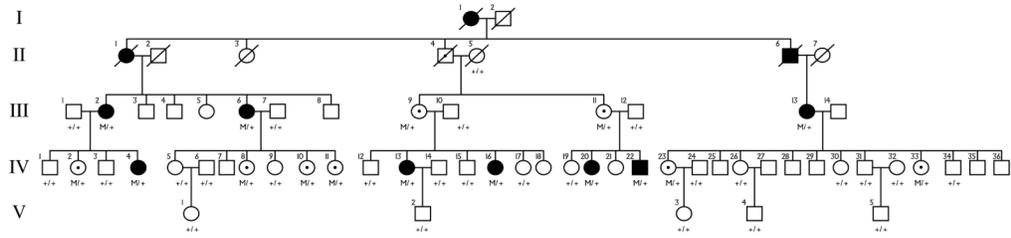
Figure 1. Schematic pedigree of family #1562

Black and white symbols indicate affected and unaffected individuals, respectively. Dots indicate unaffected carriers of the mutation. The c.1374+654C>G mutation is represented with "M"; wild-type *PRPF31* alleles with "+".

Figure 2. Genomic localization and sequence of the VNTR containing the c.1374+654C>G variant

(A) Numbered rectangles represent *PRPF31* exons. The c.1374+654C>G mutation is comprised within the first of 7 56-bp consecutive repeat elements (shown as small rectangles), constituting a VNTR that starts at position 638 of intron 13. The drawing is to scale. (B) Sequence and structure of the VNTR. Repeated elements are shown in alternating red and blue colors, and the c.1374+654C>G substitution is highlighted in yellow. The newly-created donor splice site is underlined, while the portion of intron 13 that is retained in mutant isoforms is indicated in italics. Numbers on the left refer to the relative position of nucleotides, with respect to the beginning of intron 13.
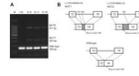
Figure 3. PRPF31 mRNA isoforms in LCLs from members of family #1562
(A) A representative image of RT-PCR products, run on a 1.5% agarose gel, derived from the exon 13-exon 14 mRNA segment from 2 asymptomatic (III-9, III-11) and 1 affected (IV-16) individuals from family #1562, as well as from 1 CEPH individual (Ctrl). "M" is 100 bp ladder. Length and names of PCR products are indicated on the right side of the gel.
(B) Schematic representation of the 3 mRNA isoforms from which PCR products are derived. Numbered rectangles and bold lines represent exons and introns of *PRPF31*, respectively. Hatched rectangles designate parts of intron 13 that were retained (starts and ends are indicated by numbers). Numbering of stop codons for each mRNA isoform is based on the wild-type sequence. Depicted exons and introns are not to scale.
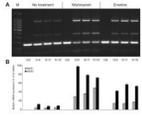
Figure 4. Rescue of NMD-sensitive mRNA isoforms MUT1 and MUT2 by wortmannin and emetine

(A) Agarose gel showing RT-PCR products for the exon 13-exon 14 *PRPF31* region, obtained from mRNA extracted from the same LCLs shown in Fig. 3. "M" is a 100 bp ladder. Apparent variable expression of wild-type bands over different conditions reflects a technical artifact due to variable and inhibitor-specific efficiency, for which wild-type transcripts compete differentially with mutant isoforms for PCR resources. (B) Quantification of PCR products by the combination of both densitometry values of agarose gel bands (ImageJ software) and the integration of peaks from capillary electrophoresis (Biocalculator software). Data shown summarize measurements of 3 PCR replicates from 3 culturing conditions on 3 independent cultures for each cell line. Grey and black bars represent the MUT1/wild-type and MUT2/wild-type ratios for each cell line and each treatment. Supplementation with either emetine or wortmannin partly restored the expression of NMD-sensitive mutant mRNA from LCLs from patient cell lines and had no effect on the control cell line.
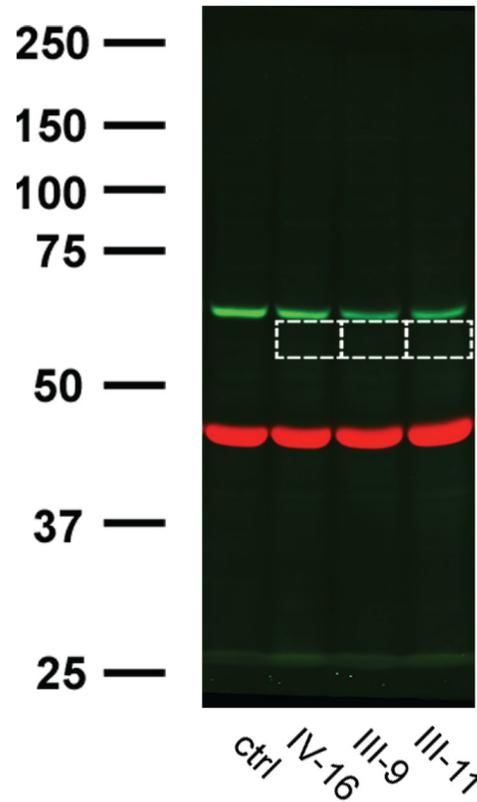
Figure 5. Protein analysis in 3 members from family #1562 and one control
SDS-PAGE (10%) of total protein extracts from LCLs from 1 affected (IV-16) and 2 asymptomatic (III-9, III-11) carriers of the mutation, and from a non-carrier of any *PRPF31* mutation (ctrl). Simultaneous infrared detection of proteins was achieved with an N-terminal anti-PRPF31 antibody (green) and a β-actin control (red). A single band corresponding to the size of wild-type PRPF31 protein was detected. White dotted boxes indicate the expected position of mutant proteins derived from MUT1 and MUT2 mRNAs.

Table1

PRPF31 heterozygous variants detected in individual IV-16

PRPF31 position[a]	Genotype[b]	Variation name
355	A/G	rs4806711:A>G
684	A/G	rs35705606:A>G
6,601	T/G	c.322+115T>G[d]
6,628	A/G	rs59977379:A>G
6,929	C/T	c.323−111C>T[d]
7,219	C/G	rs56220912:C>G
8,328	G/A	rs1058572:G>A[d]
8,626	G/A	rs56234781
10,054	−/dup CAAG	c.855+855_855+858dup[d]
10,861	T/C	rs254272:T>C
11,372	C/A	rs10424816:C>A
11,799	T/G	c.945+643T>G[d]
11,921	G/C	rs254271:G>C
12,261	G/A	rs10853869:G>A
12,295	T/C	c.946−317T>C[d]
12,794	C/T	rs34990810:C>T
13,255	G/A	rs2556367:G>A
14,563[c]	C/G	c.1374+654C>G[d]

[a]Numbering is relative to the first base of exon 1.

[b]RefSeq allele/sequenced allele.

[c]Not detected by UHT sequencing.

[d]Nucleotide numbering reflects coding DNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the *PRPF31* GenBank reference sequence NM_015629.3, according to the Human Genome Variation Society guidelines (www.hgvs.org/mutnomen).