Human Molecular Genetics, 2008, Vol. 17, No. 20 **3154–3165**
doi:10.1093/hmg/ddn212
Advance Access published on July 18, 2008

Two *trans*-acting eQTLs modulate the penetrance of *PRPF31* mutations

Thomas Rio Frio[1], Natacha Civic[1], Adriana Ransijn[1], Jacques S. Beckmann[1,2,†] and Carlo Rivolta[1,†*]

[1]Department of Medical Genetics, University of Lausanne, Rue du Bugnon 27, Lausanne 1005, Switzerland and [2]Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne 1011, Switzerland

Received April 22, 2008; Revised June 27, 2008; Accepted July 17, 2008

Dominant mutations in the gene encoding the ubiquitously-expressed splicing factor PRPF31 cause retinitis pigmentosa, a form of hereditary retinal degeneration, with reduced penetrance. We and others have previously shown that penetrance is tightly correlated with *PRPF31* expression, as lymphoblastoid cell lines (LCLs) from affected patients produce less abundant *PRPF31* transcripts than LCLs from their unaffected relatives carrying the same mutation. We have investigated the genetic elements determining the variable expression of *PRPF31*, and therefore possibly influencing the penetrance of its mutations, by quantifying *PRPF31* mRNA levels in LCLs derived from 15 CEPH families (200 individuals), representative of the general population. We found that *PRPF31* transcript abundance was a highly variable and highly heritable character. Moreover, by linkage analysis we showed that *PRPF31* expression was significantly associated with at least one expression quantitative trait locus (eQTL), spanning a 8.2-Mb region on chromosome 14q21–23. We also investigated a previously mapped penetrance factor located near *PRPF31* itself in LCLs from individuals belonging to selected families segregating *PRPF31* mutations that displayed reduced penetrance. Our results indicate that, despite its constant association with the non-mutant allele, this factor was able to modulate the expression of both *PRPF31* alleles. Furthermore, we showed that LCLs from affected patients have less *PRPF31* RNA than those of asymptomatic patients, even at the pre-splicing stage. Altogether, these data demonstrate that *PRPF31* mRNA expression and consequently the penetrance of *PRPF31* mutations is managed by diffusible compounds encoded by at least two modifiers, acting in a co-regulatory system on both *PRPF31* alleles during transcription.

INTRODUCTION

Retinitis pigmentosa (RP) is a heterogeneous group of neurodegenerative diseases affecting the retina by causing the progressive death of photoreceptor cells. RP occurs alone or as a part of more complex syndromes. Evolution of RP is an exponential progression of symptoms, usually night blindness, loss of mid-peripheral visual field, loss of peripheral vision by visual field constriction, and in many cases complete or near-complete blindness, as a final stage. Physiologically, first the outer segment of photoreceptors and then the outer nuclear layer of the retina disappear, followed by the degeneration of the inner and ganglion cell layers (1). Clinically, analysis of retinal fundi typically exhibits bone spicule-like pigmentary deposits, due to retinal pigmented epithelium migration in the retina (2), and electroretinograms show reduced or absent amplitudes and delayed signals (3). RP is genetically highly heterogeneous and mostly a monogenic disorder. It is transmitted as a Mendelian autosomal dominant, autosomal recessive or X-linked character, and in small proportions also as a digenic and a non-Mendelian trait (4). So far, mutations in almost 80 genes have been found to cause RP (http://www.sph.uth.tmc.edu/Retnet/).

Some of the genes causing RP encode proteins which are exclusively present or have a predominant role in the retina. Other RP genes are ubiquitously expressed and encode proteins that do not have any known retina-specific function. The mechanisms by which mutations in these genes, that are

*To whom correspondence should be addressed. Tel: +41 216925451; Fax: +41 216925455; Email: carlo.rivolta@unil.ch
†The authors wish it to be known that, in their opinion, the last two authors have equally contributed to this work.

sometimes essential and highly conserved, produce a retina-restricted phenotype remain unknown. PRPF31, a protein component of the spliceosome complex, is ubiquitously expressed and essential for cell survival (5). Distinct heterozygous *PRPF31* mutations were shown, however, to cause autosomal dominant RP (ADRP) with no associated syndromic symptoms (6–26), and were all found to map to the *RP11* locus on chromosome 19q13.4 (13,27), which accounts for 5% of ADRP cases (1).

A particular clinical feature of *RP11*-linked RP is the presence, in some pedigrees, of asymptomatic carriers of pathogenic mutations showing no impaired vision and normal to slightly reduced electroretinographic recordings, suggesting incomplete penetrance of symptoms. These individuals are rarely identified with a routine clinical examination, since they never develop any significant clinical manifestations, even at old ages, but can generate offspring that are affected with RP (6–12,14,18–20,22–26). Linkage analyses performed with markers located around the *RP11* locus have demonstrated that the penetrance of mutations is presumably determined by particular alleles, termed isoalleles, that are inherited from the parent who does not carry the mutation (i.e. they are *in trans* with respect to the mutation), and are located close to or within the *RP11* locus itself (Fig. 1) (12,13). These isoalleles (from now on called '*RP11*-associated isoalleles') are assumed to be present in the general population and would not produce any clinical phenotype by themselves, but may determine or modulate the pathogenic effects of *PRPF31* mutations, when present. However, *RP11*-associated isoalleles alone cannot explain penetrance in some families segregating *PRPF31* mutations, indicating that other modifiers could be involved (12).

Since retinal biopsies are impossible and *PRPF31* is ubiquitously expressed, lymphoblastoid cell lines (LCLs) from carriers of mutations have been used by us and others as an excellent biological proxy for photoreceptors in *PRPF31* molecular studies (21,28,29). Such investigations have allowed to determine that haploinsufficiency is the likely cause of the disease, as the large majority of *PRPF31* mutations result in functionally null alleles. Specifically, LCLs from unaffected controls were shown to express substantially more *PRPF31* mRNA than LCLs from carriers of *PRPF31* mutations, as mutant *PRPF31* mRNA alleles in both affected and asymptomatic patients are either degraded by the nonsense-mediated mRNA decay (29) or absent due to the deletion of one copy of the gene (19,22). Very interestingly, LCLs from asymptomatic patients express higher levels of the non-mutant *PRPF31* mRNA allele than LCLs from affected patients (21,26,28). This indicates that, regardless of the specific mutation, higher expression of functional *PRPF31* mRNA may protect asymptomatic carriers from the disease and explain incomplete penetrance. A similar mechanism has been identified in at least two other dominant diseases with incomplete penetrance. Specifically, penetrance of mutations in the α-subunit of the spectrin gene (*SPTA1*) and in the ferrochelatase gene (*FECH*), leading to hereditary elliptocytosis and erythropoietic protoporphyria, respectively, is also determined by variations in the expression of the wild-type allele, modulated by a *cis*-acting polymorphism (30,31). Consequently, it can be hypothesized that genetic modifiers can determine the clinical status of carriers of *PRPF31* mutations, by modulating the expression of non-mutant *PRPF31* alleles.

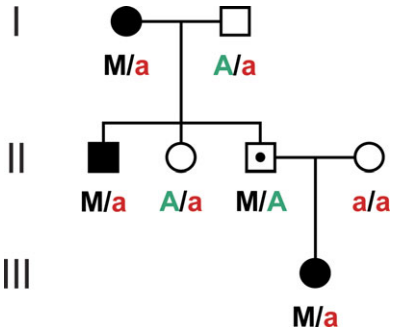

Figure 1. *RP11*-associated isoalleles and penetrance of *PRPF31* mutations, according to the model by McGee *et al*. (12). A hypothetical family segregating the *PRPF31* mutation M is depicted. *RP11*-associated isoalleles (A), inherited from the parent that does not carry the mutation and polymorphic in the general population, protect an asymptomatic patient (symbol with a dot) from the disease. Conversely, when this isoallele is not present (a), *PRPF31* mutations cause RP.

The aim of the present study is to identify the genetic elements that control the expression of *PRPF31* and therefore to gain insights into the heritable factors involved in determining the penetrance of its mutations.

RESULTS

PRPF31 expression and its heritability

Variable *PRPF31* expression was proven to be directly correlated with penetrance of its mutations (21,26,28). We first investigated whether differential amounts of *PRPF31* mRNA could be a consequence of polymorphic copy number variations (CNVs), found to be important factors in human phenotypic variability (32). A scan of the Database of Genomic Variants (33) for possible CNV sequences spanning the *PRPF31* region revealed the absence of polymorphic variations in copy number. Direct analysis by genomic DNA real-time PCR of all *PRPF31* exons in one affected and two asymptomatic patients from the same family, as well as in three control individuals, also showed that there were neither deletions nor duplications, in any of the samples (Supplementary Material, Fig. S1).

We then wanted to refine the mapping of *RP11*-associated isoalleles and, possibly, detect new modulators of *PRPF31* expression. For this purpose, we measured the expression of *PRPF31* mRNA in 200 LCLs from 15 CEPH families, by highly sensitive real-time PCR set-ups. Thirteen of these pedigrees were three-generation families (grand parents, parents and children), while the remainder were composed of parents and children only, with an overall average of 9.5 sibs per family. For each individual, we normalized the expression of *PRPF31* using *GAPDH* transcripts as endogenous controls, described as reliable standards for correlating *PRPF31* expression and clinical status of LCL donors, according to our results (not shown) and data from other groups (28). The relative expression was calculated with respect to the median *PRPF31* expressor, arbitrarily set to an expression value of 1.0. The distribution of the phenotypes showed that the amount of *PRPF31* transcripts were far from being

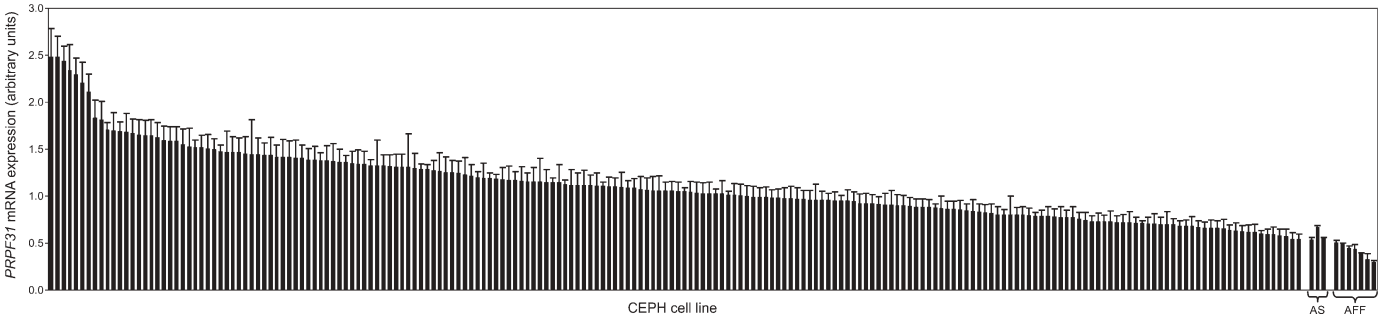

Figure 2. Quantification of *PRPF31* mRNA expression in lymphoblastoid cell lines (LCLs). Sorted expression levels of *PRPF31* mRNA in LCLs from 200 control individuals (CEPH), 3 asymptomatic (AS) and 7 affected (AFF) patients with *PRPF31* mutations are shown, as measured by real-time PCR. *PRPF31* mRNA expression is normalized to a median expressor, arbitrarily set to have a value of 1.0. Error bars refer to the standard error of the mean.

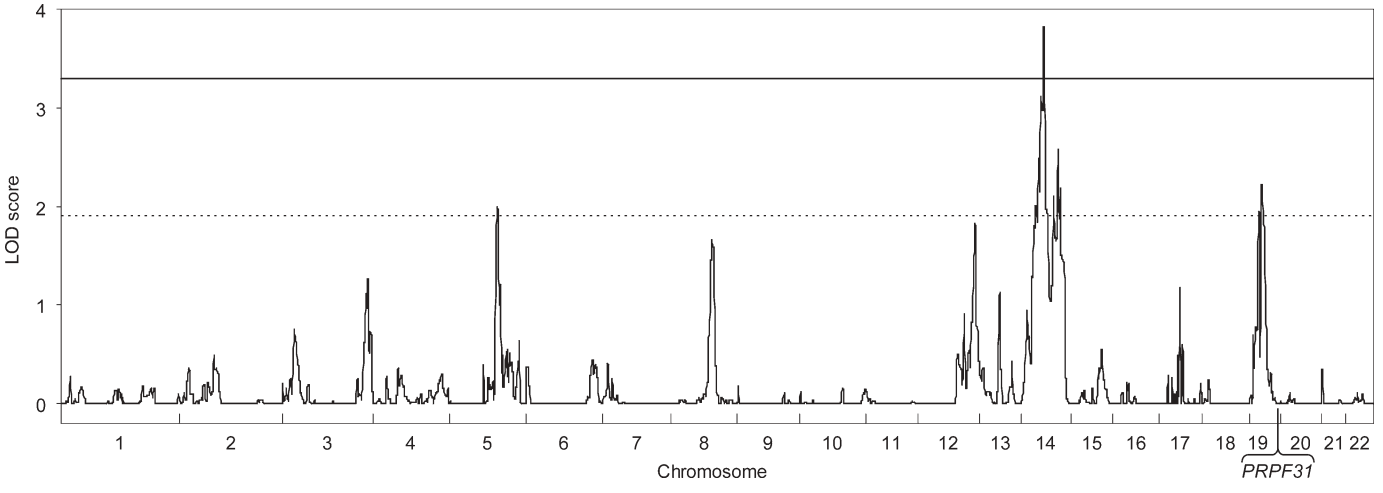

Figure 3. Results of genome-wide quantitative multipoint linkage analysis. Result plot obtained from the regress-based method implemented in the MERLIN software (53). The continuous horizontal line indicates the threshold for significant LOD scores and the dotted line is the threshold for 'suggestive linkage', as defined by Lander and Kruglyak (56). A significant peak with a LOD score of 3.83 ($P = 1 \times 10^{-5}$) was detected at 69.36 cM from the pter of chromosome 14. Four other LOD scores suggestive of linkage were observed on chromosome 5, 14 and 19 (containing two peaks). The location of *PRPF31* on chromosome 19 is indicated.

constant in the general population, as they varied from 0.53 to 2.48 arbitrary units (a.u.), representing a ~5-fold change between the lowest and the highest expressor (Fig. 2).

To confirm the robustness of these data, *PRPF31* expression was also measured in seven LCLs from clinically affected patients carrying one of the following *PRPF31* mutations: c.177+1delG, c.323-2A>G, c.877_910del, c.319C>G and c.856-2A>G, that were previously found to express pathologically low levels of *PRPF31* transcripts (21). In these cells, normalized *PRPF31* expression varied from 0.29 to 0.50 a.u. (Fig. 2) and was, as expected, lower than in the general population ($P = 2.5 \times 10^{-6}$, by *t*-test). Further, such measurements in LCLs from three asymptomatic carriers from these same families showed that *PRPF31* expression levels ranged from 0.53 to 0.67 a.u. As expected, these values were lower than, but within the same range as those resulting from LCLs from the normal population, while still being higher than those detected in LCLs from affected patients ($P = 0.01$, by *t*-test).

PRPF31 expression in the CEPH cell lines was measured to calculate heritability of this molecular phenotype. Specifically, we found heritability to be 0.45 ($P = 5.09 \times 10^{-5}$), as assessed by the SOLAR (Sequential Oligogenic Linkage Analysis Routines) software (34), indicating that variable *PRPF31* expression is a highly heritable trait in the general population. We also estimated the likelihood of detecting significant eQTL (expression quantitative trait locus) to be 70%, given the size of our CEPH panel and such a high heritability.

eQTL linkage analysis

We performed multipoint genome-wide linkage analyses using the MERLIN REGRESS and MERLIN VC algorithms (34) on all 200 CEPH samples to identify chromosomal regions potentially involved in the regulation of *PRPF31* expression. LOD scores from MERLIN REGRESS and MERLIN VC were highly correlated ($r = 0.98$, by the Pearson's test), indicating that the results obtained were indeed robust. Since the MERLIN REGRESS algorithm has been shown to be more powerful in large sibships (≥ 3) than MERLIN VC (35), the latter method was used only as an internal control. A significant linkage peak (LOD score = 3.83, $P = 1 \times 10^{-5}$, Fig. 3, Table 1) was found on chromosome 14 at 69.36 cM from pter. To assess the robustness of this result, we randomly separated 12 times the

Table 1. Summary of the genome-wide multipoint linkage analysis performed using the MERLIN REGRESS algorithm

Chromosome	Marker	Position (cM)	LOD score	*P*-value
5	rs2043833	167.14	1.99	1.2×10^{-3}
14	rs1889383	69.36	3.83	1×10^{-5}
14	rs1959287	159.1	2.58	3×10^{-4}
19	rs1862471	30.68	1.95	1.4×10^{-3}
19	rs1122713	40.69	2.23	7×10^{-4}

Only significant ($\geq$3.3) and suggestive ($\geq$1.9) LOD scores are reported.

15 families analyzed into a 'test' (8 families) and a 'validation' group (7 families). For each of these sets, eQTL analyses were re-assessed by the same methods, with markers belonging to chromosome 14. We found that, for all 12 simulations, the peak originally detected in the 15 families was always present in each of the 'test' sets and in the corresponding 'validation' sets. These results indicate that all families analyzed likely contributed to the eQTL signal on chromosome 14, and that we can confidently exclude the possibility of a false-positive result in the original set of 200 cell lines.

Four other suggestive linkage sites, indicated in Table 1, were also revealed on chromosomes 5, 14 and 19 (containing two peaks). Noteworthy, the *RP11* locus is also located on chromosome 19, however, it is very distant from both these loci. The LOD-1 95% confidence interval (CI) (36) of the significant linkage peak in the 14q22.1–23.1 region identified a 8.2-Mb interval containing 97 genes, as deduced from the genome database at NCBI (National Center for Biotechnology Information), build 36.2 (www.ncbi.nlm.nih.gov). Of these, at least 50 were previously assessed to be expressed in LCLs by whole-genome microarray expression analysis of 210 LCLs (37) and may have therefore contributed to this eQTL. Among these 50 genes, there were some transcription factors or genes involved in transcription regulation such as *CGR19*, *WDHD1* and *TXNDC1* that can potentially direct or modulate *PRPF31* mRNA expression. Furthermore, the region identified has been shown to have a major role in retinal physiology as it contains three genes: *OTX2*, *BMP4* and *SIX6* that are strongly involved in retinal development and maintenance (38–40). Seventeen out of these 50 expressed genes correspond to proteins with no known established function.

Considering this newly identified regulator of *PRPF31* expression localized on chromosome 14, designated from this point onwards as the '*RP11*-distant regulator', we propose a classical model for the function of this eQTL (Fig. 4A and B). In this model, the modifier gene, which is variably expressed because of a common polymorphism, would encode a diffusible element (e.g. a transcription factor) modulating the expression of *PRPF31*. In this scenario, irrespective of the total expression, the two alleles of *PRPF31* would always be expressed in a 50:50 ratio.

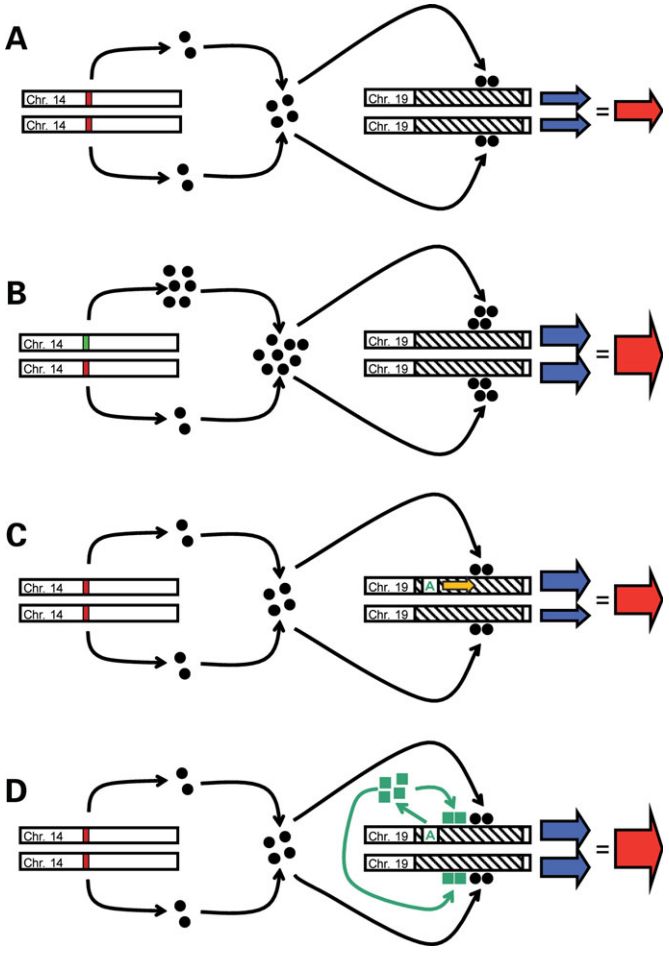

Figure 4. Models for regulation of *PRPF31* expression. Diagrammatic representation of various models by which the *RP11*-distant regulator eQTL identified in this study and *RP11*-associated isoalleles may affect the expression of *PRPF31*. Open rectangles are the graphical representations of segments of interest of chromosomes 14 and 19. Red and green rectangles on chromosome 14 correspond to polymorphic alleles of the *RP11*-distant regulator. Black circles represent proteins or other diffusible modulators produced by this eQTL. Hatched rectangles indicate the *PRPF31* locus. Blue and red arrows symbolize allelic and total *PRPF31* mRNA amounts, respectively, their thickness being proportional to the rate of expression. (**A** and **B**) Schematic representation of an individual who does not carry any *RP11*-associated isoallele. In a heterozygote individual for a polymorphic allele in the *RP11*-distant regulator, one eQTL allele produces more modifier than the other, resulting eventually in increased *PRPF31* total mRNA production. Regardless of the genotype at the *RP11*-distant regulator locus and of the total *PRPF31* mRNA production, each *PRPF31* allele contribute equally to gene expression. (**C**) Model of an individual carrying an *RP11*-associated isoallele (green 'A'). Hypothesizing that this isoallele acts *in cis*, inducing *PRPF31* expression (orange arrow), an overexpression of the corresponding *PRPF31* allele should be observed, compared to the allele *in trans*. (**D**) Alternatively, *RP11*-associated isoalleles producing a diffusible element (green squares) could act *in cis* and *in trans*, increasing with equal strength the amount of *PRPF31* pre-mRNA and mRNA derived from both alleles.

PRPF31 mRNA allelic expression in the general population

The reduced penetrance of hereditary elliptocytosis (30) and erythropoietic protoporphyria (31) is modulated in each case by the variable expression of the wild-type allele, which is in turn directly influenced by the presence of a polymorphism *in cis* with it, i.e. *in trans* with respect to the mutation. Based on these two well-documented dominant diseases and by adapting these paradigms to our current model of an *RP11*-distant regulator, we investigated whether the expression of the *PRPF31* wild-type allele could be enhanced

by *RP11*-associated isoalleles acting as *cis* regulatory elements (Fig. 4C).

To test this model, we quantified in our set of CEPH cell lines the individual expression of both *PRPF31* alleles, whenever these could be specifically identifiable by molecular tags. Specifically, we selected all SNPs within the *PRPF31* transcript sequence, from the Ensembl database (www.ensembl.org) and from our own sequencing data, that create or abolish a restriction site and are present in at least one pedigree among all tested CEPH families. Overall, we identified 2 SNPs that matched these criteria and 65 CEPH individuals from 11 families were found to carry one of these 2 SNPs. No cell line was found to carry both. One (rs1058572 in exon 7) is an A/G transition creating an *Ear*I restriction site and the second is a C/T transition in exon 14 (provisionally reported in the Ensembl database as ENSSNP6991476) creating an *Mse*I restriction site. For all cell lines displaying heterozygosity at one of these sites, we found that the allelic ratio of *PRPF31* expression was approximately 50:50, regardless of the total expression (Fig. 5). This indicates that the mRNAs derived from both *PRPF31* alleles are present in equal amounts, disfavoring the presence of any modifiers acting exclusively *in cis* with respect to a given allele in the tested set of CEPH cell lines.

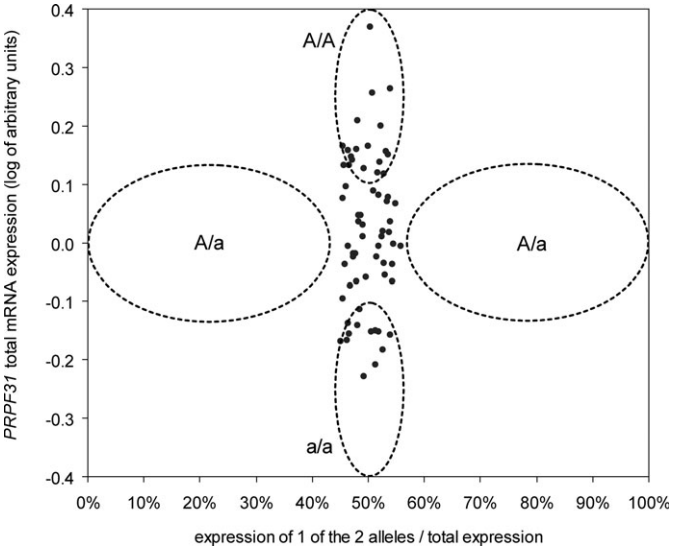

Figure 5. *PRPF31* mRNA allelic expression in CEPH cell lines. Data (black dots) are represented as the ratio of one allele over the total *PRPF31* mRNA expression, and were generated by RFLP analyses of *PRPF31* coding SNPs present in 65 heterozygous CEPH cell lines. Total expression of *PRPF31* mRNA was measured by real-time PCR. Dotted ovals delimit expected values in the hypothesized model of an exclusive *cis*-acting regulation of *PRPF31* expression by the *RP11*-associated isoalleles. Under this assumption, one allele should be more expressed than the other in LCLs from heterozygous individuals (with the genotypes Aa). 'A' and 'a' stand for the presence or absence of the *RP11*-associated isoallele, respectively, as defined in Fig. 1.

Effects of *RP11*-associated isoalleles on *PRPF31* expression

The absence of *cis*-acting *PRPF31* expression modulators in the CEPH cell lines analyzed does not exclude *per se* that some *RP11*-associated isoalleles are actually present in such a set, if they act *in cis* as well as *in trans* (Fig. 4D). Indeed, previous studies in families segregating such isoalleles did hypothesize but never demonstrated that, similar to *FECH* and *SPTA1*, these modifiers also act exclusively *in cis* (12,28). To gain insights into the mechanisms of action of *RP11*-associated isoalleles, we performed total and allelic quantification in LCLs from definite carriers of such isoalleles, with or without *PRPF31* mutations, as ascertained by previous investigations on *RP11* families, and having various clinical classifications (Table 2). We analyzed LCLs from four asymptomatic carriers and five affected patients from three *RP11* families for which individual alleles could be distinguished because of sequence variations. Given that mature mRNA from carriers of heterozygous *PRPF31* mutations is not suitable for allelic expression quantification, because only mRNA derived from wild-type alleles exists at the steady state (29), we quantified the allelic expression of *PRPF31* at the nuclear pre-mRNA level. The same was repeated on the one individual, referenced as AG261, who is a non-carrier of *PRPF31* mutations but is an obligate heterozygous carrier of an *RP11*-associated isoallele. This individual is the spouse of a carrier of a *PRPF31* mutation from family #AD5 and has five children, of whom one affected and two asymptomatic patients underwent clinical examination (8,41).

Results clearly showed that LCLs from all carriers of *RP11*-associated isoalleles (asymptomatics and AG261) had more *PRPF31* pre-mRNA than those from non-carriers, i.e. affected patients ($P = 8.38 \times 10^{-4}$, by *t*-test, Fig. 6), and that this increase pertained equally to both wild-type and mutant alleles. This indicates that, similar to *RP11*-distant regulators, *RP11*-associated isoalleles enhance *PRPF31* transcription by acting on both alleles. Because of this phenomenon, an ancillary conclusion is therefore that some of the high LCLs expressors from the CEPH panel (Fig. 2) could possibly represent unascertained carriers of *RP11*-associated isoalleles that, unlike all expressors measured in the past, do not carry any *PRPF31* mutations. This same situation is shared by cell line AG261, which would actually be the fifth highest expressor if its total mRNA expression is compared with all 200 CEPH cell lines analyzed initially (not shown).

DISCUSSION

Dominant mutations in *PRPF31* (or *RP11*) leading to RP are subject to reduced penetrance (6–12,14,18–20,22–25). This phenomenon, which is not due to the presence of regional CNVs, has in turn been shown to be determined by the expression levels of *PRPF31* mRNA derived from the wild-type allele (21,26,28). Although photoreceptors and lymphoblasts are very different cell types, a number of studies have demonstrated a strong link between penetrance of *PRPF31* mutations and elevated *PRPF31* expression in LCLs derived from asymptomatic carriers of mutations with respect to affected patients (21,28,29). To investigate genetic determinants of *PRPF31* expression we used an unbiased approach, independent from the presence of *PRPF31* mutations. Specifically, we took advantage of the fact that CEPH cell lines are widely genotyped and have often been used as a powerful system to detect genetic linkage for quantitative traits

Table 2. Lymphoblastoid cell lines from families showing incomplete penetrance of *PRPF31* mutations

Family ID	Individual ID	*PRPF31* mutation	Clinical status	Reference
1562	13189	Not characterized	Affected	Berson *et al*. (6,7)[b], McGee *et al*. (12), Rivolta *et al*. (21).
	12943		Asymptomatic	
	13191		Asymptomatic	
2474	14284	c.877_910del	Affected	Berson *et al*. (7)[b], McGee *et al*. (12), Rivolta *et al*. (21).
	13991		Asymptomatic	
AD5	AG293	c.1115_1125del	Affected	Moore *et al*. (8)[b], Evans *et al*. (41)[b], Al Maghteh *et al*. (10), Vithana *et al*. (13).
	AG305		Affected	
	AG307		Affected	
	AG316		Asymptomatic	
	AG340		Asymptomatic	
	AG261		Control[a]	

[a]Obligate heterozygous carrier of an *RP11*-associated isoalleles.
[b]Clinical description.

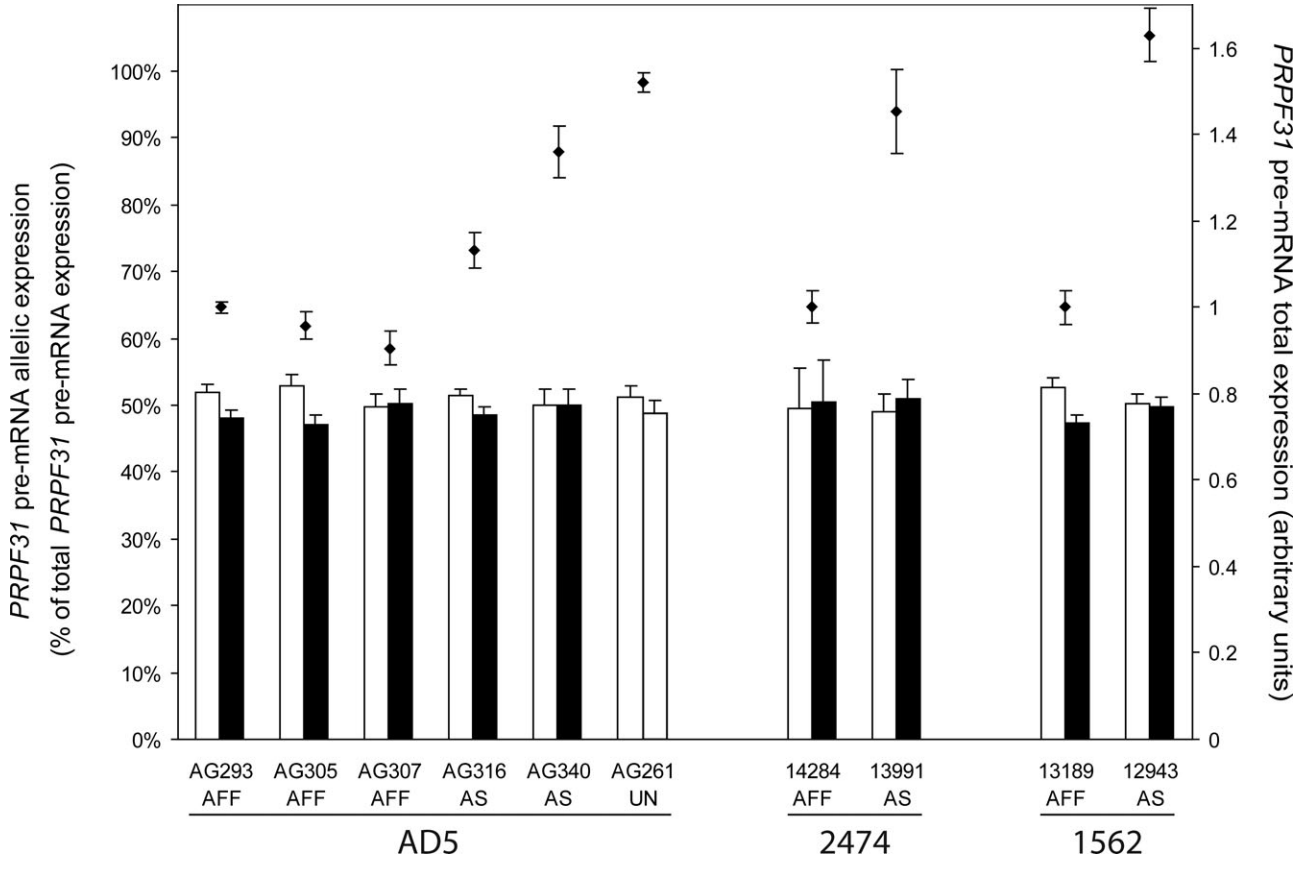

Figure 6. *PRPF31* pre-mRNA total and allelic quantification in LCLs from carriers of *RP11*-associated isoalleles. AD5, 2474 and 1562 indicate the family codes for individuals AG261, AG293, AG305, AG307, AG316, AG340, 14284, 13991, 12943 and 13189. AFF, AS and UN refer to affected and asymptomatic patients, or the unaffected control AG261, respectively. Diamond symbols indicate total *PRPF31* pre-mRNA expression (right ordinates). Bars represent allelic quantifications, as percents of total expression, of pre-mRNA from either the *PRPF31* wild-type (white) or mutant (black) allele (left ordinates). Error bars are standard errors of the mean. In all families, LCLs from carriers of an *RP11*-associated isoallele (asymptomatics and AG261) express more *PRPF31* pre-mRNA than those from affected patients, but all cell lines express the same ratio of allelic pre-mRNA, regardless of the presence of mutations and of the clinical classification of the donors. Total pre-mRNA expression is normalized relatively to the LCLs from affected individuals AG293, 14284 and 13189 for families #AD5, #2474 and #1562, respectively.

(42,43), including eQTLs (37), and analyzed LCLs from 200 control individuals from 15 families by real-time PCR. Our data show that the steady-state expression of *PRPF31* mRNA is variable and heterogeneous within the general population (Fig. 2). As expected, LCLs from seven affected patients, carrying different mutations and used as positive controls, had lower amounts of *PRPF31* mRNA than LCLs sampled from the general, unaffected population, since the mRNA derived from the mutant *PRPF31* allele was inactivated by nonsense-mediated mRNA decay (29). As a further

control, LCLs from three asymptomatic patients were shown to have amount of *PRPF31* mRNA within the range of the unaffected population and higher than those from affected patients (Fig. 2), as predicted (21,28). By significantly extending the size of the samples representing the unaffected population, we therefore confirm and reinforce what was previously demonstrated in families with *PRPF31* mutations, i.e. that low expression of *PRPF31* mRNA is the cause of the disease and that incomplete penetrance is mediated by modifiers modulating *PRPF31* mRNA expression.

We next demonstrated that *PRPF31* mRNA expression was highly heritable among the examined families ($H2r = 0.45$, $P = 5.09 \times 10^{-5}$), indicating that genetic elements are predominantly involved in modulating the amounts of *PRPF31* transcripts. We performed a whole-genome non-parametric linkage analysis to map eQTLs (Fig. 3), and identified a genomic region on chromosome 14q22.1–23.1 (LOD score = 3.83, $P = 1 \times 10^{-5}$, Table 1) that was significantly linked to variation in *PRPF31* mRNA expression levels. The 95% CI encompasses an 8.2-Mb DNA sequence and contains 97 genes. Fifty-eight of these genes have been analyzed in CEPH trios expression studies: 50 were indeed expressed in LCLs (37), 17 of which are not functionally characterized. Among these 50 genes, some encode for transcription factors or proteins that are involved in regulation of transcription such as: CGR19, WDHD1 and TXNDC1. Moreover, this region plays a major role in retinal physiology as it contains three genes: *OTX2*, *BMP4*, and *SIX6* that are strongly involved in human retinal development and maintenance (38–40). Interestingly, the gene closest to the SNP yielding the maximum LOD score is *OTX2* (*Drosophila* orthodenticle homolog 2), which has a major role in the embryonic formation of the human eye. Specifically, it controls the development of photoreceptors and determines their fate (44) and also regulates the expression of rhodopsin (45), a function that could be in turn a causal determinant of RP. However, *OTX2* is not expressed in LCLs (37) and it is therefore unlikely to have contributed to the observed eQTL. In short, no obvious candidate genes could be identified at this stage but our data provide an excellent starting point for future investigations. It is worth noticing, however, that the large majority of genes present in this region (57 of the 97) produce proteins or hypothetical proteins for which no precise function is currently known.

We considered a classical model for the regulation of *PRPF31* expression by the chromosome 14 eQTL, also referred to as the *RP11*-distant regulator. Within this model, we hypothesized that this modifier could be variably expressed, depending on polymorphic variants in its sequence, and in turn modulate *PRPF31* expression. As a consequence, we expected both *PRPF31* alleles to be expressed in a 50:50 ratio (Fig. 4A and B), independently from the total *PRPF31* expression. This was indeed the case, as allelic measurements of *PRPF31* expression in all 65 CEPH LCLs that were suitable for such investigations revealed that, although they expressed very variable amounts of total *PRPF31* mRNA (from ~0.6 to 2.4 a.u.), they each expressed both alleles with equal strength (Fig. 5).

Previous studies located another modifier onto chromosome 19, within or near the *PRPF31* gene. Indeed, reduced penetrance of *PRPF31* mutations was previously demonstrated by sib pair analysis in families with asymptomatic and affected patients to be significantly linked to alleles present within the *RP11* locus itself or associated to a closely linked locus (*RP11*-associated isoalleles) (12,13). Specifically, in a meta-analytical study of 26 sibling pairs from four different families with *RP11*-linked RP, McGee *et al*. found that 10 sib pairs with the same *PRPF31* mutation and the same clinical status (either affected or asymptomatic) shared the same wild-type allele inherited from the parent who did not carry the mutation (12). Conversely, 13 of 16 sib pairs with divergent phenotypes, despite being carriers of the same mutation, inherited different haplotypes from the healthy non-carrier parent. Since penetrance of *PRPF31* mutations is determined by *PRPF31* mRNA expression, by measuring *PRPF31* mRNA levels in CEPH cell lines we expected to find a signal on chromosome 19, near the *RP11* locus. Yet, no significant LOD score was observed in the region where these *RP11*-associated isoalleles were mapped. Suggestive linkage peaks were, however, found on chromosome 19, but as far as ~50–60 cM away from the *PRPF31* sequence (Fig. 3). Failure to detect this modifier in our linkage analysis could be due to sampling and size effects. Specifically, we arbitrarily selected healthy CEPH families whereas other studies have selectively targeted *RP11* families that contained asymptomatic patients, and this may have introduced an ascertainment bias. There are no reliable estimates on the frequency of these *RP11*-associated isoalleles in the general population. Hence, it is impossible to estimate the likelihood of identifying them in the studied CEPH families. However, it is an intriguing possibility that the seven highest CEPH expressors (Fig. 2), showing very distinctive expression profiles, may indeed be carriers of *RP11*-associated isoalleles.

We next wanted to specifically investigate *RP11*-associated isoalleles, and we focused our research on LCLs derived from definite carriers of such isoalleles, ascertained from previously described *RP11* families. Similar to *PRPF31*-linked RP, the penetrance of the hereditary elliptocytosis (30) and the erythropoietic protoporphyria (31) is modulated by the expression of the allele *in trans* to the mutant allele. In the latter two diseases, the amount of wild-type mRNA is regulated by a *cis*-acting polymorphism. Based on this model, we investigated whether the *RP11*-associated isoalleles would also act as *in cis* elements. This model, depicted in Fig. 4C, could explain the asymptomatic status of carriers of *RP11*-associated isoalleles since wild-type *PRPF31* alleles would be overexpressed when the isoalleles are present. Under the assumption of a *cis*-mediated effect, the two alleles of *PRFP31* would not be expressed in a 50:50 ratio. To test this, and circumvent problems associated with nonsense-mediated mRNA decay, we performed total and allele-specific pre-mRNA quantifications in cell lines from five affected and from four asymptomatic patients. Our results showed that *PRPF31* transcripts derived from both alleles were present approximately in a 50:50 ratio, regardless of the total *PRPF31* expression, indeed higher in asymptomatics, and of the patients' clinical status (Fig. 6). The same held true for *PRPF31* pre-mRNA from LCLs from an individual (AG261) who does not carry any *PRPF31* mutation and is an obligate heterozygous carrier of an *RP11*-associated isoallele (Fig. 6). In addition, total mRNA from LCLs from this individual was elevated

with respect to other normal controls, making AG261 the fifth expressor if compared with the original CEPH set. Our data thus suggest that *RP11*-associated isoalleles do modulate *PRPF31* expression by acting on both alleles, with equal strength (Fig. 4D), unlike the classical models of hereditary elliptocytosis and erythropoietic protoporphyria (30,31). This isoalleles-based regulation would most likely imply a diffusible compound, a protein or a regulatory RNA, exercising its activity at a very early stage of RNA synthesis, since its effects are detectable at the *PRPF31* pre-mRNA level. Specifically, it is possible that this diffusible modulator could enhance *PRPF31* transcription or, conversely, be a less efficient expression downregulator. This latter scenario would indeed be in perfect agreement with a model of putative *PRPF31* downregulation by sense-antisense pairing between two neighboring transcripts, *TFPT* and *PRPF31*, that has been suggested by Anandalakshmi *et al*. (46). Furthermore, such modulation is independent from the presence of *PRPF31* mutations, as demonstrated by the analysis of AG261. However, overexpression of mutant *PRPF31* mRNA in both patients and asymptomatic carriers has virtually no effect, as the large majority of *PRPF31* mutations are degraded by nonsense-mediated mRNA decay (29). The net overall functional result of the presence of *RP11*-associated isoalleles in carriers of mutations is therefore the increase of mature *PRPF31* transcripts from the wild-ype allele only, conferring to asymptomatic patients resistance to the disease (21,28).

In light of all our results, we propose that *PRPF31* expression is managed according to a co-regulatory model, based on the effect of at least two modulators whose mode of action and relative impact on expression regulation remain to be determined. Within this model, the product of a genetic element located in the 14q21–23 region is proposed to regulate the transcription of both *PRPF31* alleles, an event that can probably be defined as the most common situation, at least according to our arbitrary sampling of the general population. Furthermore, independently from the presence of modulators from chromosome 14, *RP11*-associated isoalleles could also play a role in increasing the amounts of pre-mRNA from both *PRPF31* copies during the transcription process (Fig. 4D). The study by McGee *et al*. (12) showed that there were 3 out of 26 sibling pairs who had different phenotypes and inherited the same *in trans* haplotypes, indicating that other modifiers, such as the *RP11*-distant regulator, could indeed be determinants of penetrance, independently from *RP11*-associated isoalleles. This notion is also supported by observations that carriers of the same *PRPF31* mutation who share the same clinical classification may display variations in severity of the disease (8,11,23–25).

Penetrance is frequently determined by either the influence of alleles *in trans* and in close proximity with respect to the mutation [hereditary elliptocytosis (30), erythropoietic protoporphyria (31) and breast and ovarian cancer (47)] or by one or several modifier genes at loci unlinked to the gene causing the disease [autosomal recessive deafness (48) and cystic fibrosis (49)]. *PRPF31*-linked RP appears to be a disorder for which incomplete penetrance of dominant mutations could be tightly influenced by a combinatorial effect of very specific regulators, nearby as well as distant.

MATERIALS AND METHODS

Cell lines and cell culture

Epstein-Barr virus transformed LCLs from 200 members of 15 CEPH families (50), namely families #102, #884, #1328, #1331, #1332, #1333, #1334, #1340, #1341, #1345, #1346, #1347, #1362, #1408 and #13292 were purchased from the Coriell Cell repository (http://ccr.coriell.org). Cells were grown at a density of $0.4-1 \times 10^6$ cells/ml in RPMI 1640 with Glutamax I medium supplemented with 15% of foetal bovine serum and 1% penicillin/streptomycin mix. LCLs from three families segregating various *PRPF31* mutations are reported in Table 2. LCLs derived from affected patients AG293, AG305, AG307, 13189 and 14284 and asymptomatic patients 12943, 13191 and 13991 from three families each with different mutations in *PRPF31* were obtained and cultured as described previously (21,29). Asymptomatic individual 13991 (III-7) is the father of 14284 (IV-8), who is affected [family #2474, Fig. 1 in McGee *et al*. (12)]. In family #1562, affected individual 13189 (IV-30) is the daughter of 13191 (III-12), who is asymptomatic as is her sister, 12943 (III-15) [Fig. 1 in McGee *et al*. (12)]. LCLs representative of asymptomatic carriers used for the experiment depicted in Fig. 2 were from individuals 12943, 13191, and 13991. Asymptomatic patient cell lines (AG316, AG340) and control cell line AG261 from a family (#AD5) segregating the c.1115_1125del mutation [Fig. 1 in Moore *et al*. (8), Evans *et al*. (41) and Vithana *et al*. (28)] were purchased from the European Collection of Cell Cultures. AG293 is the affected child of AG316, who is an asymptomatic carrier of the c.1115_1125del mutation. Another branch of the family is represented by AG340 (asymptomatic) and AG261 (non-carrier) (28). Classification of individuals in specific clinical categories (affected patients, asymptomatic carriers of mutations and controls) was the same as the one originally made by the authors who first examined these individuals and reported their diagnostic assessment (6–8). These clinical evaluations were made at a time when the causative gene responsible for the disease was still unknown and were therefore not biased by the presence or absence of specific *PRPF31* genotypes.

RNA extraction and retrotranscription

Total and nuclear RNA from 10 million cultured cells was extracted and retrotranscribed as previously described (29). Briefly, all mRNA analyses were performed on cDNA synthesized from total RNA using an oligo-dT, whereas products of retrotranscription of nuclear RNA, treated with RNAse-free DNase I (Roche), were retrotranscribed with random hexamers and used for pre-mRNA analyses. To confirm that genomic DNA was absent from cDNA preparations, control reactions lacking the reverse transcriptase enzyme (−RT) were always performed. These preparations were run in parallel to and used as negative controls for every PCR and real-time PCR reaction.

PRPF31 mRNA quantification by real-time PCR

Quality of total RNA was assessed through 260/280 nm and 260/240 nm absorbance ratios and by calculating the 28S/18S ratio using the HDA-GT12 Multi-Channel Genetic Analyzer

(eGene, Inc.). Only RNA with a 28S/18S ratio above 1.4 was retained for further investigations. Expression of *PRPF31* mRNA (Genbank accession no. NM_015629.2) was measured by quantitative real-time PCR using an ABI Prism 7900 HT Real-time PCR system (Applied Biosystems). Reactions were performed in a final volume of 20 μl containing TaqMan Universal PCR Master Mix (Applied Biosystems) and ~80 ng of retrotranscribed mRNA. Primers and probe were designed to specifically amplify cDNA derived from *PRPF31* mRNA (Supplementary Table) and a pre-developed TaqMan Endogenous Control (VIC/MGB probe) (Applied Biosystems) was used to amplify *GAPDH* cDNA derived from mRNA (Genbank accession no. NM_002046.3). Reactions were set-up in 384-well plates by a Freedom Evo 200 robot (Tecan). Each plate contained the two premixes, able each to amplify either *PRPF31* or *GAPDH* cDNA, and one copy of each sample. Furthermore, replicated twice on each plate, two serial dilutions of samples composed by a mix of cDNA from CEPH cell lines were added to measure the efficiency of the two PCR set-ups. This checked the accuracy of pipetting (i.e. same Ct values for the two replicates of standards in a same plate) and the reproductibility and robustness of each PCR set-up (i.e. same Ct values for standards and same PCR efficiencies on different plates). Overall, *PRPF31* and *GAPDH* expression from each sample was measured five times on five different plates for a total of ~2000 independent real-time PCR amplifications. Biological reproducibility was tested by randomly selecting 82 cell lines from the analyzed CEPH individuals and by growing each one of them again as two independent cultures. RNA for each of these 164 samples was re-extracted and re-assessed for both *PRPF31* and *GAPDH* expression by following the same procedures. The same replications were applied to all LCLs derived from families segregating *PRPF31* mutations.

All plates were analyzed using the same threshold to ensure uniformity of the analyses. Ct values for all samples were confirmed to be within the Ct range of the serial dilutions. Since efficiencies of the two PCR were approximately equal, *PRPF31* mRNA expression was normalized with respect to *GAPDH*, using the ΔΔCt method, for each sample and was expressed as a relative quantity, with respect to the median *PRPF31* mRNA expressor. Outliers were defined using the Chauvenet's criterion and removed.

Genotypes, heritability and linkage analyses

The CEPH cell panel of SNP markers was obtained from the CEPH genotype database V10.0 (http://www.cephb.fr/cephdb/). A genetic map consisting of 2713 SNP markers genotyped for the 15 CEPH families and with known NCBI position (dbSNP build 124) was established. The SNP centiMorgan map positions were interpolated from the Rutgers Linkage-Physical Map (http://compgen.rutgers.edu/maps/) (51). The marker set was screened with PedCheck (52). No Mendelian inconsistencies or unlikely genotypes were detected.

To estimate the part of the phenotypic variance that is due to total additive genetic variance, heritability was calculated using the polygenic function of the SOLAR software (34). To identify eQTLs involved in the regulation of *PRPF31* expression, genome-wide multipoint linkage analyses were performed using two different methods implemented within the MERLIN software (53): a variance components algorithm (MERLIN-VC) and a regression-based algorithm (MERLIN-REGRESS) (35).

The distribution of *PRPF31* expression was right-skewed (skewness = 1.2) and, as computed by the SOLAR software, had a residual kurtosis of 1.9, which denotes a significant shift from the normal distribution (data not shown). Such a departure from the Gaussian curve has been shown to cause inflated type I errors during linkage analyses (54,55). Therefore, we applied a $\log_{10}$ transformation to the raw phenotypes, and obtained a set of values showing a final residual kurtosis of -0.2 and a skewness of -0.2, more typical of a normal distribution, as confirmed by a Kolmogorov–Smirnov test (data not shown). Therefore, the $\log_{10}$ transformed values of *PRPF31* expressions were used as a phenotype for linkage analyses.

LOD score significance thresholds were set as suggested by Lander and Kruglyak (56), i.e. a LOD score ≥ 3.3 ($P = 4.9 \times 10^{-5}$) was defined to correspond to a genome-wide 'significant linkage' and a LOD score ≥ 1.9 ($P = 1.7 \times 10^{-3}$) to indicate 'suggestive linkage'.

mRNA and pre-mRNA allelic quantification by SNP analysis

Regions containing SNPs that create restriction sites in the mRNA and genomic DNA were amplified using specific primers to generate PCR products of equal sizes on the genomic DNA and cDNA (Supplementary Table). Heterozygous carriers were identified by analyzing their genomic DNA using restriction fragment length polymorphism (RFLP). PCR were performed in a total volume of 25 μl, containing 1× Expand High Fidelity PCR System buffer with $MgCl_2$, 100 μM dNTP mix, 200 nM of each primer and 1 U of Expand High Fidelity PCR System (Roche). PCR cycling conditions were 94°C for 2 min followed by 40 cycles of 15 s at 94°C, 30 s at 60°C and 1 min at 68°C and by a final elongation step of 10 min at 68°C. RFLP analysis was done in a total volume of 16.25 μl by incubating 10 μl of PCR products with 1× appropriate NEB Buffer, 1× BSA when required, and 10 U of restriction enzyme at 37°C for 2 h.

To quantify pre-mRNA or mRNA allelic expression, for each SNP analyzed, it was first necessary to create quantification standards. Two different genomic DNA templates from CEPH cell lines that were homozygous for either the major or minor SNP alleles were mixed to obtain quantification standards in the following ratios: 70/30, 65/35, 60/40, 55/45, 50/50, 45/55, 40/60, 35/65 and 30/70, at concentrations that could generate as much PCR products as cDNA (to avoid differences in signal strength). cDNA from samples, quantification standards and controls (homozygous and heterozygous genomic DNA, cDNA derived from homozygotes for either SNP alleles) were amplified on the same plate. PCR products were then digested by the relevant restriction enzyme. RFLP analysis was performed by capillary electrophoresis, i.e. by running digested PCR products on the HDA-GT12 Multi-Channel Genetic Analyzer (eGene, Inc.) and quantification was achieved using the Biocalculator software (eGene, Inc.).

PCR products from homozygous variations of genomic DNA and cDNA were used as controls for the digestion.

The logarithm of the values of the quantification standards was plotted against ratios measured by the Biocalculator software. Linear regression analysis showed that the R^2 was at least 0.99. Ratios calculated by the software were reported on this regression curve to determine the allelic expression of each sample. Quantification of heterozygous DNA was confirmed to be in a ~50:50 ratio. Pre-mRNA allelic quantifications were performed by RFLP for family #1562 (6) (using SNP rs1058572) and for the control individual AG261 (using the intronic SNP rs11673377).

Pre-mRNA allelic quantification by real-time PCR

Oligonucleotides specific for either the mutant or the non-mutant cDNA derived from pre-mRNA from patients carrying *PRPF31* mutations were used for allele-specific real-time PCR (Supplementary Table). PCR products were subcloned into pcDNA3.1/V5-His TOPO TA vectors (Invitrogen). The resulting plasmids were sequenced using BigDye Cycle Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) and used to generate quantification standards, by combining non-mutant and mutant sequences in the following ratios: 90/10, 80/20, 70/30, 60/40, 50/50, 40/60, 30/70, 20/80, 10/90. Reactions were performed using an ABI Prism 7500 Sequence Detector (Applied Biosystems) in a final volume of 20 μl containing Power Sybr Green Master Mix (Applied Biosystems) and ~600 ng of retrotranscribed nuclear RNA. PCR efficiencies were determined by amplifying standard dilutions of patient cDNA and derived plasmids. Efficiencies were confirmed to be identical for cDNA and plasmids for the same allele (mutant or non-mutant). To quantify the expression of mutant and non-mutant alleles, cDNA from LCLs from patients, the heterozygous genomic DNA carrying the same *PRPF31* mutation, and plasmid quantification standards were run in triplicate on the same plate for each family sharing the same mutation.

Expression analysis used the relative standard curve method, with wild-type pre-mRNA as reference and mutant pre-mRNA as target. Standards and genomic DNA from LCLs from patients were also quantified to validate the accuracy of the allele-specific real-time PCR. Pre-mRNA allelic quantifications were performed by allele-specific real-time PCR for carriers of *PRPF31* mutations in families #AD5 (8) and #2474 (7).

Pre-mRNA quantification by real-time PCR

To quantify the expression of *PRPF31* pre-mRNA from nuclear RNA, PCR primers were designed to specifically amplify cDNA derived only from pre-mRNA (Supplementary Table). Reactions were performed using an ABI Prism 7500 Sequence Detector in a final volume of 20 μl containing either the Power Sybr Green Master Mix or the TaqMan Universal PCR Master Mix (Applied Biosystems) and ~600 ng of retrotranscribed nuclear RNA. The pre-developed TaqMan Endogenous Control (VIC/MGB probe) that amplifies *GAPDH* cDNA derived from mRNA was used to normalize *PRPF31* pre-mRNA expression during quantification. Standard curves were generated, for both *PRPF31* pre-mRNA and *GAPDH* mRNA, using a 5-fold serial dilution over a 375-fold range of a pool of cDNA samples to be analyzed. Similar efficiencies for the two PCR reactions allowed the relative quantification of the amount of *PRPF31* pre-mRNA in cell lines derived from patients by the ΔΔCt method.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* Online.

FUNDING

This work was supported by the Swiss National Science Foundation grant 310000-109620 and by funds from the SVS (Sciences, Vie, Société) Program.

ACKNOWLEDGEMENTS

We would like to thank Dr Corinne Loeuillet for providing the majority of CEPH cell lines, Diana Marek for help on data normalization, and Dr Nicholas Wade for careful revision of this manuscript. We are particularly indebted to Dr Eliot L. Berson from Harvard Medical School for providing important clinical information and very constructive comments.

Conflict of Interest statement. None declared.

REFERENCES

1. Hartong, D.T., Berson, E.L. and Dryja, T.P. (2006) Retinitis pigmentosa. *Lancet*, **368**, 1795–1809.
2. Li, Z.Y., Possin, D.E. and Milam, A.H. (1995) Histopathology of bone spicule pigmentation in retinitis pigmentosa. *Ophthalmology*, **102**, 805–816.
3. Berson, E.L. (1993) Retinitis pigmentosa. The Friedenwald Lecture. *Invest. Ophthalmol. Vis. Sci.*, **34**, 1659–1676.
4. Rivolta, C., Sharon, D., DeAngelis, M.M. and Dryja, T.P. (2002) Retinitis pigmentosa and allied diseases: numerous diseases, genes, and inheritance patterns. *Hum. Mol. Genet.*, **11**, 1219–1227.
5. Weidenhammer, E.M., Singh, M., Ruiz-Noriega, M. and Woolford, J.L Jr. (1996) The PRP31 gene encodes a novel protein required for pre-mRNA splicing in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **24**, 1164–1170.
6. Berson, E.L., Gouras, P., Gunkel, R.D. and Myrianthopoulos, N.C. (1969) Dominant retinitis pigmentosa with reduced penetrance. *Arch. Ophthalmol.*, **81**, 226–234.
7. Berson, E.L. and Simonoff, E.A. (1979) Dominant retinitis pigmentosa with reduced penetrance. Further studies of the electroretinogram. *Arch. Ophthalmol.*, **97**, 1286–1291.
8. Moore, A.T., Fitzke, F., Jay, M., Arden, G.B., Inglehearn, C.F., Keen, T.J., Bhattacharya, S.S. and Bird, A.C. (1993) Autosomal dominant retinitis pigmentosa with apparent incomplete penetrance: a clinical, electrophysiological, psychophysical, and molecular genetic study. *Br. J. Ophthalmol.*, **77**, 473–479.
9. Xu, S., Nakazawa, M., Tamai, M. and Gal, A. (1995) Autosomal dominant retinitis pigmentosa locus on chromosome 19q in a Japanese family. *J. Med. Genet.*, **32**, 915–916.
10. Al Maghtheh, M., Vithana, E., Tarttelin, E., Jay, M., Evans, K., Moore, T., Bhattacharya, S. and Inglehearn, C.F. (1996) Evidence for a major retinitis pigmentosa locus on 19q13.4 (RP11) and association with a unique bimodal expressivity phenotype. *Am. J. Hum. Genet.*, **59**, 864–871.
11. Nakazawa, M., Xu, S., Gal, A., Wada, Y. and Tamai, M. (1996) Variable expressivity in a Japanese family with autosomal dominant retinitis

pigmentosa closely linked to chromosome 19q. *Arch. Ophthalmol.*, **114**, 318–322.
12. McGee, T.L., Devoto, M., Ott, J., Berson, E.L. and Dryja, T.P. (1997) 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.*, **61**, 1059–1066.
13. Vithana, E.N., Abu-Safieh, L., Allen, M.J., Carey, A., Papaioannou, M., Chakarova, C., Al Maghtheh, M., Ebenezer, N.D., Willis, C., Moore, A.T. *et al.* (2001) A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell*, **8**, 375–381.
14. Martinez-Gimeno, M., Gamundi, M.J., Hernan, I., Maseras, M., Milla, E., Ayuso, C., Garcia-Sandoval, B., Beneyto, M., Vilela, C., Baiget, M. *et al.* (2003) Mutations in the pre-mRNA splicing-factor genes PRPF3, PRPF8 and PRPF31 in Spanish families with autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.*, **44**, 2171–2177.
15. Wang, L., Ribaudo, M., Zhao, K., Yu, N., Chen, Q., Sun, Q. and Wang, Q. (2003) Novel deletion in the pre-mRNA splicing gene PRPF31 causes autosomal dominant retinitis pigmentosa in a large Chinese family. *Am. J. Med. Genet. A.*, **121**, 235–239.
16. Xia, K., Zheng, D., Pan, Q., Liu, Z., Xi, X., Hu, Z., Deng, H., Liu, X., Jiang, D. and Xia, J. (2004) A novel PRPF31 splice-site mutation in a Chinese family with autosomal dominant retinitis pigmentosa. *Mol. Vis.*, **10**, 361–365.
17. Lu, S.S., Zhao, C., Cui, Y., Li, N.D., Zhang, X.M. and Zhao, K.X. (2005) [Novel splice-site mutation in the pre-mRNA splicing gene PRPF31 in a Chinese family with autosomal dominant retinitis pigmentosa]. *Zhonghua Yan Ke Za Zhi*, **41**, 305–311.
18. Sato, H., Wada, Y., Itabashi, T., Nakamura, M., Kawamura, M. and Tamai, M. (2005) Mutations in the pre-mRNA splicing gene, PRPF31, in Japanese families with autosomal dominant retinitis pigmentosa. *Am. J. Ophthalmol.*, **140**, 537–540.
19. Abu-Safieh, L., Vithana, E.N., Mantel, I., Holder, G.E., Pelosini, L., Bird, A.C. and Bhattacharya, S.S. (2006) A large deletion in the adRP gene PRPF31: evidence that haploinsufficiency is the cause of disease. *Mol. Vis.*, **12**, 384–388.
20. Chakarova, C.F., Cherninkova, S., Tournev, I., Waseem, N., Kaneva, R., Jordanova, A., Veraitch, B.K., Gill, B., Colclough, T., Nakova, A. *et al.* (2006) Molecular genetics of retinitis pigmentosa in two Romani (Gypsy) families. *Mol. Vis.*, **12**, 909–914.
21. Rivolta, C., McGee, T.L., Rio Frio, T., Jensen, R.V., Berson, E.L. and Dryja, T.P. (2006) Variation in retinitis pigmentosa-11 (PRPF31 or RP11) gene expression between symptomatic and asymptomatic patients with dominant RP11 mutations. *Hum. Mutat.*, **27**, 644–653.
22. Sullivan, L.S., Bowne, S.J., Seaman, C.R., Blanton, S.H., Lewis, R.A., Heckenlively, J.R., Birch, D.G., Hughbanks-Wheaton, D. and Daiger, S.P. (2006) Genomic rearrangements of the PRPF31 gene account for 2.5% of autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.*, **47**, 4579–4588.
23. Ghazawy, S., Springell, K., Gauba, V., McKibbin, M.A. and Inglehearn, C.F. (2007) Dominant retinitis pigmentosa phenotype associated with a new mutation in the splicing factor PRPF31. *Br. J. Ophthalmol.*, **91**, 1411–1413.
24. Taira, K., Nakazawa, M. and Sato, M. (2007) Mutation c. 1142 del G in the PRPF31 gene in a family with autosomal dominant retinitis pigmentosa (RP11) and its implications. *Jpn J. Ophthalmol.*, **51**, 45–48.
25. Waseem, N.H., Vaclavik, V., Webster, A., Jenkins, S.A., Bird, A.C. and Bhattacharya, S.S. (2007) Mutations in the gene coding for the pre-mRNA splicing factor, PRPF31, in patients with autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.*, **48**, 1330–1334.
26. Liu, J.Y., Dai, X., Sheng, J., Cui, X., Wang, X., Jiang, X., Tu, X., Tang, Z., Bai, Y., Liu, M. *et al.* (2008) Identification and functional characterization of a novel splicing mutation in RP gene PRPF31. *Biochem. Biophys. Res. Comm*, **367**, 420–426.
27. Al Maghtheh, M., Inglehearn, C.F., Keen, T.J., Evans, K., Moore, A.T., Jay, M., Bird, A.C. and Bhattacharya, S.S. (1994) Identification of a sixth locus for autosomal dominant retinitis pigmentosa on chromosome 19. *Hum. Mol. Genet.*, **3**, 351–354.
28. Vithana, E.N., Abu-Safieh, L., Pelosini, L., Winchester, E., Hornan, D., Bird, A.C., Hunt, D.M., Bustin, S.A. and Bhattacharya, S.S. (2003) Expression of PRPF31 mRNA in patients with autosomal dominant retinitis pigmentosa: a molecular clue for incomplete penetrance? *Invest. Ophthalmol. Vis. Sci.*, **44**, 4204–4209.
29. Rio Frio, T., Wade, N.M., Ransijn, A., Berson, E.L., Beckmann, J.S. and Rivolta, C. (2008) Premature termination codons in PRPF31 cause retinitis pigmentosa via haploinsufficiency due to nonsense-mediated mRNA decay. *J. Clin. Invest.*, **118**, 1519–1531.
30. Wilmotte, R., Marechal, J., Morle, L., Baklouti, F., Philippe, N., Kastally, R., Kotula, L., Delaunay, J. and Alloisio, N. (1993) Low expression allele alpha LELY of red cell spectrin is associated with mutations in exon 40 (alpha V/41 polymorphism) and intron 45 and with partial skipping of exon 46. *J. Clin. Invest.*, **91**, 2091–2096.
31. Gouya, L., Puy, H., Robreau, A.M., Bourgeois, M., Lamoril, J., Da Silva, V., Grandchamp, B. and Deybach, J.C. (2002) The penetrance of dominant erythropoietic protoporphyria is modulated by expression of wildtype FECH. *Nat. Genet.*, **30**, 27–28.
32. Beckmann, J.S., Estivill, X. and Antonarakis, S.E. (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat. Rev. Genet.*, **8**, 639–646.
33. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. and Lee, C. (2004) Detection of large-scale variation in the human genome. *Nat. Genet.*, **36**, 949–951.
34. Almasy, L. and Blangero, J. (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.*, **62**, 1198–1211.
35. Sham, P.C., Purcell, S., Cherny, S.S. and Abecasis, G.R. (2002) Powerful regression-based quantitative-trait linkage analysis of general pedigrees. *Am. J. Hum. Genet.*, **71**, 238–253.
36. Lander, E.S. and Botstein, D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.
37. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C. *et al.* (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, **315**, 848–853.
38. Bernier, G., Panitz, F., Zhou, X., Hollemann, T., Gruss, P. and Pieler, T. (2000) Expanded retina territory by midbrain transformation upon overexpression of Six6 (Optx2) in Xenopus embryos. *Mech. Dev.*, **93**, 59–69.
39. Behesti, H., Holt, J.K. and Sowden, J.C. (2006) The level of BMP4 signaling is critical for the regulation of distinct T-box gene expression domains and growth along the dorso-ventral axis of the optic cup. *BMC Dev. Biol.*, **6**, 62.
40. Fossat, N., Le Greneur, C., Beby, F., Vincent, S., Godement, P., Chatelain, G. and Lamonerie, T. (2007) A new GFP-tagged line reveals unexpected Otx2 protein localization in retinal photoreceptors. *BMC Dev. Biol.*, **7**, 122.
41. Evans, K., al-Maghtheh, M., Fitzke, F.W., Moore, A.T., Jay, M., Inglehearn, C.F., Arden, G.B. and Bird, A.C. (1995) Bimodal expressivity in dominant retinitis pigmentosa genetically linked to chromosome 19q. *Br. J. Ophthalmol.*, **79**, 841–846.
42. Malhotra, A., Cromer, K., Leppert, M.F. and Hasstedt, S.J. (2005) The power to detect genetic linkage for quantitative traits in the Utah CEPH pedigrees. *J. Hum. Genet.*, **50**, 69–75.
43. Loeuillet, C., Deutsch, S., Ciuffi, A., Robyr, D., Taffe, P., Munoz, M., Beckmann, J.S., Antonarakis, S.E. and Telenti, A. (2008) In vitro whole-genome analysis identifies a susceptibility locus for HIV-1. *PLoS Biol.*, **6**, e32.
44. Nishida, A., Furukawa, A., Koike, C., Tano, Y., Aizawa, S., Matsuo, I. and Furukawa, T. (2003) Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development. *Nat. Neurosci.*, **6**, 1255–1263.
45. Koike, C., Nishida, A., Ueno, S., Saito, H., Sanuki, R., Sato, S., Furukawa, A., Aizawa, S., Matsuo, I., Suzuki, N. *et al.* (2007) Functional roles of Otx2 transcription factor in postnatal mouse retinal development. *Mol. Cell. Biol.*, **27**, 8318–8329.
46. Anandalakshmi, V., Divya, V., Victor, Y., Bhattacharya, S.S., Vithana, E.N. and Aung, T. (2007) Investigation of sense-antisense pairing of PRPF31 and TFPT as a putative mechanism for reduced levels of wild type PRPF31 transcripts. *Asian J. Ophthalmol.*, **9** (Suppl. 1), 226.
47. Ginolhac, S.M., Gad, S., Corbex, M., Bressac-De-Paillerets, B., Chompret, A., Bignon, Y.J., Peyrat, J.P., Fournier, J., Lasset, C., Giraud, S. *et al.* (2003) BRCA1 wild-type allele modifies risk of ovarian cancer in carriers of BRCA1 germ-line mutations. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 90–95.
48. Riazuddin, S., Castelein, C.M., Ahmed, Z.M., Lalwani, A.K., Mastroianni, M.A., Naz, S., Smith, T.N., Liburd, N.A., Friedman, T.B.,

Griffith, A.J. *et al.* (2000) Dominant modifier DFNM1 suppresses recessive deafness DFNB26. *Nat. Genet.*, **26**, 431–434.
49. Gropman, A.L. and Adams, D.R. (2007) Atypical patterns of inheritance. *Semin. Pediatr. Neurol.*, **14**, 34–45.
50. Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J.M. and White, R. (1990) Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics*, **6**, 575–577.
51. Kong, X., Murphy, K., Raj, T., He, C., White, P.S. and Matise, T.C. (2004) A combined linkage-physical map of the human genome. *Am. J. Hum. Genet.*, **75**, 1143–1148.
52. O'Connell, J.R. and Weeks, D.E. (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.*, **63**, 259–266.
53. Abecasis, G.R., Cherny, S.S., Cookson, W.O. and Cardon, L.R. (2002) Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.*, **30**, 97–101.
54. Allison, D.B., Neale, M.C., Zannolli, R., Schork, N.J., Amos, C.I. and Blangero, J. (1999) Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *Am. J. Hum. Genet.*, **65**, 531–544.
55. Feingold, E. (2002) Regression-based quantitative-trait-locus mapping in the 21st century. *Am. J. Hum. Genet.*, **71**, 217–222.
56. Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.*, **11**, 241–247.