

Mutational screening of splicing factor genes in cases with autosomal dominant retinitis pigmentosa

Paola Benaglio,¹ Patricia Fernandez San Jose,^{2,3} Almudena Avila-Fernandez,^{2,3} Giulia Ascari,¹ Shyana Harper,⁴ Gaël Manes,⁵ Carmen Ayuso,^{2,3} Christian Hamel,⁵ Eliot L. Berson,⁴ Carlo Rivolta¹

¹Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; ²Department of Genetics, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital (IIS-FJD, UAM), Madrid, Spain; ³Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), ISCIII, Madrid, Spain; ⁴The Berman-Gund Laboratory for the Study of Retinal Degenerations, Harvard Medical School, Massachusetts Eye and Ear, Boston, MA; ⁵INSERM U1051, Institut des Neurosciences de Montpellier, Hôpital Saint Eloi, Montpellier, France

Purpose: Mutations in genes encoding proteins from the tri-snRNP complex of the spliceosome account for more than 12% of cases of autosomal dominant retinitis pigmentosa (adRP). Although the exact mechanism by which splicing factor defects trigger photoreceptor death is not completely clear, their role in retinitis pigmentosa has been demonstrated by several genetic and functional studies. To test for possible novel associations between splicing factors and adRP, we screened four tri-snRNP splicing factor genes (*EFTUD2*, *PRPF4*, *NHP2LI*, and *AAR2*) as candidate disease genes.

Methods: We screened up to 303 patients with adRP from Europe and North America who did not carry known RP mutations. Exon-PCR and Sanger methods were used to sequence the *NHP2LI* and *AAR2* genes, while the sequences of *EFTUD2* and *PRPF4* were obtained by using long-range PCRs spanning coding and non-coding regions followed by next-generation sequencing.

Results: We detected novel missense changes in individual patients in the sequence of the genes *PRPF4* and *EFTUD2*, but the role of these changes in relationship to disease could not be verified. In one other patient we identified a novel nucleotide substitution in the 5' untranslated region (UTR) of *NHP2LI*, which did not segregate with the disease in the family.

Conclusions: The absence of clearly pathogenic mutations in the candidate genes screened in our cohort suggests that *EFTUD2*, *PRPF4*, *NHP2LI*, and *AAR2* are either not involved in adRP or are associated with the disease in rare instances, at least as observed in this study in patients of European and North American origin.

The most common form of hereditary retinal blindness is retinitis pigmentosa (RP), which affects about 1 in 4,000 people worldwide. The disease typically begins with patients experiencing night blindness, due to the early involvement of rod photoreceptors, and progresses with a decrease in the visual field and loss of central vision, due to the degeneration of cone photoreceptors [1]. Patients affected with RP display clinical heterogeneity regarding age of onset, degree of severity, rate of progression, and other secondary manifestations. These differences are partly explained by the different genes and mutations that cause RP. To date, more than 60 genes have been associated with non-syndromic RP, with about 3,000 mutations reported in total; however, a substantial fraction of cases are negative for mutations in known disease genes [2]. The inheritance mode is classically monogenic: dominant (about 30–40%), recessive (about 50–60%),

X-linked (5–15%), and a smaller fraction of non-Mendelian or complex inheritance [1].

The functions of RP genes can be diverse: some genes are specific for retinal function such as phototransduction and retinal metabolism, while others have a more general function in cell development and maintenance [3]. A particular category that exemplifies the complexity of the molecular genetics of RP is represented by a few highly conserved and ubiquitously expressed pre-mRNA splicing factors.

Splicing consists of consecutive reactions occurring in the nucleus and leading to the removal of introns from pre-mRNA to form mature mRNA. A macromolecular complex, referred to as the spliceosome, ensures the fidelity and the correct timing of these reactions. The core components of the spliceosome are five small nuclear ribonucleoproteins (snRNP), U1, U2, U4, U5, and U6 [4], that assemble on the pre-mRNA in an ordered, stepwise manner. U1 snRNP first recognizes the 5' splice site, and U2 binds to the branch point; then the U4/U6.U5 tri-snRNP complex is recruited, and finally U1 and U4 are released, leading to catalytic activation [5].

Correspondence to: Carlo Rivolta, Department of Medical Genetics, University of Lausanne Rue du Bugnon 27, 1005 Lausanne, Switzerland; Phone: +41-21-6925451; FAX: +41-21-6925455; email: carlo.rivolta@unil.ch

To date, six splicing factors genes have been found to be mutated in patients with adRP: *PRPF8* (RP13; ID: 10594, OMIM 607300) [6], *PRPF31* (RP11; ID: 26121, OMIM 606419) [7], *PRPF3* (RP18; ID: 9129, OMIM 607301) [8], *PAP-1* (RP9; ID: 6100, OMIM 607331) [9], *SNRNP200* (RP33; ID: 23020, OMIM 601664) [10,11], and *PRPF6* (ID: 24148, OMIM 613979) [12]. The prevalence of mutations in adRP cases is estimated to be about 8% for *PRPF31*, 2–3% for *PRPF8*, 1–4% for *PRPF3*, 1.6% for *SNRNP200*, and to be rare for *PAP-1* and *PRPF6* [1,13,14], globally accounting for more than 12% of all adRP cases.

All these genes have a high level of protein sequence conservation up to yeast and belong to the U4/U6.U5 tri-snRNP complex. The growing evidence of a major role of these particular splicing factors suggested that other partners of the complex could be meaningful candidate genes for adRP. Indeed, these mutations have been discovered through linkage analysis and positional cloning for the first two genes discovered (*PRPF8* and *PRPF31*), followed by the sequencing of other splicing factor genes in linkage intervals or in large cohorts of patients. This latter strategy, commonly referred to as the candidate gene approach [15], has been (and still is) instrumental for discovering several new RP genes. For instance, the role of the *PRPF6* gene in adRP was found with this approach, via the sequencing of the coding sequence in a cohort of 200 American patients [12].

Following the same rationale, we screened four candidate genes from the tri-snRNP complex in up to 303 patients with adRP with unknown molecular diagnosis and previously found to be negative for mutations in the most common adRP genes or hotspots. We selected the genes *EFTUD2* (ID: 9343, OMIM 603892), *PRPF4* (ID: 9128, OMIM 607795), *NHP2L1* (ID: 4809, OMIM 601304) and *AAR2* (ID: 25980), because of their physical or functional interaction with known RP-linked splicing factors. In particular, *EFTUD2* encodes for an essential GTPase, hSnu114, homolog of *Saccharomyces cerevisiae* Snu114p, which forms a stable complex with the *SNRNP200* and *PRPF8* products (i.e., hBrr2 and PRPF8, both involved in adRP) [16]. hSnu114 regulates hBrr2 at the dissociation step of U4 from U6 and is necessary for spliceosome disassembly after splicing [17]. The *AAR2* gene encodes Aar2p, which competes with hBrr2 in the binding of the C-terminal region of PRPF8 before the maturation of the U5 snRNP, supposedly regulating its assembly [18]. The 15.5-kDa protein (Snu13p in yeast), encoded by the *NHP2L1* gene, binds to the 5'-stem-loop of U4 snRNA probably playing a role in the late phase of the spliceosome assembly [19]. Finally, the PRPF4 protein forms a complex with PRPF3 in the U4/U6 snRNP complex,

and its downregulation was found to induce photoreceptor defects in a zebrafish model, similarly to PRPF31 [20,21].

For the genetic screening we took advantage of a method that combines classical exon-PCR for the small genes and long-range PCR followed by next-generation sequencing for the large genes. The latter approach provides a cost- and time-effective alternative to the Sanger method and adapts well to routine genetic screenings in large sets of samples.

METHODS

Samples and patients: The subjects analyzed in this cohort belong to three groups of unrelated patients affected with autosomal dominant retinitis pigmentosa. One hundred and ninety-one samples were collected at the Berman-Gund Laboratory, Harvard Medical School, Massachusetts Eye and Ear and are mostly of North American origin. They were previously screened and found to be negative for exonic mutations by Sanger sequencing in 90% of all known adRP genes. One hundred and fifteen were collected in Spain (Servicio de Genética, IIS Fundación Jiménez Díaz University Hospital, Madrid) and were negative to a genotyping microarray that assessed known RP mutations [22]. Ninety-six were from France (INSERM U1051, Institut des Neurosciences de Montpellier, Hôpital Saint Eloi, Montpellier) and before this study were sequenced and found to be negative for the ten most frequently mutated genes or hotspots (*RHO* (ID: 6010, OMIM 180380), *RDS* (ID: 5961, OMIM 179605), *PRPF31* (ID: 26121, OMIM 606419), *RPI* (ID: 6101, OMIM 603937), *PRPF8* (ID: 10594, OMIM 607300), *IMPDH1* (ID: 3614, OMIM 146690), *NRL* (ID: 4901, OMIM 162080), *PRPF3* (ID: 9129, OMIM 607301), *NR2E3* (ID: 10002, OMIM 604485), and *SNRNP200* (ID: 23020, OMIM 601664) [23]. DNA was extracted from peripheral leukocytes and quantified. For technical reasons, only a subset of these samples could undergo the complete screening of the four genes but all (402 individuals) were analyzed for putative mutations in specific exons. Control DNA samples were obtained from 95 individuals with no history of retinal degeneration and 96 unrelated healthy individuals between age 34 and 92, purchased from the Coriell Institute for Medical Research. All subjects provided written, informed consent, and the study was conducted in adherence with the Declaration of Helsinki. This research was approved by the Institutional Review Boards of our respective Universities or Hospitals: University of Lausanne, Fundación Jiménez Díaz University Hospital, Institut des Neurosciences de Montpellier, Harvard Medical School, and the Massachusetts Eye and Ear.

Library preparation and next-generation sequencing: Genes *EFTUD2* and *PRPF4* were sequenced with long-range PCR

(LR-PCR) followed by next-generation sequencing (NGS), using Illumina instruments (San Diego, CA). Five and two LR-PCRs were generated to amplify the entire 51- and 20-kb regions of each gene, respectively, for a total of 71-kb targeted region. LR-PCRs were obtained individually for each sample using TaKaRa LA Taq polymerase (Takara Bio, Shiga, Japan) with GC buffer and 1 μ M of the primers reported in Appendix 1. The following cycling conditions were used: 94 °C for 1 min followed by 30 cycles at 98 °C for 5 s and 68 °C for 15 min, and final extension of 72 °C for 10 min. For each sample, the seven LR-PCRs were pooled into a single tube, after their quantity was estimated on agarose 1% gel. They were subsequently purified using DNA Clean and Concentrator columns (Zymo Research, Orange, CA). Only the DNA samples from the three cohorts that yielded seven clear PCR bands underwent NGS, resulting in 200 samples in total.

Library preparation and sample barcoding were performed as described by Adey et al. [24] using the Nextera DNA Sample Prep Kit (Epicenter, Madison, WI) and 48 barcodes adapted to Illumina platforms [25], following the manufacturer's instructions. Fourteen tagged samples were sequenced as a pool in one lane of the GAII instrument for testing purposes, after which two runs of the HiSeq instrument (one lane for each run) were used to sequence two pools of 48 and 47 barcoded samples each. After the Nextera products were integrated by Illumina, we processed 91 additional samples using the Nextera XT DNA Sample Preparation (Illumina) protocol, reagents, and barcodes, and sequenced the samples as a unique pool with one Miseq instrument run.

Analysis and variant calling of sequences from next-generation sequencing: We mapped the reads obtained from NGS to the reference sequence of the genes (GRCh37.p10 assembly) with the CLC Genomics Workbench package, v. 5.5 (CLC bio, Aarhus, Denmark). The parameters were in a way that a read could align only if it had at least a 90% identity for the 90% of its length. A more relaxed setting (80% identity over 70% of its length) was also tried. Single nucleotide variant calling and small insertion and deletion calling were achieved by imposing a minimum frequency of discordant bases of 20%, with minimum coverage of five nucleotides and an average base quality of 20 Phred. The analyses were performed as a batch of all individual samples (200), and the obtained variants were annotated with the hg19_snp137 track from the [UCSC Genome Browser](#).

Analysis of variant identified: To exclude polymorphisms, we consulted the databases dbSNP, 1000 Genomes, Exome Variant Server, Complete Genomics' 42 control individuals, and exome sequencing data from 500 individuals from the CoLaus cohort [26]. Missense changes were analyzed with

the online package **PON-P**, which integrates the results of the most common prediction software, including PolyPhen and SIFT [27]. The effect of intronic changes was evaluated with the Shannon Human Splicing Pipeline, kindly offered to us as a free trial by Cytogenomix (London, Canada) and implemented in the CLC software [28], and the NNSPLICE 0.9 algorithm [29].

Sanger sequencing and restriction analysis: The genes *NHP2L1* and *AAR2* were screened with Sanger sequencing of the coding exons. PCR reactions were obtained with the GoTaq polymerase (Promega, Madison, WI) standard protocol and 0.25 μ M of the primers reported in Appendix 2. Reactions were purified from excess primers and nucleotides with ExoSAP-IT (Affymetrix, Santa Clara, CA) and subjected to sequencing reactions using the Big Dye VI.1 Terminator Kit (Applied Biosystems, Foster City, CA) and an ABI automated DNA sequencer (Applied Biosystems). Sequences were analyzed using the CLC Genomics Workbench (CLC bio). The same procedure was applied to validate the novel changes in specific exons identified with NGS, cosegregation analysis, and screening of controls and additional patients. Primers used for these purposes are listed in Appendix 2. In some instances, controls and additional patients were tested using restriction enzymes when a particular nucleotide change abolished or created a restriction site. In particular, exon 5 of *PRPF4* was tested with *MscI*, exon 8 of *EFTUD2* with *HahI*, and exon 1 of *NHP2L1* with *MluI* (New England Biolabs, Ipswich, MA).

RESULTS

Screening of PRPF4 and EFTUD2: We obtained LR-PCR products spanning the genes *PRPF4* and *EFTUD2* for a total of 200 unrelated individuals diagnosed with adRP (Table 1). Seventy-nine patients were from North America, 71 from France, and 50 from Spain. Following multiplexed runs of NGS instruments, we analyzed the sequencing reads by alignment to the reference genomic sequences of the targeted genes. Since different instruments were used, different samples had different coverage depths. Specifically, the samples sequenced with HiSeq had higher coverage than the ones sequenced with MiSeq, due to the lower throughput and higher number of samples sequenced with the latter (Appendix 3). With the exception of a few samples, the targeted region was optimally covered for reliable variant calling. This consisted in the detection of single nucleotide variations and small insertions and deletions by the CLC Genomics algorithm. After merging the results obtained individually for each sample, we obtained a total figure of 1,195 variants identified, 591 of which are annotated variants

TABLE 1. GENES ANALYZED IN THIS STUDY AND METHODS.

Symbol	Protein	snRNP complex	# exons	Screening method	# screened patients
<i>EFTUD2</i>	Elongation factor Tu GTP binding domain containing 2- 116 kDa	U5 [13]	28	Nextera-NGS	200
<i>PRPF4</i>	PRP4 pre-mRNA processing factor 4 homolog – 60 kDa	U4/U6 [17]	14	Nextera-NGS	200
<i>NHP2L1</i>	NHP2 non-histone chromosome protein 2-like 1- 15.5 kDa	U4/U6.U5 [16]	4	Sanger	303
<i>AAR2</i>	AAR2 splicing factor homolog	U5 [15]	4	Sanger	187

present at different frequencies in the analyzed cohort. By restricting the analysis to exonic changes, we identified in total six missense variants, of which one was later found to be a false positive due to low coverage, as ascertained with Sanger sequencing (not shown). The remaining five variants are listed in Table 2.

Within the *PRPF4* sequence, we identified two annotated variants, p.His78Arg and p.Pro187Ala (Table 2). The first one corresponded to dbSNP entry rs1138958 and was present in 60 European heterozygotes from the Exome Variant Server database; therefore, we considered the variant to be non-pathogenic. The second variant, found in a single individual (ID: 001–417), involved nucleotide c.559C (NM_004697.3, exon 5), which was flagged in dbSNP as entry rs187531407 and referred to a CCC>TCC (p.Pro187Ser) change, found only in two non-validated 1000 Genomes reports. We ascertained that one of these reports (a low-coverage genome in an African sample) was a false positive, following validation with direct Sanger sequencing on the original DNA sample (ID: NA18933, Coriell DNA repository). Moreover, although the nucleotide is the same, the base change in the patient in our

cohort was different compared to rs187531407. More specifically, we identified a CCC>GCC change, which resulted in p.Pro187Ala. Proline 187 is not fully conserved across different species and, according to predictions with different tools, the likelihood of pathogenicity of p.Pro187Ala is uncertain (Table 3). We followed up this change by analyzing controls and available relatives. Public databases, as well as sequencing of in-house controls, did not reveal the presence of this change. We then screened an additional 202 patients with adRP from the same cohorts for this specific variant, but this change was not found in any additional individuals. The affected sibling (ID: 226–1953) of the index patient also carried the same change, but other family members were not available for further segregation analysis. Since the base change is located within exon 5, at a 5-bp distance from the donor splice site, we also checked if splicing of this exon was affected. Bioinformatic prediction was negative, and reverse-transcription (RT)–PCR of patient’s cDNA did not reveal missplicing events. In the absence of additional elements, we could neither exclude nor validate this change as a potential mutation.

TABLE 2. VARIANT OUTPUT FROM NGS SCREENING OF *PRPF4* AND *EFTUD2* AFTER FILTERING FOR CODING CHANGES.

Gene	Sample count	Genomic Position	Coding region change	Amino acid change	DATABASE
<i>PRPF4</i>	3/200	9:116049532	NM_004697.4:c.233A>G	NP_004688.2:p. His78Arg	rs1138958
<i>PRPF4</i>	1/200	9:116053770	NM_004697.4:c.559C>G	NP_004688.2:p. Pro187Ala	rs187531407
<i>EFTUD2</i>	1/200	17:42953357	NM_004247.3:c.814A>G	NP_004238.3:p. Thr272Ala	rs150633454
<i>EFTUD2</i>	1/200	17:42956968	NM_004247.3:c.658C>T	NP_004238.3:p. Arg220Cys	-
<i>EFTUD2</i>	1/200	17:42963986	NM_004247.3:c.238A>C	NP_004238.3:p. Ile80Leu	-

Genomic coordinates refer to assembly GRCh37.p10. Numbering of coding region starts at A of the ATG. Human variation database search included dbSNP 137, 1000 Genomes Project, Exome Variant Server, Complete Genomics control samples and 500 exomes from the CoLaus cohort.

TABLE 3. CHARACTERIZATION OF NOVEL CHANGES IDENTIFIED IN THE COMPLETE SCREENING.

Patient's origin	Gene	Change	Polyphen prediction	PON-P prediction	Controls	Segregation
German	<i>PRPF4</i>	p.Pro187Ala	Probably damaging	Unclassified, Probability of pathogenicity: 0.27	0/189	Positive
American Indian-French Canadian/Irish	<i>EFTUD2</i>	p.Arg220Cys	Probably damaging	Pathogenic, Probability of pathogenicity: 0.91	0/150	Positive
-	<i>EFTUD2</i>	p.Ile80Leu	Benign	Neutral, Probability of pathogenicity: 0.02	Not available	Not available
Italian	<i>NHP2L1</i>	chr22:42078408 NM_005008.2:c.-46G>A		Creation of upstream out of frame ORF	0/150	Negative

Prediction of pathogenicity was made using Polyphen and PON-P [22]. In addition to database consultation, in-house controls were performed by direct Sanger sequencing (*PRPF4*) or restriction enzymes (*EFTUD2* and *NHP2L1*).

Of the three missense changes found in the *EFTUD2* gene two were novel and present in single individuals (p.Arg220Cys and p.Ile80Leu) while one (p.Thr272Ala) was a rare variant found in two heterozygote control African samples from Exome Variant Server (rs150633454; Table 2). The p.Arg220Cys change (ID: 001–492) was confirmed with Sanger sequencing and predicted to be damaging by several prediction tools (Table 3). The residue is in fact highly conserved, from human to yeast. The change was not found in public variation databases, in 150 in-house controls tested, and in the remaining patients from the other cohorts consisting of the 202 patients with adRP. Although we verified that the patient's healthy sister (ID: 226–2008) and son (ID: 226–2009) did not carry this change, the unavailability of other family members prevented us from investigating this variant further. The novel p.Ile80Leu missense and the rare p.Thr272Ala were predicted to be neutral, based on conservation and strength of change, and considered non-pathogenic. Moreover, for the p.Thr272Ala change it was possible to perform cosegregation analysis, and the results were negative.

Since with LR-PCRs we amplified coding and non-coding regions of the target genes, we tested whether any novel variant identified could affect sequences important for splicing signals, including those located in deep intronic regions. We analyzed 1,008 variants in exons and introns of the *EFTUD2* and *PRPF4* genes with the Shannon Human Splicing Pipeline [28]. Only single nucleotide substitutions, but not insertions or deletions, could be tested with this method. No change inactivated or reduced the strength of the natural splice sites. Four hundred and thirty-four variants were predicted to alter the sequence information of cryptic splice sites. By filtering for variations that were not polymorphic and that resulted in the creation of a donor or acceptor splice site with greater strength than the natural splice site,

only five variants remained (Appendix 4). Two were likely false positives because they were present only in one read out of fivefold coverage. For the remaining three, other predictions were made using the NNSPLICE algorithm and did not agree with the one of the Shannon pipeline. According to NNSPLICE, in fact, in two cases the new cryptic splice sites were still weaker than the natural ones, and in one case the already existing cryptic site was weakened by the change. We therefore concluded that there were insufficient elements to study these variants further.

Finally, a second run of variant calling was performed on alignments obtained with less stringent criteria, to exclude the possibility of false negatives due to too rigid mapping parameters. This analysis increased by two times the number of known SNPs identified (999 versus 476) and by three times the number of non-reported changes (3,752 versus 1,195), indicating a gain in sensitivity but also a decrease in specificity (Appendix 5). In fact, when we analyzed only the coding changes, in addition to the variants found with previous mappings, we obtained eight false positives, all found in the same sample and localized in a stretch of wrongly aligned reads, as became clear from inspection of the mapping.

Screening of NHP2L1 and AAR2: Two additional genes, *NHP2L1* and *AAR2*, were also selected. Because of their relatively small size, they were screened by Sanger sequencing (Table 1). *NHP2L1* consists of two coding exons and two alternative 5' untranslated regions (UTRs) containing the start codon. Sequencing of the four exons in 303 patients (182 from the United States, 90 from France, and 31 from Spain) revealed only one novel change introducing an ATG start codon in the 5' UTR, which was found in one patient (ID: 001–245) and not in the control population. This change was potentially interesting because it creates an upstream open

reading frame (uORF), the effect of which could be to reduce the rate of translation from the downstream, canonical ATG [30]. However, this change did not segregate with RP in the family.

The sequence of the *AA2R* gene was negative for novel variations in the American cohort (187 samples analyzed). Notably, we found a patient (ID: 001-156) with a frameshift change reported in the Exome Variant Server (NM_015511.3: c.351_352insC). The inspection of DNA variations in *AA2R* in the general population revealed the presence of several truncating variants, which indicates that this gene tolerates haploinsufficiency and that therefore its role is molecular pathology of adRP is unlikely.

DISCUSSION

The adRP-linked splicing factors PRPF31, PRPF3, PRPF8, PRPF6, and hBrr2 are all components of the U4/U6.U5 tri-snRNP, suggesting that there is a common mechanism of pathogenesis in RP related to dysfunction of this complex. It has been shown that mutations in genes encoding these proteins impair the assembly of the tri-snRNP complex [31,32] or affect catalytic activation of the spliceosome [33], leading to pre-mRNA splicing defects and eventually to cell death [34-36]. Because of their higher requirement of RNA processing, photoreceptor cells are particularly sensitive to the accumulation of splicing defects, compared to other tissues or organs [36]. Mutations are thought to act through a haploinsufficiency mechanism because many determine either truncation and degradation of the protein and the transcript [37] or their instability and accumulation in Cajal bodies [34-36].

In this work, we wanted to investigate the hypothesis that interacting proteins of the same functional complex could also have a role in adRP, by screening their DNA sequences in well-characterized cohorts of dominant patients previously analyzed for the most prevalent RP genes. We selected components of the tri-snRNP complex that, based on functional studies, were found to regulate or interact with splicing factors that were already associated with adRP. We used an NGS-based approach that allowed a fast and parallel analysis of these few candidate genes in a large set of patients, enabling in principle the identification of very rare mutations, which are expected in the case of this disease. In recent years, the strategies aiming at identifying the molecular causes of Mendelian diseases, including RP, have shifted toward genome-wide sequencing of patients followed by unbiased or gene-driven prioritization of mutations. However, these approaches are more powerful when analyzing recessive conditions or dominant diseases with no genetic

heterogeneity. For a dominant disease with considerable genetic heterogeneity the computational analysis and validation elements necessary to find a significant association with heterozygous changes become more important [38]. Therefore, we reasoned that for autosomal dominant RP screening many samples for candidate genes could still be an effective option. For our study, next-generation sequencing has been a practical tool for performing targeted resequencing quickly and comprehensively, even when applied to a hypothesis-driven strategy such as the candidate gene approach.

Nevertheless, the sequencing of the coding exons of the genes *NHP2L1* and *AAR2* and of exons and introns of *EFTUD2* and *PRPF4* revealed a few variants that could have an effect at the protein level and that were absent from the general population. Only the p.Arg220Cys missense in the *EFTUD2* gene was predicted to be damaging by multiple predictive tools; however, its putative pathogenicity could not be demonstrated in our patient with RP. Moreover, during the course of this screening the same gene was linked by exome sequencing to a class of rare and sporadic congenital malformation syndromes, in particular to mandibulofacial dysostosis with microcephaly (OMIM 610536) [39,40]. In these patients the mutations were de novo heterozygous missense, frameshift, and null alleles. Although certain phenotypic variability was observed for the *EFTUD2* mutations [41], it seems that they affect early developmental stages and lead to much more dramatic phenotypes than RP. However, it cannot be excluded that other mutations may have milder effects and trigger the same photoreceptor cell death pathway as for RP-linked splicing factors. The unique novel amino acid substitution in the *PRPF4* gene (p.Pro187Ala) was difficult to interpret in terms of pathogenicity in the absence of additional genetic or functional elements, but the evidence that downregulation of this protein in zebrafish leads to splicing defects and photoreceptor degeneration [21] still suggests that the gene might have a role in RP, perhaps with low frequency.

In conclusion, we did not find proof that the genes *EFTUD2*, *PRPF4*, *NHP2L1* and *AAR2* are associated with adRP in patients of European and North American origin. However, we cannot exclude that very rare pathogenic mutations exist in these genes in the same ethnic groups or in other populations, in virtue of the high genetic heterogeneity of the disease, the increasingly low frequency of mutations detected in novel RP genes, and of geographical effects.

Note added in proof: While the current article was under review, mutations in *PRPF4* were identified as a cause of dominant RP in a Chinese cohort of patients, highlighting in fact a possible population-specific effect [42].

APPENDIX 1. PRIMERS USED FOR LONG-RANGE PCR.

To access the data, click or select the words “[Appendix 1.](#)”

APPENDIX 2. PRIMERS USED FOR SHORT RANGE PCR AND SEQUENCING

To access the data, click or select the words “[Appendix 2.](#)”

APPENDIX 3. SUMMARY OF METRICS OF NGS SCREENING RUNS AND PER SAMPLE STATISTICS.

To access the data, click or select the words “[Appendix 3.](#)”

APPENDIX 4. RESULTS OF SPLICE SITES ANALYSIS USING THE SHANNON HUMAN SPLICING PIPELINE.

To access the data, click or select the words “[Appendix 4.](#)”

APPENDIX 5. TOTAL NUMBER OF VARIANTS IDENTIFIED BY TWO ALIGNMENTS WITH DIFFERENT MAPPING CRITERIA.

To access the data, click or select the words “[Appendix 5.](#)”

ACKNOWLEDGMENTS

For this study we acknowledge the help of the teams of Drs. Keith Harshman and Marzanna Künzli from the Lausanne Genomic Technologies Facility and the Functional Genomic Centre Zurich, respectively, for the sequencing experiments; Dr. Peter Rogan (Cytognomix) for providing free access to the Shannon Human Splicing Pipeline, used to analyze the variants of this study. We are grateful for access to exome sequence data from the CoLaus cohort, which was sequenced as part of a partnership between the Wellcome Trust Sanger Institute, the CoLaus principal investigators and the Quantitative Sciences Department of GlaxoSmithKline. This study was supported by the following research grants: the Swiss National Science Foundation (Grant Number 310030_138346), the Gebert-Rüf Foundation (Rare Diseases – New Technologies Grant), the Foundation Fighting Blindness USA, CIBERER 06/07/0036, FIS PI 13/00226, RD 09–0076–00101 (Retics Bionbank) and Rio Hortega CM12/00013 from the Spanish Ministry of Health, ONCE and FUNDALUCE.

REFERENCES

- Hartong DT, Berson EL, Dryja TP. Retinitis pigmentosa. *Lancet* 2006; 368:1795-809. [PMID: 17113430].

- Daiger SP, Sullivan LS, Bowne SJ. Genes and mutations causing retinitis pigmentosa. *Clin Genet* 2013; 84:132-41. [PMID: 23701314].
- Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res* 2010; 29:335-75. [PMID: 20362068].
- Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 2003; 12:5-14. [PMID: 12887888].
- Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 1998; 92:315-26. [PMID: 9476892].
- McKie AB, McHale JC, Keen TJ, Tarttelin EE, Goliath R, van Lith-Verhoeven JJ, Greenberg J, Ramesar RS, Hoyng CB, Cremers FP, Mackey DA, Bhattacharya SS, Bird AC, Markham AF, Inglehearn CF. Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet* 2001; 10:1555-62. [PMID: 11468273].
- Vithana EN, Abu-Safieh L, Allen MJ, Carey A, Papaioannou M, Chakarova C, Al-Magthteh M, Ebenezer ND, Willis C, Moore AT, Bird AC, Hunt DM, Bhattacharya SS. A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol Cell* 2001; 8:375-81. [PMID: 11545739].
- Chakarova CF, Hims MM, Bolz H, Abu-Safieh L, Patel RJ, Papaioannou MG, Inglehearn CF, Keen TJ, Willis C, Moore AT, Rosenberg T, Webster AR, Bird AC, Gal A, Hunt D, Vithana EN, Bhattacharya SS. Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet* 2002; 11:87-92. [PMID: 11773002].
- Keen TJ, Hims MM, McKie AB, Moore AT, Doran RM, Mackey DA, Mansfield DC, Mueller RF, Bhattacharya SS, Bird AC, Markham AF, Inglehearn CF. Mutations in a protein target of the Pim-1 kinase associated with the RP9 form of autosomal dominant retinitis pigmentosa. *Eur J Hum Genet* 2002; 10:245-9. [PMID: 12032732].
- Zhao C, Bellur DL, Lu S, Zhao F, Grassi MA, Bowne SJ, Sullivan LS, Daiger SP, Chen LJ, Pang CP, Zhao K, Staley JP, Larsson C. Autosomal-dominant retinitis pigmentosa caused by a mutation in SNRNP200, a gene required for unwinding of U4/U6 snRNAs. *Am J Hum Genet* 2009; 85:617-27. [PMID: 19878916].
- Li N, Mei H, MacDonald IM, Jiao X, Hejtmancik JF. Mutations in ASCC3L1 on 2q11.2 are associated with autosomal dominant retinitis pigmentosa in a Chinese family. *Invest Ophthalmol Vis Sci* 2010; 51:1036-43. [PMID: 19710410].
- Tanackovic G, Ransijn A, Ayuso C, Harper S, Berson EL, Rivolta C. A missense mutation in PRPF6 causes impairment of pre-mRNA splicing and autosomal-dominant retinitis pigmentosa. *Am J Hum Genet* 2011; 88:643-9. [PMID: 21549338].
- Fahim AT, Daiger SP, Weleber RG. Retinitis Pigmentosa Overview. In: Pagon RA, Adam MP, Bird TD, Dolan CR, Fong

- CT, Smith RJH, Stephens K, editors. GeneReviews. Seattle (WA)1993.
14. Bowne SJ, Sullivan LS, Avery CE, Sasser EM, Roorda A, Duncan JL, Wheaton DH, Birch DG, Branham KE, Hecklively JR, Sieving PA, Daiger SP. Mutations in the small nuclear riboprotein 200 kDa gene (SNRNP200) cause 1.6% of autosomal dominant retinitis pigmentosa. *Mol Vis* 2013; 19:2407-17. [PMID: 24319334].
 15. Dryja TP. Gene-based approach to human gene-phenotype correlations. *Proc Natl Acad Sci USA* 1997; 94:12117-21. [PMID: 9342372].
 16. Achsel T, Ahrens K, Brahms H, Teigelkamp S, Luhrmann R. The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol Cell Biol* 1998; 18:6756-66. [PMID: 9774689].
 17. Liu S, Rauhut R, Vornlocher HP, Luhrmann R. The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA* 2006; 12:1418-30. [PMID: 16723661].
 18. Weber G, Cristao VF, de LAF, Santos KF, Holton N, Rappsilber J, Beggs JD, Wahl MC. Mechanism for Aar2p function as a U5 snRNP assembly factor. *Genes Dev* 2011; 25:1601-12. [PMID: 21764848].
 19. Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R, Luhrmann R. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J* 1999; 18:6119-33. [PMID: 10545122].
 20. Lauber J, Plessel G, Prehn S, Will CL, Fabrizio P, Groning K, Lane WS, Luhrmann R. The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA* 1997; 3:926-41. [PMID: 9257651].
 21. Linder B, Dill H, Hirmer A, Brocher J, Lee GP, Mathavan S, Bolz HJ, Winkler C, Lagerbauer B, Fischer U. Systemic splicing factor deficiency causes tissue-specific defects: a zebrafish model for retinitis pigmentosa. *Hum Mol Genet* 2011; 20:368-77. [PMID: 21051334].
 22. Blanco-Kelly F, Garcia-Hoyos M, Corton M, Avila-Fernandez A, Riveiro-Alvarez R, Gimenez A, Hernan I, Carballo M, Ayuso C. Genotyping microarray: mutation screening in Spanish families with autosomal dominant retinitis pigmentosa. *Mol Vis* 2012; 18:1478-83. [PMID: 22736939].
 23. Hamel CP. Gene discovery and prevalence in inherited retinal dystrophies. *C R Biol* 2014; 337:160-6. [PMID: 24702842].
 24. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, Mackenzie AP, Caruccio NC, Zhang X, Shendure J. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol* 2010; 11:R119-[PMID: 21143862].
 25. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010; 2010:t5448-prot. [PMID: 20516186].
 26. Firmann M, Mayor V, Vidal PM, Bochud M, Pecoud A, Hayoz D, Paccaud F, Preisig M, Song KS, Yuan X, Danoff TM, Stirnadel HA, Waterworth D, Mooser V, Waeber G, Vollenweider P. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord* 2008; 8:6-[PMID: 18366642].
 27. Olatubosun A, Valiaho J, Harkonen J, Thusberg J, Vihinen M. PON-P: integrated predictor for pathogenicity of missense variants. *Hum Mutat* 2012; 33:1166-74. [PMID: 22505138].
 28. Shirley BC, Mucaki EJ, Whitehead T, Costea PI, Akan P, Rogan PK. Interpretation, stratification and evidence for sequence variants affecting mRNA splicing in complete human genome sequences. *Genomics Proteomics Bioinformatics* 2013; 11:77-85. [PMID: 23499923].
 29. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997; 4:311-23. [PMID: 9278062].
 30. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci USA* 2009; 106:7507-12. [PMID: 19372376].
 31. Makarova OV, Makarov EM, Liu S, Vornlocher HP, Luhrmann R. Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J* 2002; 21:1148-57. [PMID: 11867543].
 32. Huranová M, Hnilicova J, Fleischer B, Cvackova Z, Stanek D. A mutation linked to retinitis pigmentosa in HPRP31 causes protein instability and impairs its interactions with spliceosomal snRNPs. *Hum Mol Genet* 2009; 18:2014-23. [PMID: 19293337].
 33. Mozaffari-Jovin S, Wandersleben T, Santos KF, Will CL, Luhrmann R, Wahl MC. Inhibition of RNA Helicase Brr2 by the C-Terminal Tail of the Spliceosomal Protein Prp8. *Science* 2013; [PMID: 23704370].
 34. Schaffert N, Hossbach M, Heintzmann R, Achsel T, Luhrmann R. RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO J* 2004; 23:3000-9. [PMID: 15257298].
 35. Comitato A, Spampanato C, Chakarova C, Sanges D, Bhattacharya SS, Marigo V. Mutations in splicing factor PRPF3, causing retinal degeneration, form detrimental aggregates in photoreceptor cells. *Hum Mol Genet* 2007; 16:1699-707. [PMID: 17517693].
 36. Tanackovic G, Ransijn A, Thibault P, Abou Elela S, Klinck R, Berson EL, Chabot B, Rivolta C. PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Hum Mol Genet* 2011; 20:2116-[PMID: 21378395].
 37. 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* 2008; 118:1519-31. [PMID: 18317597].

38. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. *Genome Biol* 2011; 12:228-[\[PMID: 21920049\]](#).
39. Lines MA, Huang L, Schwartzentruber J, Douglas SL, Lynch DC, Beaulieu C, Guion-Almeida ML, Zechi-Ceide RM, Gener B, Gillessen-Kaesbach G, Nava C, Baujat G, Horn D, Kini U, Caliebe A, Alanay Y, Utine GE, Lev D, Kohlhase J, Grix AW, Lohmann DR, Hehr U, Bohm D, Majewski J, Bulman DE, Wieczorek D, Boycott KM. Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. *Am J Hum Genet* 2012; 90:369-77. [\[PMID: 22305528\]](#).
40. Luquetti DV, Hing AV, Rieder MJ, Nickerson DA, Turner EH, Smith J, Park S, Cunningham ML. "Mandibulofacial dysostosis with microcephaly" caused by EFTUD2 mutations: expanding the phenotype. *Am J Med Genet A* 2013; 161A:108-13. [\[PMID: 23239648\]](#).
41. Gordon CT, Petit F, Oufadem M, Decaestecker C, Jourdain AS, Andrieux J, Malan V, Alessandri JL, Baujat G, Baumann C, Boute-Benejean O, Caumes R, Delobel B, Dieterich K, Gaillard D, Gonzales M, Lacombe D, Escande F, Manouvrier-Hanu S, Marlin S, Mathieu-Dramard M, Mehta SG, Simonic I, Munnich A, Vekemans M, Porchet N, de Pontual L, Sarnacki S, Attie-Bitach T, Lyonnet S, Holder-Espinasse M, Amiel J. EFTUD2 haploinsufficiency leads to syndromic oesophageal atresia. *J Med Genet* 2012; 49:737-46. [\[PMID: 23188108\]](#).
42. Chen X, Liu Y, Sheng X, Tam PO, Zhao K, Chen X, Rong W, Liu Y, Liu X, Pan X, Chen LJ, Zhao Q, Vollrath D, Pang CP, Zhao C. PRPF4 mutations cause autosomal dominant retinitis pigmentosa. *Hum Mol Genet* 2014; 23:2926-39. [\[PMID: 24419317\]](#).

Articles are provided courtesy of Emory University and the Zhongshan Ophthalmic Center, Sun Yat-sen University, P.R. China. The print version of this article was created on 18 June 2014. This reflects all typographical corrections and errata to the article through that date. Details of any changes may be found in the online version of the article.