



ARTICLE

Complete exon-intron structure of the RPGR-interacting protein (RPGRIP1) gene allows the identification of mutations underlying Leber congenital amaurosis

Sylvie Gerber¹, Isabelle Perrault¹, Sylvain Hanein¹, Fabienne Barbet¹, Dominique Ducroq¹, Imad Ghazi², Dominique Martin-Coignard³, Corinne Leowski⁴, Tessa Homfray⁵, Jean-Louis Dufier², Arnold Munnich¹, Josseline Kaplan^{*1} and Jean-Michel Rozet¹

¹Unité de Recherches sur les Handicaps Génétiques de l'Enfant. Hôpital Necker–Enfants Malades, Paris, France; ²Service d'Ophthalmologie, Hôpital Necker–Enfants Malades, Paris, France; ³Centre Hospitalier Régional du Mans, France; ⁴Institut National des Jeunes Aveugles, Paris, France; ⁵Saint George's Hospital Medical School, London, UK

Leber congenital amaurosis (LCA) is a genetically heterogeneous autosomal recessive condition responsible for congenital blindness or greatly impaired vision since birth. So far, six LCA loci have been mapped but only 4 out of 6 genes have been identified. A genome-wide screen for homozygosity was conducted in seven consanguineous families unlinked to any of the six LCA loci. Evidence for homozygosity was found in two of these seven families at the 14q11 chromosomal region. Two retinal specific candidate genes were known to map to this region, namely the neural retina leucine zipper (NRL) and the retinitis pigmentosa GTPase regulator interacting protein (RPGRIP1). No mutation of the NRL gene was found in any of the two families. Thus, we determined the complete exon-intron structure of the *RPGRIP1* gene. *RPGRIP1* encompasses 24 coding exons, nine of which are first described here with their corresponding exon-intron boundaries. The screening of the gene in the two families consistent with linkage to chromosome 14q11 allowed the identification of a homozygous null mutation and a homozygous missense mutation, respectively. Further screening of LCA patients unlinked to any of the four already identified LCA genes ($n=86$) identified seven additional mutations in six of them. In total, eight distinct mutations (5 out of 8 truncating) in 8/93 patients were found. So far this gene accounts for eight out of 142 LCA cases in our series (5.6%). *European Journal of Human Genetics* (2001) 9, 561–571.

Keywords: Leber congenital amaurosis; homozygosity mapping; RPGR-interacting protein (RPGRIP1) gene; complete gene structure

Introduction

Originally described by Theodore Leber in 1869, Leber congenital amaurosis (LCA, MIM 204000) is an autosomal recessive condition distinct from other retinal dystrophies

and responsible for congenital blindness.¹ The diagnosis is usually made at birth or during the first months of life in an infant with total blindness or greatly impaired vision, normal fundus and non recordable electroretinogram (ERG,²). It is usually accepted that LCA accounts for 5% of all retinal dystrophies.³

LCA is a genetically heterogeneous condition, and four disease genes have been hitherto identified: *GUCY2D* (retGC-1),⁴ *RPE65*,^{5,6} *CRX*^{7–9} and *AiPL1*¹⁰. Taken together these four genes account for only one third of the LCA cases

*Correspondence: J Kaplan, Unité de Recherches sur les Handicaps Génétiques de l'Enfant. Hôpital Necker–Enfants Malades, 149 rue de Sèvres, 75743, Paris Cedex 15, France. Tel: +33 1 44 49 51 61; Fax: +33 1 47 34 85 14; E-mail: Kaplan@necker.fr

Received 1 May 2001; revised 17 May 2001; accepted 22 May 2001

of our series. Two additional loci have been reported, namely LCA3 and LCA5 on chromosomes 14q24 and 6q11-q16, respectively.^{11,12}

In an effort to identify novel LCA loci, we have undertaken a genome-wide search for homozygosity in seven consanguineous and multiplex families with no mutation in known LCA genes, and unlinked to LCA3 or LCA5. We found evidence of homozygosity at the D14S283 and D14S275 loci on chromosome 14q11 in two out of seven families. Interestingly, the gene encoding the 'retinitis pigmentosa GTPase regulator interacting protein' (RPGRIP1) was found to map to the 14q11 chromosomal region.^{13,14}

Here, we report the characterization of the complete exon-intron structure of the RPGRIP1 gene and the identification of mutations in two families consistent with linkage to chromosome 14q11 as well as in six out of 86 unrelated patients. Taken together, our results show that RPGRIP1 accounts for eight out of 142 LCA families (5.6%), while the other four reported genes account for 49 out of 142 LCA families (34.5%).

Patients and methods

Patients

In addition to the seven consanguineous and multiplex families selected for the genome-wide search for homozygosity, 86 families were studied of which 14 were multiplex, 11 were consanguineous, and five were consanguineous and multiplex. All patients fulfilled the minimal criteria for diagnosis of LCA described elsewhere.^{19,20}

Genome-wide search for homozygosity

A genome-wide search for homozygosity was undertaken independently in all patients of the seven LCA families ($n=17$) using the 382 pairs of fluorescent oligonucleotides of the Genescan Linkage Mapping Set, Version II (Perkin Elmer Cetus) under conditions recommended by the manufacturer. Amplified fragments were electrophoresed and analysed on an automatic sequencer (ABI 377). The polymorphic markers have an average spacing of 10 cM throughout the genome.

Mutation screening of NRL in the two families linked to 14q11

The three exons (two coding exons) of the NRL gene were amplified and sequenced using specific primers designed from the intronic sequences nearby the intron-exon junctions (data not shown, available on request).

Characterization of the complete exon-intron structure of the human RPGRIP1 gene

Human retinal total RNA (4.5 μ g) was heated for 10 min at 65°C, in a buffer (19 μ l) containing 1 mM dithiothreitol, 5 mM MgCl₂, 1 mM dNTP, 2.5 μ M random hexamers, 1 \times buffer II, 1 unit RNase Inhibitor (GeneAmp RNA PCR Kit, Applied Biosystems, Foster City, USA). After ice quenching (5 min),

the RNAs were reverse transcribed by adding 2.5 units of MuLV Reverse Transcriptase (GeneAmp RNA PCR Kit, Applied Biosystems, Foster City, USA) for 1.5 h at 42°C.

Reverse transcribed mRNAs (1 μ l) were subsequently submitted to successive PCR amplifications using specific forward primers designed from exons-4, -3, -2, -1 and 1 of the C14000056 clone and a unique reverse primer located in exon 10 of the RPGRIP1 gene (Figure 2). Reactions were carried out in a buffer (20 μ l) containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 200 μ M of each dNTP, 0.5 μ M of each primer and 0.75 U of Taq DNA polymerase (Life Technologies). After an initial denaturation for 5 min at 96°C, denaturation was at 96°C for 20 s, annealing was for each combination of primers at 60°C for 15 s and extension was at 72°C for 30 s, for 34 cycles followed by a final extension, for 7 min at 72°C. The 1 kb fragment amplified between primers PF1/PR was loaded onto a 1% low melting agarose gel and purified by phenol-chloroform extraction and recovered by ethanol precipitation. The purified fragment was directly sequenced using the Big Dye Terminator Cycle Sequencing Kit (ABI Prism).

The genomic structure of the RPGRIP1 gene was determined by comparing the PF1/PR 1 kb sequence with the genomic sequence of the C14000056 clone.

Mutation screening of the RPGRIP1 in the two families linked to 14q11

The 24 exons encoding RPGRIP1 were amplified using specific primers designed from the intronic sequences nearby the intron-exon junctions (Table 1). The PCR and sequencing conditions are similar to that described above except for annealing temperature which was at 56°C for each exon.

Segregation analysis of the G2237A mutation

The G2237A mutation created a *TaqI* restriction site. Exon 15 of the RPGRIP1 gene was amplified as described in the previous section. PCR products were digested for 1 h at 65°C using restriction enzyme *TaqI* (5 u) in the buffer supplied by the manufacturer (Life Technologies). Digested fragments were electrophoresed on a 3% agarose low melting and 1% agarose gel.

Screening of a large series of LCA patients

The search for mutations was carried out by single strand conformation polymorphism (SSCP) analysis and/or denaturing high-pressure liquid chromatography (DHPLC).

For SSCP analysis, genomic DNA (200 ng) was submitted to PCR amplification using 0.5 μ M of specific primers (Table 1) and alpha ³³P dCTP (0.07 μ l, 10 mCi/ml NEN). Amplified DNA was mixed with an equal volume of formamide loading dye (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol), denatured for 10 min at 95°C, loaded onto a polyacrylamide gel (30 \times 40 \times 0.04 cm MDE, TEBU) and electrophoresed at room temperature for 14 h at 4 W, and at 4°C for 14 h at 8 W. Gels were transferred onto

Table 1 PCR primers designed for the amplification of the 24 exons encoding RPGRIPI

Exon number	Forward sequence (5'–3')	Reverse sequence (5'–3')	Size of amplified product (bp)
1	GACATCCTAAAGTTGCATG	GTTCCACAGTGAGAGTTC	231
2	CTCTCTGGACAAGATGTG	AAATTTAAGGAGAACTCTA	245
3	TAACTGTCTATGAAAGGAGAAG	AGTCCTTCCCAGTGTCTT	366
4	GTTCCGGAGGGTACTGTT	CTTCCCTGATCATGCTGAA	263
5	CCTCGACATGTACCAAGGT	TTCCTCTGAGATGGAGGAA	411
6	AGGGCATAGTCAAGGAGAA	CTGAATTGTGGCTTCTCATA	282
7	GACTACTTGGCAAGCTAGG	TACTTGGAGATGAACATAGAT	282
8	CGTGCTGAGTGATATGACC	ACACATTCTAACATCCCTGA	257
9	GAAATCCTGTGCAGGGGAA	TCATAGTAACACTGCTCTG	175
10	ACGTCATATCACACCCCTT	GAGGTAGAGGATGCCACA	257
11	TGGTGATAAATACTACAGAAT	GCCAGAGATCTCTCTGCAG	261
12	CCTGTCATATTTATACTCCCT	TTCCACGTTTCTGTATC	240
13	TTGGTTTTAGGCCACTGAGA	CTGTGGAAGGGTCCCGAA	247
14	GAAAGAGCTCCCTACCCTT	GGAAATTCTGCATTGGTGC	550
15	CTTGCCACACCATCTGTAT	GTTTGGCTGAGGCTCCTC	250
16	ACAGTCTCAAGCTGCCCTT	AGACAACACTGGGAAGAGG	442
17	CTTATTTTCATGTGATCAGGTC	CATAGGATTGGCAGAGATC	278
18	GTTGTAAACTACCAGCTTG	GGGACACTACACCCACAA	301
19	CCAAGATATTACCAGCTATG	AGCCTGATCTCGTGATCTG	258
20	GGCTAAAGTGCTTTGAAACA	TATCGTCTTATCTCGTATGC	194
21	CTTGGAGCCTCACTAACC	TTCATCAGACTTCCTCACC	287
22	CACTGCAACAGTATATGATTC	CTGTCTCAATATCTCCTTTG	181
23	GCATTAAGAGTATCAACAGTG	TTAGGATGATTTCTTTGGAA	228
24	AGTGTTCAACTGAGTGATGC	ACTATGGTTCCCTCAGAGAC	200

3 MM Whatman paper, dried and autoradiographed with Kodak BIOMAX-MR films for 16–48 h.

For DHPLC analysis, heteroduplex formation was induced by heat denaturation of PCR products at 94°C for 10 min, followed by gradual reannealing from 94°C to 25°C over 30 min. DHPLC analysis was performed with the WAVE DNA fragment analysis system (Transgenomic). PCR products were eluted at a flow rate of 0.9 ml/min with a linear acetonitrile gradient. The values of the buffer gradients (buffer A, 0.1 M triethylammoniumacetate; buffer B, 0.1 M triethylammoniumacetate/25% acetonitrile), start and end points of the gradient, and melting temperature predictions were determined by the WAVEMAKER software (Transgenomic). Analysis per sample took 7.5 min, including regeneration and re-equilibration to the starting conditions. Optimal run temperatures were empirically determined. Mobile-phase temperatures were assessed within a 5°C window above and below the suggested run temperature on the basis of each fragment's characteristic melting profile.

PCR fragments displaying SSCP and/or DHPLC abnormal profiles were further sequenced as described above.

Results

Genome-wide search for homozygosity

The 382 polymorphic markers of the Genescan Linkage Mapping Set Version II were tested in seven consanguineous families. For each of the seven families, evidence of homozygosity was found for several independent loci on

various chromosomes. Subsequently, polymorphic markers flanking the various homozygote loci were studied in the corresponding families. These analyses allowed to exclude several chromosomal regions since (i) homozygosity for the marker only reflected a lack of informativity in the parents, and surrounding markers were heterozygotes in the patients, or (ii) the parents were indeed informative for a given marker but flanking markers in the immediate vicinity ($\Theta \sim 0\%$) were heterozygotes in the patients (data not shown, available on request). Conversely, we found evidence for homozygosity at the D14S283 and D14S275 loci in all patients of two unrelated families (SED and NIL families, Figure 1). Homozygosity for this chromosomal region was confirmed by flanking markers and the shortest area of homozygosity was defined by the D14S261 and D14S257 interval (Figure 1). This hint of mapping led us to search for candidate genes in this chromosomal region. Interestingly, two genes highly expressed in the retina were found to map to in this genetic interval.

Screening of the neural retina-specific leucine zipper (NRL) gene

A search for mutations in the neural retina-specific leucine zipper gene¹⁵ was undertaken in the homozygote patients of the two consanguineous families linked to this region (Families SED and NIL; Figure 1). Direct sequencing of the three exons of the NRL gene revealed no deleterious change in any of the patients, supporting the exclusion of this gene in the two families.

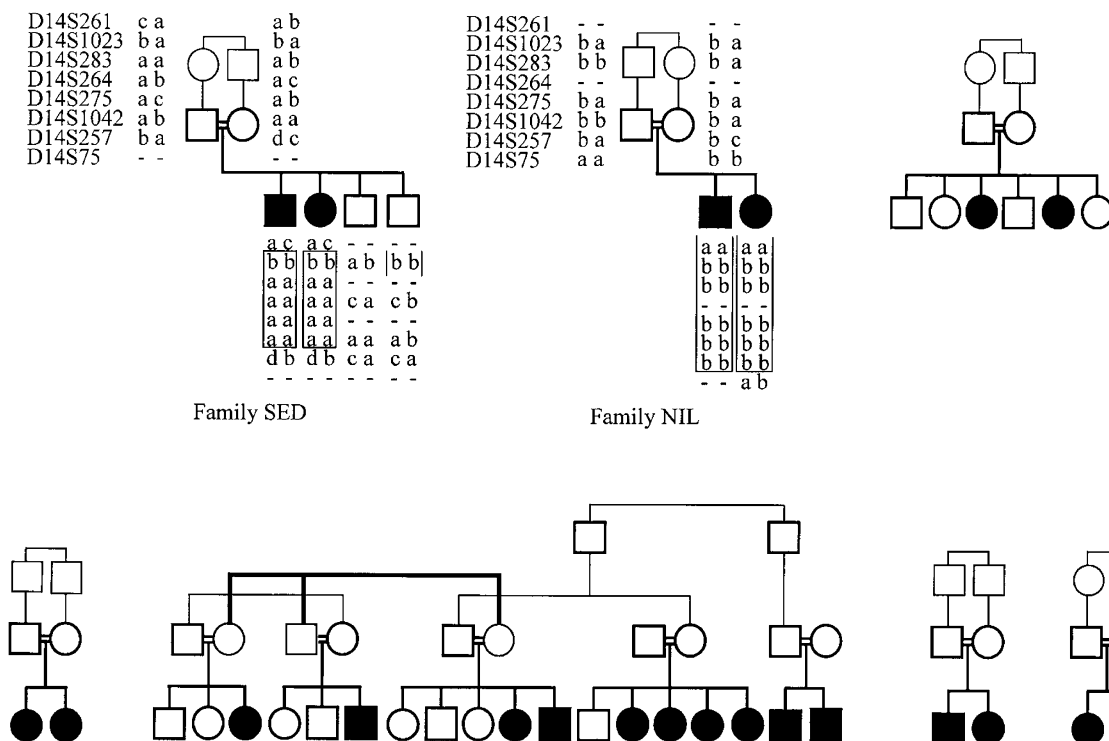


Figure 1 Pedigree of the seven consanguineous and multiplex families ascertained for the genome-wide search for homozygosity and haplotypes of two of them consistent with linkage to the 14q11 region. Regions of homozygosity are squared.

Characterization of the complete exon-intron structure of the human RPGRIP1 gene

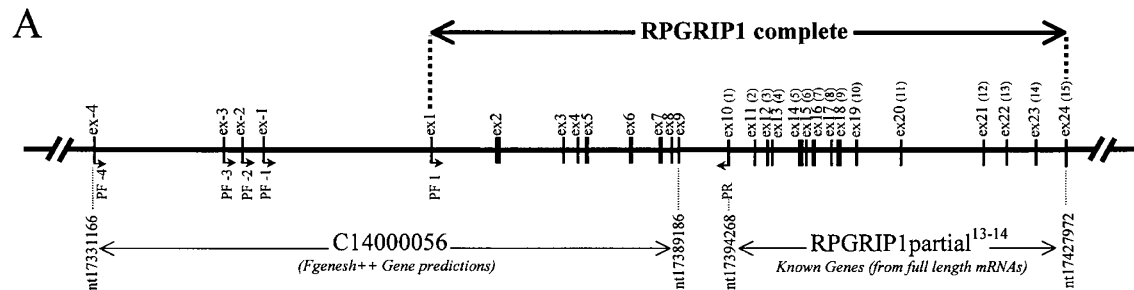
Two groups have recently reported the identification of a full length human RPGRIP1 cDNA encoding a 902 amino acid protein.^{13,14} Sequence alignment of this predicted protein with the RPGRIP1 orthologues showed that the mouse and bovine counterparts were extending forward the N-terminal region (Entrez Protein Sequence Search accession numbers AAG10247, NP_076368, respectively). Subsequently, we tried to determine whether the human gene stretched upstream the 5' end of this cDNA sequence, especially as several alternative transcripts were described.^{13,14,18} A search in the Human Genome Project Working Draft at USC using the Draft Genome Browser v4 allowed the identification of a 58 kb sequence (C14000056, nt17331166–17389186; Figure 2A), 5 kb upstream the 5' end of the originally described RPGRIP1 gene (referred to as RPGRIP1 partial, nt17394268–17427972; Figure 2A). This clone was predicted to contain 13 exons (Fgenesh++ Gene Predictions; Figure 2A). The cDNA sequence including exons 1–9 contained an open reading frame (ORF, start: nt17364655) encoding a 384 amino acid sequence highly similar to the N-terminal part of the murine RPGRIP1 protein.

No PCR amplifications of reverse transcribed human retinal RNA were obtained when using primers PF-4/PR, PF-

3/PR, PF-2/PR, PF-1/PR, giving support to the view that predicted exons-4 to -1 of the C14000056 sequence did not belong to the RPGRIP1 gene (Figure 2A). Conversely, a 1321 bp fragment was amplified using primers PF 1/PR (Figure 2A). Direct sequencing of the PF 1/PR fragment supported the location of the translation initiation codon at position nt17364655 (referred to as nt1 in Figure 3) and showed that (i) the predicted intron 8/exon 9 boundary of the C14000056 clone was to be shifted 6 bp in the 5' direction, (ii) the four bases at the 3' end of the C14000056 were intronic and (iii) the 5 kb sequence between the original RPGRIP1 and C14000056 sequences corresponded to an intron (Figure 2A). In total, the RPGRIP1 gene consists in 24 exons and 23 introns (Figure 3). The predicted product would yield a protein product of 1287 amino acids (Figure 3). Specific primers were designed from the intronic sequences in order to screen the 24 coding exons in all patients of our series (Table 1).

Identification of RPGRIP1 mutations in LCA patients

Very recently, analysis of radiation hybrids have resulted in the assignment of the RPGRIP1 gene to chromosome 14q11, at the D14S932 locus.^{13,14} Therefore, this gene was regarded as an attractive candidate gene by virtue of both position and function. A search for mutations was undertaken in



From the Human Genome Project Working Draft at USC - Draft Genome Browser v4 - Chromosome 14, Bases 17331166-17427972

B

Human: 1 MSHLVDTSGDLFVRDIDATPLVLPASKGKNMKTQPLSRMNEELEDSEFFRLREDHMLVKELSWKQDEIKXXXXXXXXXXXXXXXXXXVAEEAAPLSETARRGQKAGWRQRLSMHQRPQ 120
M HL++ DLPVRD D+ PL L + GKN++ QF L RMN++EL L E+ LVKE S KQ+D+ + A L E RR + W P
Mouse: 1 MQHLLIYMPEDLPVRDTSPL-LKGTSGKNVRAQPHLGRMNQKELNCRRLHLHEEPTLVKEPSFKQDKNRNRRRTNVQRSTTTQPDRLTL---AVLQEPERR--RRPVWSASPSAPSAPP 114

Human: 121 MHRLQG---HFHCVGFPASP--RAQPRVQVGHRLHTAGAPVPEKPRGRDRLSYTAAPPSEKHEATNEN--RGEVASKFSELVXXXXXXXXXXXXXMAKPIGLCMPSAHIMASNTMQVE 235
+ G H + P++ AQPRV G R H AG P DR S+TAPP+FK++ ++N R E+ +PS+L H M +++ VE
Mouse: 115 RAPVPGRKAHVQRLCPSTAVGSAQPRVHAGRRLLPHIAG-----PNDRRSHTPAPFAFDYVADKNTRIEITREPSQLT-----HTMTTDSHVE 197

Human: 236 EPPKSPEKMPKPNFEQRSSLECAQKAAELRASIKKVELIRLKKLLHERNASLVMTKAQLTEVQAYETLLQKNQGISAAHEALLKQVNELRAELKEESKKAVALSKSQLEDVSILOM 355
E P+SPEK K E EQRSS EC QKAAELRASIKK VELIRLKKLL ERN SL T+AQLT VQAYE LLQKNQGISL AH A L QVNEL+REL EESKKAVAL++QL DVSILQ+
Mouse: 198 EIPRSPEKT-SKVEKPEQRSSECTQKAAELRASIKKVELIRLKKLLQERNTSLAATEAQLTRVQAYEDLLQKNQGISLTAHNAFLSQVNELKAESESKKAVSLRTQLGDVSILOQ 316

Human: 356 TLKEFQERVEDLEKERKLLNDNDYKLLS 384
TLKEFQ RVEDLEKERKLL+D+YD+LLES
Mouse: 317 TLKEFQVRVEDLEKERKLLSDSYDRLLS 345

Figure 2 (A) Identification of nine additional exons of the human RPGRIP1 gene. Based on the Human Genome Project Working Draft, the 34 kb sequence of chromosome 14 lying between nt17394268 and nt17427972 contained the 15 RPGRIP1 exons independently reported in 2000 by Roepman *et al*¹³ and Boylan and Wright¹⁴ ('RPGRIP1 partial', exons 1–15 in brackets). Five kilobases upstream the 5' end of this sequence, the 58 kb long C14000056 clone (nt17331166–17389186) was predicted to contain 13 exons (ex-4, ex-3, ex-2, ex-1, ex1 to ex9). Nine of them (ex1 to ex9) were found to represent the nine first exons of the complete RPGRIP1 gene which consists in 24 encoding exons. PF-4, PF-3, PF-2, PF-1 and PF 1 show the location of forward primers alternatively used with the reverse primer PR to amplify reverse transcribed human retinal mRNA. (B) Conservation of the predicted amino acid sequence encoded by exons 1–9 of the C14000056 clone and the N terminal sequence of the mouse orthologue. The predicted exons 1–9 of the C14000056 clone contains an ORF encoding a 382 amino acid sequence similar to the mouse RPGRIP1 N-terminal region.

homozygote patients of the two families consistent with linkage to 14q11 by direct sequencing of the 15 coding exons originally reported^{13,14} as well as the nine additional exons reported here.

The two patients of the first family of Moroccan origin were found to carry a homozygous missense mutation in exon 15 at nucleotide 2237 (G2237A; Family SED, Figure 4A). This mutation changed a glycine (non polar, uncharged amino acid) into a glutamic acid (charged amino acid) in the protein (G746E). The glycine at position 746 is present in both human 1a and 1b RPGRIP1 isoforms.^{13,14} In addition, it is a highly conserved amino acid, which lies in a very conserved region in both bovine and murine sequences. This base change created a restriction site for the Taq1 enzyme. Restriction analysis of the entire family showed that the mutation segregated with the disease (Figure 4A) and was absent in 252 normal chromosomes. In the second family (Family NIL, Figure 1), we found a homozygous deletion of nt 511 in exon 4 leading to a premature stop codon, 19 amino

acids downstream, in both affected children (Figure 4B). This deletion was found to be inherited from both carrier parents and was not identified in any of the 252 normal chromosomes studied.

Subsequently, we undertook a search for mutations in a large panel of 86 LCA patients unrelated to any of the four already identified LCA genes and found RPGRIP1 gene mutations in six additional families. In one family of compound origin (French, English, German and Bulgarian), the affected child was found to harbour a heterozygote 2 bp insertion (2566+2 insTT) in exon 16, leading to a premature stop codon, three amino acids downstream (Family LOB, Figure 5). The second mutation was found to be a heterozygote insertion of a guanine in exon 23, at position 3629 (ins 3629+1G; Figure 5). This latter mutation resulted in a frameshift and the apparition of a stop codon, five amino acids downstream. The segregation of the two mutations in the family revealed that the 2566+2TT and 3629+1G insertions were inherited from the carrier mother and father,

respectively (Family LOB, Figure 5). In addition, these two mutations resulted into an abnormal DHPLC profile allowing to demonstrate their absence in 252 normal chromosomes.

In another family of Belgian origin, the affected child was found to be a compound heterozygote as well. The first mutation, a C–T transition at nucleotide 2668 (exon 16)

a		
Exon 1	Exon 2	
ATGTCACATCTGGTGGACCTACATCAGGAGACTTCCAGTTAGAGACATAGATGCTATACCTCTGGTGTACACGCTCAAAAGCTAAG		90
M S H L V D P T S G D L P V R D I D A I P L V L P A S K G K		30
AATATGAAACTCAACACCCTTGAGCAGGATGAACGGGAGGAATTGGAGGACAGTTTCTTTCGACTTCGCGAAGATCAGATGTTGGTG		180
N M K T Q P P L S R M N R E E L E D S F F R L R E D H M L V		60
AAGGAGCTTTCTTGAAGCAACAGGATGAGATCAAAAGCTGAGGACCACTTGCTGCGGTTGACCGCTGCTGGCCGGGACCTGCGGGTC	Exon 3	270
K E L S W K Q Q D E I K R L R T T L L R L T A A G R D L R V		90
GCGGAGGAGCGCGCGCTCTCGGAGACCGCAAGCGCGGAGGCGGAGTGGCGGACGCGCTCTCCATGCACAGCGCCCGCAG		360
A E E A A P L S E T A R R G Q K A G W R Q R L S M H Q R P Q		120
ATGCACCGACTGCAAGGCACTTCCACTGCGTCGCGCTGCCAGCCCCCGCGCGCCAGCCTCGCGTCCAAGTGGGACACAGACAGCTC		450
M H R L Q G H F H C V G P A S P R R A Q P R V Q V G H R Q L		150
CACACAGCGGTGCACCGTGCAGGAGAACCAAGAGGGGCGCAAGGAGGAGCTGAGCTACACAGCCCTCCATCGTTTAAGGAGCAT	Exon 4	540
H T A G A P V P E K P K R G P R D R L S Y T A P P S F K E H		180
GCGACAAATGAAACAGAGGTGAAGTAGCCAGTAACCCAGTGAACCTGTTTCTGGTTCTAACAGCATAATTTCTTTCAGCAGTGTCTATA	Exon 5	630
A T N E N R G E V A S K P S E L V S G S N S I I S F S S V I		210
AGTATGGCTAAACCCATTGGTCTATGCATGCCTAACAGTGCACATCATGGCCAGCAATACCATGCAAGTGAAGAGCCACCAAGTCT		720
S M A K P I G L C M P N S A H I M A S N T M Q V E E P P K S		240
CCTGAGAAATGTGCCTAAAGATGAAATTTTGAACAGAGAAGCTCATTTGGAGTGTGCTCAGAAGGCTGCAGAGCTTCGAGCTTCCATT	Exon 6	810
P E K M W P K D E N F E Q R S S L E C A Q K A A E L R A S I		270
AAAGAGAAGGTAGAGCTGATTCGACTTAAGAAGCTCTTACATGAAAGAAATGCTTATTGGTTATGACAAAAGCACAATTAACAGAAGTT		900
K E K V E L I R L K K L L H E R N A S L V M T K A Q L T E V		300
CAAGAGCTCATACGAAACCTTGCTCCAGAAGATCAGGGAATCCTGAGTGCAGCCCATGAGGCCCTCTCAAGCAAGTGAATGAGCTCAGG	Exon 7	990
Q E A Y E T L L Q K N Q G I L S A A H E A L L K Q V N E L R	Exon 8	330
GCAGAGCTGAAGGAAGAAAGCAAGAAGGCTGTGAGCTTGAAGAGCCAACCTGGAAGATGTGTCTATCTTGACAGTACTCTGAAGGAGTTT	Exon 9	1080
A E L K E E S K K A V S L K S Q L E D V S I L Q M T L K E F		360
CAGGAGAGGTGAAGATTTGAAAGAAAGCAAAATTTGCTGAATGACAATTTAGCAAACTCTTAGAAGATGCTGGACAGCAGTGAC	Exon 10	1170
Q E R V E D L E K E R K L L N D N Y D K L L E S M L D S S D		390
AGCTCCAGTCAGCCCCACTGGAGCAACGAGCTCATAGCGGAACAGCTACAGCAGCAAGTCTCTCAGCTGCAGGATCAGCTGGATGCTGAG		1260
S S Q P H W S N E L I A E Q L Q Q Q V S Q L Q D Q L D A E		420
CTGGAGGACAAGAGAAAGTTTACTTGAAGCTGTCCAGGGAGAAAGCTCCAAATGAGGATCTGAAGCTTGAAGTACCAACATACTTCTAG	Exon 11	1350
L E D K R K V L L E L S R E K A Q N E D L K L E V T N I L Q		450
AAGCATAACAGGAAGTAGAGCTCTCCAAATGCAGCCCAATTTCCCAACCTCTGACAGGCAATCTGAACAGCCACTCACCCAGCT		1440
K H K Q E V E L L Q N A A T I S Q P P D R Q S E P A T H P A		480
GTATTGCAAGAGAACTCAGATCGAGCTCAAGTGAACCCAAACCAAGAGAAAGAACTGCTCCAGGTGCTAAATGAGTTGCAAGTA	Exon 12	1530
V L Q E N T Q I E P S E P K N Q E E K K L S Q V L N E L Q V		510
TCACACGCAAGACCATTTGAACTAGAAAAGACAGGACATGCTTATCTGCAGCGCAAAATCAACGTGTGTTATCAGCAGGAACTG	Exon 13	1620
S H A E T T L E L E K T R D M L I L Q R K I N V C Y Q E E L		540
GAGGCAATGATGACAAAAGCTGACAATGATAATAGAGATCACAAGAAAAGCTGGAGAGGTGACTCGACTACTAGACCTCAAGAATAAC		1710
E A M M T K A D N D N R D H K E K L E R L T R L L D L K N N		570
CGTATCAAGCAGCTGGAAGGTATTTAAGAAGCCATGACCTTCCAACTCTGACAGCTCAAAGATGTTGCTTATGGCACCCGACCGTTG	Exon 14	1800
R I K Q L E G I L R S H D L P T S E Q L K D V A Y G T R P L		600
TCGTTATGTTTGAAGAACTGCCAGCCCATGGAGATGAGGATAAAGTGGATATTTCTCTGCTGCATCAGGTTGAGAATCTTTTGAACATG		1890
S L C L E T L P A H G D E D K V D I S L L H Q G E N L F E L		630
CACATCCACAGGCTTCTGACATCTGCCGCCCTAGCTCAGGCTGGAGATACCAACCTACCACTTTCTGCACCTATTCTCTTATGAC		1980
H I H Q A F L T S A A L A Q A G D T Q P T T F C T Y S F Y D		660
TTTGAAACCCACTGTACCCATTATCTGTGGGGCCACAGCCCTCTATGACTTCACCTCCAGTATGTGATGGAGACAGATTCGCTTTTC		2070
F E T H C T P L S V G P Q P L Y D F T S Q Y V M E T D S L F		690
TTACACTACCTTCAAGAGGCTTCAGCCCGGCTTGACATACACAGGCCATGGCCAGTGAACACAGCACTCTTGTGAGGATGGATTGTC		2160
L H Y L D I H Q A M A S E H S T L A A G W I C		720
TTTGACAGGCTGCTAGAGACTGTGAGAGAGTCCATGGCTTGCCACACTGATTCGAGCTGGTGGAGAAGTTCGGGTTCTAGAGTAC	Exon 15	2250
F D R V L E T V E K V H G L A T L I G A G G E E F G V L E Y		750
TGGATGAGGCTGCGTTTCCCCATAAAACCCAGCCTACAGCGTGCATAAAGCAAGAAAGCCAGGTTCTACCTGTCAACCGATGTGCTT		2340
W M R L R F P I K P S L Q A C N K R K K A Q V Y L S T D V L		780
GGAGGCGGAAGGCCAGGAAGAGGAGTTCAGATCGGAGCTTGGGAACCTCAGAAGAGCTGTGGATTGAAATACCAAGTGTGTGGC	Exon 16	2430
G G R K A Q E E E F R S E S W E P Q N E L W I E I T K C C G		810
CTCCGGAGTCGATGGTGGGAACCTCAACCCAGTCCATATGCTGTACCGCTTCTTACCTTTCTGACCATGACACTGCCATCATCCCA		2520
L R S R W L G T Q P S P Y A V Y R F F T F S D H D T A I I P		840

resulted in the substitution of an arginine into a stop codon (R890X). This non-sense mutation is expected to truncate more than half of the protein. The second mutation was a missense mutation in exon 21 at nucleotide 3341 (A3341G, Family NES; Figure 5). This latter mutation changed an aspartic acid (charged amino acid) into a glycine (non polar, uncharged amino acid) in the protein (D1114G). The aspartic acid at position 1114 is present in all known isoforms of the protein and is highly conserved in bovine and murine sequences. In addition, this amino acid lies in the highly conserved and predicted globular C-terminal region corresponding to the cognate RPGR-interacting domain (RID). Each mutation in this patient was inherited from a healthy parent and was absent in 252 normal chromosomes. Interestingly, a patient born to consanguineous parents of Moroccan origin was homozygous for the same missense mutation (D1114G, Family MOU, Figure 5).

In a family of French origin, the affected child was found to carry a heterozygote C–T transition (C1525T) in exon 12 resulting in the substitution of a glutamine for a stop codon (Q509X, Family GRA, Figure 5). The mutation was found to

be inherited from the healthy father. In an other French family, the LCA patient was found to carry a heterozygote 4 bp insertion in exon 12 (1501+4 insTGTC; Family HER, Figure 5), leading to a frameshift and a premature stop codon, eight amino acids downstream. This mutation was inherited from his healthy mother. In both families, the RPGRIP1 mutation identified was not found in any of the 252 normal chromosomes studied.

Finally, in a family of French origin, two first cousins were affected with LCA. One of them (III2) was found to carry a heterozygote 3 bp in frame deletion in exon 24 (del nt 3835–3837; Family GUY, Figure 5). The mutation led to the deletion of the glutamic acid located eight amino acids upstream to the stop codon (del E1279). This amino acid, found in both RPGRIP1 a and b isoforms, is conserved in bovine and murine sequences and lies in the highly conserved RID domain. The E1279 deletion identified in patient III2 was inherited from his healthy father (II1) and was therefore absent in his affected cousin (III4, Figure 5). Both mutations identified in this family resulted in an abnormal DHPLC profile that was not found in any of the 252 normal chromosomes studied. For the last three families,

b

```

GCCAGTAACAACCCCTACTTTAGAGACCAGGCTCGATTCCAGTGCTTGTGACCTCTGACCTGGACCATTTATCTGAGACGGGAGGCTTG 2610
A S N N P Y F R D Q A R F P V L V T S D L D H Y L R R E A L 870

TCTATACATGTTTTGATGATGAAGACTTAGAGCCTGGCTCGTATCTTGGCCGAGCCGAGTGCCTTTACTGCCTCTTGCAGAAATGAA 2700
S I H V F D D E D L E P G S Y L G R A R V P L L P L A K N E 900

Exon 17
TCTATCAAAAGTGTATTTAACCTCACTGACCCCTGCAGAGAAACCCACGGATCTATTCAAGTGCRACTGGATTGGAAGTTTCCCTACATA 2790
S I K G D I F L T D P A E K P N G S I Q V Q L D W K F P Y I 930

CCCCCTGAGAGCTTCCTGAAACCGAAGCTCAGACTAAGGGGAAGGATACCAAGGACAGTTCAAAGATCTCACTGAAGAGGAAAAGGCT 2880
P P E S F L K P E A Q T K G K D T K D S S K I S S E E E K A 960

Exon 18
TCATTTCCTTCCCAATGATCAGATGGCATCTCCTGAGGTTCCCATTAAGCTGGCCAGTATCGATCTAAGAGAAACCTCCTCATGGGGGA 2970
S F P S Q D Q M A S P E V P I E A G Q Y R S K R K P P H G G 990

GAAAGAAAGGAGAAGGAGCACCAGGTTGTGAGCTACTCAAGAAGAAAACATGGCAAAAGAATAGGTGTCAAGGAAAGATAGAAATGGAG 3060
E R K E K E H Q V V S Y S R R K H G K R I G V Q G K N R M E 1020

Exon 19
TATCTTAGCCTTAACATCTTAATGGAATACACCACAGCAGGTGAATTACACTGAGTGAAGTTCTCAGAGACTAACAGCTTCATAGGT 3150
Y L S L N I L N G N T P Q Q V N Y T E W K F S E T N S F I G 1050

Exon 20
GATGGCTTTAAAAATCAGCACGAGGAGAGAAATGACATTATCCCATTCAGCACTGAAACAGAAGAACCTCTACATCCTGTAATGAC 3240
D G F K N Q H E E E E M T L S H S A L K Q K E P L H P V N D 1080

AAAGATCCTCTGAACAAGGTTCTGAAGTCAGTGAAGCACAACATACCGACAGTGATGATGTCATAGTGCCACCCATGTCTCAGAAATAT 3330
K E S S E Q G S E V S E A Q T T D S D D V I V P P M S Q K Y 1110

Exon 21
CCTAAGGCATTTACAGAGAAGATGTGCATTGAAATGTCTCCTGGCCTTCTACCCAGAGGCAGAAGTGTCTGATGAGAACATATAA 3420
P K A D S E K M C I E I V S L A F Y P E A E V M S D E N I K 1140

CAGGTGTATGTGGAGTACAAATTTACGACCTACCCTTGTGGAGACAGAGACTCCAGTGTCCCTAAGGAAGCCTAGGGCAGGAGAAGAA 3510
Q V Y V E Y K F Y D L P L S E T E T P V S L R K P R A G E E 1170

Exon 22
ATCCACTTTCACTTTAGCAAGGTTAATAGACCTGGACCCACAGGAGCAGCAAGGCCGAAGGCGGTTTCTGTTCGACATGCTGAATGGACAA 3600
I H F H F S K V I D L D P Q E Q Q G R R R F L F D M L N G Q 1200

Exon 23
GATCCTGATCAAGGACATTTAAAGTTTACAGTGGTAAGTGATCCTCTGGATGAAGAAAGAAAGATGTGAAGAAGTGGGATATGCATAT 3690
P D Q G H L K F T V V S D P L D E E K K E C E E V G Y A Y 1230

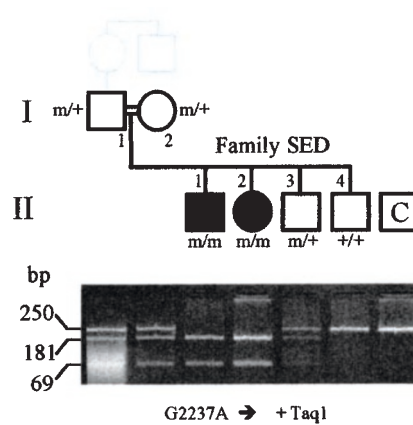
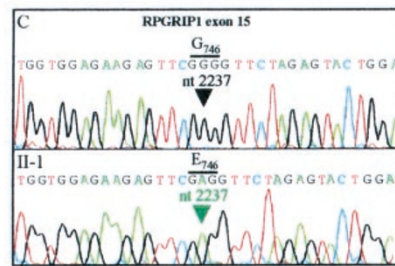
Exon 24
CTTCAACTGTGGCAGATCCTGGAGTCAGGAAGAGATATTTAGAGCAAGAGCTAGACATTTGTTAGCCCTGAAGATCTGGCTACCCCAATA 3780
L Q L W Q I L E S G R D I L E Q E L D I V S P E D L A T P I 1260

GGAAGGCTGAAGGTTTCCCTTCAAGCAGCTGTCTCCATGCTATTACAAGGAGATGACTGAAGATTTGTTTTCATGA 3861
G R L K V S L Q A A A V L H A I Y K E M T E D L F S * 1287

```

Figure 3 cDNA and predicted amino acid sequence of the human RPGRIP1 gene. The translation initiation codon is double underlined while the translation initiation codon reported by Roepman *et al*¹³ and Boylan and Wright¹⁴ is underlined. Intron-exon boundaries are shown.

A



B

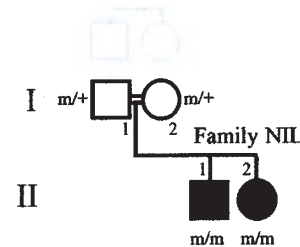
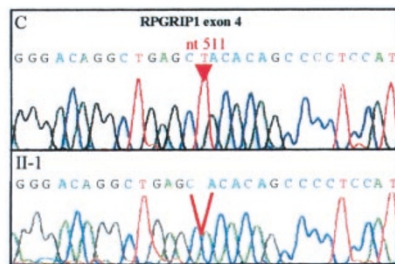


Figure 4 Identification and segregation analysis of the homozygous RPGRIP1 mutations identified in two multiplex and consanguineous families compatible with linkage on chromosome 14q11. (A) Right panel: identification of a homozygous missense mutation G→A in exon 15 at nucleotide 2237 (G746E) in family SED; C=control; II1=LCA patient. (A) Left panel: restriction analysis of the mutation. The G2237A substitution created a Taq1 site. PCR amplified genomic DNA was digested with using Taq1. Homozygous patients (II1 and II2) display two fragments of 181 and 69 bp. Heterozygous carriers of the G2237A (I1, I2, II3) show three fragments of 250 (undigested), 181 and 69 bp. Individuals homozygous for the wild type allele (II4 and the control: C) display the undigested fragment of 250 bp, only. m=mutant allele; +=wild type allele. (B) Identification and segregation of the delta 511 identified in the NIL family. m=mutant allele; +=wild type allele.

the second mutation, searched by direct sequencing, remains unidentified.

Discussion

LCA is a relatively common disease with a high rate of consanguinity. We have long assumed that this condition could be accounted for by mutations in various genes. Indeed, six LCA genes have been hitherto mapped namely, LCA1 and LCA4 on chromosome 17p13.1,^{10,16,17} LCA2 on chromosome 1p31,⁵ LCA3 on chromosome 14q24,¹¹ LCA5 on chromosome 6q11-q16¹² and LCA6 on chromosome 19q13.3.⁷ Among them, four out of six have been identified: the retinal specific guanylate cyclase gene (retGC1, LCA1,⁴), the gene encoding the 65-kD protein specific to the retinal pigment epithelium (RPE65, LCA2,⁵), the cone-rod homeo box-containing gene (CRX, LCA6,⁷) and the gene encoding

the arylhydrocarbon receptor interacting protein-like 1, only 2 Mb centromeric from LCA1 (LCA4,¹⁰). LCA3 and LCA5 respectively account for the disease in a consanguineous Saudi Arabian LCA family and a multigenerational kindred of Old Order River Brethren, a religious isolate descended from Swiss immigrants to America in the 1750s. All together, the four identified genes accounted for only one third of the LCA cases in our series.

We have found homozygosity for polymorphic markers located on chromosome 14q11 in two unrelated consanguineous families. Two known candidate genes were reported to map to this chromosomal region. One of them, NRL, was excluded in both families. In the RPGRIP1 gene, we found a homozygous missense mutation in one of the two families (Family SED) but direct sequencing of the 15 known exons (13–14) failed to detect any mutation in the second family (Family NIL).

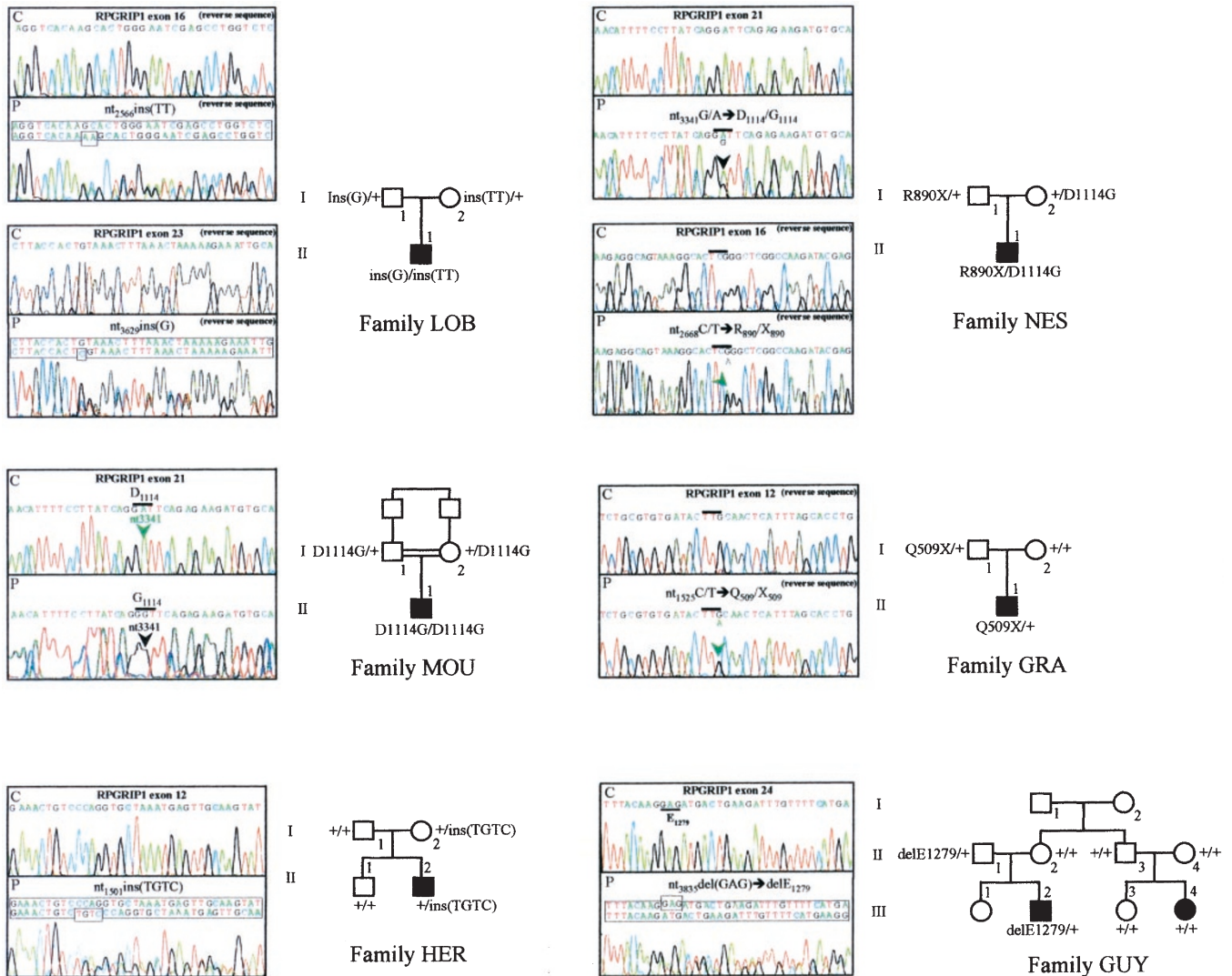


Figure 5 Identification and segregation of the RPGRIP1 mutations in LCA families. C=control; P=patient; +=wild type allele.

To determine whether the missense mutation identified in family SED involved a conserved amino acid, we aligned the bovine, murine and human sequences. This analysis revealed that this amino acid was conserved in all three species, and that the murine and bovine proteins were longer than the human one in the N-terminal region. This last finding along with the lack of detection of mutation in the NIL family prompted us to search for additional exons in the 5' region of the originally reported RPGRIP1 sequence (13–14). This study allowed the identification of nine additional coding exons, one of which was found to be homozygously mutated in the NIL family. Further screening of the 24 RPGRIP1 exons

in 86 additional LCA patients allowed the identification of mutations in six of them.

In total, we report here the identification of mutations in a novel LCA causing gene, in eight unrelated families unlinked to any of the four already identified genes. Among the nine mutations identified, eight were different (five truncating mutations, one in frame deletion of 3 bp and two missense mutations). Three of the eight patients, born to consanguineous parents, harboured homozygous mutations, two out of eight were compound heterozygotes of which one harboured a truncating mutation on each allele of the gene and in three out of eight individuals one single mutation was detected

(two truncating mutations and one in frame deletion of 3 bp), and the direct sequencing of all RPGRIP1 exons and intron-exon boundaries failed to detect the second mutation.

Among the 142 unrelated LCA patients of our series, 49 were found to harbour mutations in one of the four already reported genes (34.5%) and eight in the RPGRIP1 gene (5.6%). Among the eight different mutations identified in this last gene, five out of eight are expected to result in the truncation of the protein. Although it was unlikely that these mutations were not disease-causing, we have studied 126 normal controls by DHPLC and evidenced that, conversely to the LCA patients harbouring these truncating mutations, the normal controls did not display abnormal DHPLC profiles. The three others (two missense mutations and one in frame 3 bp deletion) concern highly conserved amino acids in bovine and murine sequences and lie in highly conserved regions (two in the RPGR-interacting domain). Hence, the identification of null mutations and of mutations that involve highly conserved amino acids suggest that the interaction of RPGRIP1 with RPGR is severely altered in the retina of patients with LCA.

RPGRIP1 is a novel protein that has been shown to specifically interact with RPGR in bovine, mouse and human.^{13,14,18} Several alternatively spliced transcripts have been identified in the three species. Some transcripts have a retina-restricted expression and interact specifically with RPGR *in vivo* and *in vitro*.¹³ Sequence analyses of all isoforms revealed the presence of a highly conserved and predicted globular C-terminal region corresponding to the cognate RPGR-interacting domain (RID) and two coiled-coil domains at the N terminus.¹³ Database analyses of the translated products revealed that the coiled-coil domains have significant homology to counterpart domains of proteins involved in membrane and vesicular trafficking.¹³ Interestingly, it has been shown that, in mouse, RPGRIP1 co-localised with RPGR in the connecting cilia of both rod and cone photoreceptors. It has been proposed that RPGRIP1 serves as a scaffold to anchor regulatory complexes including RPGR within the connecting cilium. In addition, RPGRIP1 might be an essential factor so that RPGR can normally play its putative role in maintaining the polarised protein distribution across the connecting cilium by facilitating directional transport or restricting redistribution.¹⁸

In conclusion, we report here the identification of a fifth LCA gene involved in a physiological pathway different from the previously reported ones and resulting from a novel mechanism. Hopefully, the identification of the exact role of RPGRIP1 in photoreceptor will improve our understanding of this complex retinal pathology.

While this work was in the reviewing stage, Dryja *et al*²¹ reported similar findings. Indeed, the study of the 24 RPGRIP1 coding exons allowed the identification of mutations in four out of 57 LCA families. Among them, three harboured two null RPGRIP1 alleles and one was compound heterozygote for missense mutations.

Taken together, these two independent studies strongly support the involvement of the RPGRIP1 gene in about 6–7% of LCA cases.

Acknowledgments

This work has been supported by the Foundation Fighting Blindness, the Foundation for Retinal Research, the Association Retina France and the Fondation Liliante Bettencourt-Schueller.

Electronic-database information

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/OMIM/>; Human Genome Project Working Draft at USC, <http://genome.ucsc.edu/>; Entrez Protein Sequence Search, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>

References

- 1 Leber T: Leber Retinitis pigmentosa und angeborene Amaurose. *Albrecht von Graefes Arch Ophthalmol* 1869; **15**: 1–25.
- 2 Franceschetti A, Dieterle P: Importance diagnostique et pronostique de l'électrorétinogramme (ERG) dans les dégénérescences tapéto-rétiniennes avec rétrécissement du champ visuel et héméralopie. *Confin Neurol* 1954; **14**: 184–186.
- 3 Kaplan J, Bonneau D, Frezal J, Munnich A, Dufier JL: Clinical and genetic heterogeneity in retinitis pigmentosa. *Hum Genet* 1990; **85**: 635–642.
- 4 Perrault I, Rozet JM, Calvas P *et al*: Retinal-specific guanylate cyclase gene mutations in Leber's congenital amaurosis. *Nat Genet* 1996; **14**: 461–464.
- 5 Marlhens F, Bareil C, Griffioen JM *et al*: Mutations in RPE65 cause Leber's congenital amaurosis. *Nat Genet* 1997; **17**: 139–141.
- 6 Morimura H, Fishman GA, Grover SA, Fulton AB, Berson EL, Dryja TP: Mutations in the RPE65 gene in patients with autosomal recessive retinitis pigmentosa or Leber congenital amaurosis. *Proc Natl Acad Sci USA* 1998; **95**: 3088–3093.
- 7 Freund CL, Wang QL, Chen S *et al*: De novo mutations in the CRX homeobox gene associated with Leber congenital amaurosis. *Nat Genet* 1998; **18**: 311–312.
- 8 McInnes RR, Freund CL, Chen S *et al*: De novo mutations in the CRX homeobox gene associated with Leber congenital amaurosis (LCA). *Exp Eye Res* 1998; **67**: A81.
- 9 Swaroop A, Wang QL, Wu W *et al*: Leber congenital amaurosis caused by a homozygous mutation (R90W) in the homeodomain of the retinal transcription factor CRX: direct evidence for the involvement of CRX in the development of photoreceptor function. *Hum Mol Genet* 1999; **8**: 299–305.
- 10 Sohocki MM, Bowne SJ, Sullivan LS *et al*: Mutations in a novel photoreceptor-pineal gene on 17p cause Leber congenital amaurosis (LCA4). *Nature Genet* 2000; **24**: 79–83.
- 11 Stockton DW, Lewis RA, Abboud EB *et al*: A novel locus for Leber congenital amaurosis on chromosome 14q24. *Hum Genet* 1998; **103**: 328–333.
- 12 Dharmaraj S, Li Y, Robitaille JM *et al*: A novel locus for Leber congenital amaurosis maps to chromosome 6q. *Am J Hum Genet* 2000; **66**: 319–326.
- 13 Roepman R, Bernoud-Hubac N, Schick DE *et al*: The retinitis pigmentosa GTPase regulator (RPGR) interacts with novel transport-like proteins in the outer segments of rod photoreceptors. *Hum Mol Genet* 2000; **9**: 2095–2105.
- 14 Boylan JP, Wright AF: Identification of a novel protein interacting with RPGR. *Hum Mol Genet* 2000; **9**: 2085–2093.

- 15 Farjo Q, Jackson A, Pieke-Dahl S *et al*: Human bZIP transcription factor gene NRL: structure, genomic sequence, and fine linkage mapping at 14q11.2 and negative mutation analysis in patients with retinal degeneration. *Genomics* 1997; **45**: 395–401.
- 16 Camuzat A, Dollfus H, Rozet JM *et al*: A gene for Leber's congenital amaurosis maps to chromosome 17p. *Hum Mol Genet* 1995; **4**: 1447–1452.
- 17 Camuzat A, Rozet JM, Dollfus H *et al*: Evidence of genetic heterogeneity of Leber's congenital amaurosis (LCA) and mapping of LCA1 to chromosome 17p13. *Hum Genet* 1996; **97**: 798–801.
- 18 Hong DH, Yue G, Adamian M, Li T: A retinitis pigmentosa GTPase regulator (RPGR)–interacting protein is stably associated with the photoreceptor ciliary axoneme and anchors RPGR to the connecting cilium. *J Biol Chem* [epub ahead of print] 2000.
- 19 Perrault I, Rozet JM, Gerber S *et al*: Leber congenital amaurosis. *Molecular Genetics and Metabolism* 1999; **68**: 200–208.
- 20 Kaplan J, Rozet JM, Perrault I, Munnich A: Leber congenital amaurosis. The Metabolic and Molecular Bases of Inherited Disease, 8th Edition, 2000; **Chapter 237**: 5947–5954.
- 21 Dryja TP, Adams SM, Grimsby JL *et al*: Null rpgr1 alleles in patients with leber congenital amaurosis. *Am J Hum Genet* 2001; **68**: 1295–1298.