Mutations in a human homologue of *Drosophila crumbs* cause retinitis pigmentosa (RP12)

Anneke I. den Hollander[1], Jacoline B. ten Brink[2], Yvette J.M. de Kok[1], Simone van Soest[2], L. Ingeborgh van den Born[2,3], Marc A. van Driel[1], Dorien J.R. van de Pol[1], Annette M. Payne[4], Shomi S. Bhattacharya[4], Ulrich Kellner[5], Carel B. Hoyng[6], Andries Westerveld[7], Han G. Brunner[1], Elisabeth M. Bleeker-Wagemakers[2], August F. Deutman[6], John R. Heckenlively[8], Frans P.M. Cremers[1] & Arthur A.B. Bergen[2]

Retinitis pigmentosa (RP) comprises a clinically and genetically heterogeneous group of diseases that afflicts approximately 1.5 million people worldwide. Affected individuals suffer from a progressive degeneration of the photoreceptors, eventually resulting in severe visual impairment. To isolate candidate genes for chorioretinal diseases, we cloned cDNAs specifically or preferentially expressed in the human retina and the retinal pigment epithelium (RPE) through a novel suppression subtractive hybridization (SSH) method[1,2]. One of these cDNAs (*RET3C11*) mapped to chromosome 1q31–q32.1, a region harbouring a gene involved in a severe form of autosomal recessive RP characterized by a typical preservation of the para-arteriolar RPE (RP12; ref. 3). The full-length cDNA encodes an extracellular protein with 19 EGF-like domains, 3 laminin A G-like domains and a C-type lectin domain. This protein is homologous to the *Drosophila melanogaster* protein crumbs (CRB), and denoted CRB1 (crumbs homologue 1). In ten unrelated RP patients with preserved para-arteriolar RPE, we identified a homozygous *Alu*Y insertion disrupting the ORF, five homozygous missense mutations and four compound heterozygous mutations in *CRB1*. The similarity to CRB suggests a role for CRB1 in cell-cell interaction and possibly in the maintenance of cell polarity in the retina. The distinct RPE abnormalities observed in RP12 patients suggest that *CRB1* mutations trigger a novel mechanism of photoreceptor degeneration.

To isolate novel retinal disease genes, we constructed cDNA libraries enriched for retina- and RPE-specific genes using SSH (refs 1,2). We sequenced 440 *Rsa*I cDNA fragments and selected 107 cDNAs, 60 of which showed no database match, for expression profile analysis by a semi-quantitative RT-PCR. Thirty-three cDNAs were specifically or preferentially expressed in retina or RPE and were mapped in the human genome by radiation hybrid mapping[1]. The retina-specific cDNA clone *RET3C11* mapped to 1q31–q32.1 near marker SHGC-31318 in the Stanford G3 Radiation Hybrid map and was localized on YACs 772C10 and 755E11 in a 3-cM region harbouring the RP12 gene (Fig. 1*a*). RP12 is a clinically distinct and severe form of autosomal recessive RP. Affected individuals experience night blindness from early childhood and progressive visual field loss. On ophthalmoscopy and with fluorescein angiography, typical preservation of the RPE adjacent to and under the retinal arterioles is noted, whereas there is a general loss of RPE throughout the retina. Due to early macular involvement, patients experience severe visual impairment before the age of 20 years[4,5].

We obtained the full-length cDNA sequence by screening a retina cDNA library with the *RET3C11*

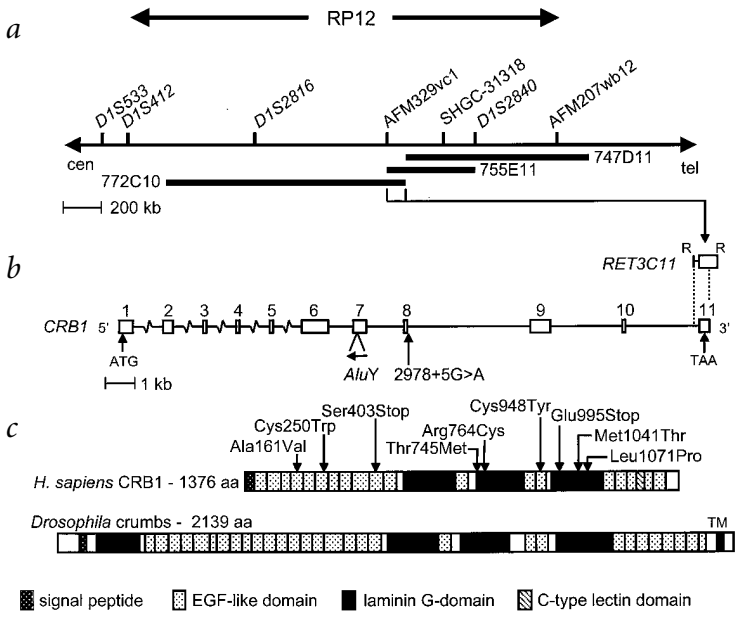

Fig. 1 Chromosomal map of the critical chromosomal region for RP12 and the structure of *CRB1* and protein product. ***a***, Localization of the RP12 gene in a 3-cM region between markers *D1S412* and AFM207wb12 (ref. 3). The *RET3C11* subtraction clone was mapped near marker SHGC-31318 by radiation hybrid mapping, and localized on YACs 772C10 and 755E11. ***b***, The intron/exon structure of *CRB1*. The size of the introns between exons 1 and 6 were not determined. The centromere/telomere orientation of *CRB1* is not known. The *RET3C11* subtraction clone represents an *Rsa*I (R) fragment of an unspliced RNA transcript. Mutation analysis in RP patients with preserved para-arteriolar RPE revealed an *Alu*Y insertion in exon 7 and a 5′ splice site mutation in intron 8. ***c***, The CRB1 protein structure and missense/nonsense mutations found in RP12 patients. The CRB1 protein is 35% identical and 55% similar to *Drosophila* CRB. TM, transmembrane region.

[1]*Department of Human Genetics, University Hospital Nijmegen, Geert Grooteplein 10, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.* [2]*The Netherlands Ophthalmic Research Institute, Amsterdam, The Netherlands.* [3]*The Rotterdam Eye Hospital, Rotterdam, The Netherlands.* [4]*Institute of Ophthalmology, University College London, London, UK.* [5]*Department of Ophthalmology, University Clinic Benjamin Franklin, Free University Berlin, Berlin, Germany.* [6]*Department of Ophthalmology, University Hospital Nijmegen, Nijmegen, the Netherlands.* [7]*Institute of Human Genetics, University of Amsterdam, Amsterdam, the Netherlands.* [8]*Department of Ophthalmology, University of California, Los Angeles, California, USA. Correspondence should be addressed to F.P.M.C. (e-mail: F.Cremers@antrg.azn.nl).*

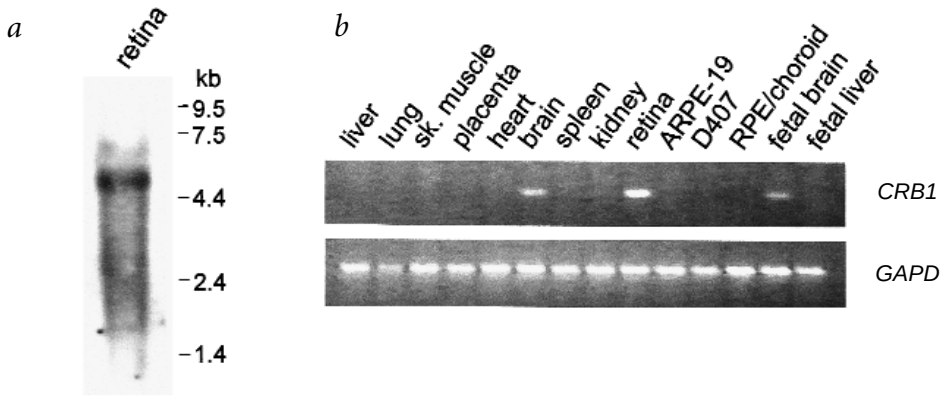

Fig. 2 Expression of *CRB1* in human tissues and RPE cell lines. ***a***, A northern blot containing poly(A)+ RNA from human retina was incubated with a *CRB1* cDNA fragment. The transcript size is approximately 5 kb. ***b***, RT-PCR on total RNA from ten human adult tissues, two fetal tissues and two RPE cell lines (ARPE-19, ref. 29; and D407, ref. 30). *CRB1* is expressed specifically in retina, adult brain and fetal brain. *GAPD* serves as a control.

subtraction clone and reiterated 5′-RACE. The consensus cDNA sequence consists of 4,361 bp and encodes a predicted protein of 1,376 amino acids (Fig. 1*c*). A 235-bp fragment of the cDNA was used as a probe in a northern-blot analysis, which revealed a 5-kb transcript in neural retina (Fig. 2*a*). Using RT-PCR, expression was detected in neural retina, brain and fetal brain, but not in RPE/choroid, two RPE cell lines, fetal liver, liver, lung, skeletal muscle, placenta, heart, spleen and kidney (Fig. 2*b*). The consensus cDNA sequence contains a small poly(A) tail but lacks an upstream polyadenylation site, suggesting that the 3′ end of the cDNA remains to be determined.

The predicted protein sequence contains a signal peptide[6], 19 EGF-like domains[7] contained in 4 clusters, 3 laminin A G-like or ALPS (agrin, laminin, perlecan, slit) domains[8,9] and a C-type lectin (CTL) domain[10] (Fig. 1*c*). The presence of EGF-like domains, a signal peptide and the absence of a transmembrane region suggests that it is an extracellular protein. Database searches with the predicted protein sequence revealed highest homology to *Drosophila* CRB (35% identity, 55% similarity). The protein was therefore denoted CRB homologue 1 (CRB1). CRB is larger (2,139 aa), has 30 EGF-like and 4 laminin G-like domains and is a transmembrane protein[11]. The typical arrangement of laminin G-like domains flanked by EGF-like domains, however, is conserved between CRB1 and CRB (Fig. 1*c*).

On the basis of its preferential expression in the retina and its co-localization with the RP12 locus, we considered *CRB1* a candidate gene for RP12. The intron/exon structure of *CRB1* was determined by ligation-mediated PCR (ref. 12) and exon-exon PCR. *CRB1* consists of 11 exons spanning at least 40 kb; the splice junctions follow the AG/GT rule. We designed 25 primer pairs for mutation analysis of the *CRB1* ORF and splice junctions (Table 1, see http://genetics.nature.com/supplementary_info/).

We diagnosed 15 unrelated RP patients with para-arteriolar preservation of the RPE. Of these, patient 22147 belongs to a large consanguineous pedigree previously investigated in the search for the RP12 locus[13]. The other families were too small for conclusive haplotype analysis, but patients of five additional families (24868, 25977, 25983, RP112 and RP0136) showed homozygosity for markers in the RP12 region. We carried out single-strand conformation polymorphism (SSCP) analysis and nucleotide sequencing using DNA from these 15 patients.

Mutations are summarized (Table 2) and indicated in *CRB1* (Fig. 1*b*) or the predicted protein (Fig. 1*c*). In patient 25977, a homozygous insertion of an *Alu* repeat DNA element was found in exon 7 at nt 2,320 (Fig. 3*a*). The *Alu* element belongs to the *AluY* subfamily[14], is orientated in the antisense direction, contains a more than 70-nt poly(A) tail and is flanked by a 12-bp direct repeat consisting of *CRB1* nt 2,309–2,320. This insertion is present heterozygously in the unrelated parents and two siblings of patient 25977 (Fig. 3*b*), but not in 185 healthy controls. The *Alu* sequence shows a 5′ truncation that was also observed in a pathologic *Alu* element inserted in *F9* (ref. 15). The *Alu* insertion disrupts the ORF, and most likely results in the inactivation of *CRB1*.

We identified a homozygous non-conservative Met1041Thr missense mutation in patient 22147 (Fig. 4*a*). As expected, the mutation co-segregates with the RP12 phenotype (Fig. 4*b*). The same mutation was found heterozygously in 1 of 100 healthy control individuals living in the same region as patient 22147. This is not unexpected, assuming an allele frequency for this mutation of 1 in 200 and considering the small size of the population in this region (~1.5 million).

Mutations in other patients include two stop mutations, a splice-site mutation and several missense mutations (Table 2 and Fig. 1). The 2978+5G→A 5′ splice-site mutation of exon 8 found in patient 25710 lowers the splice potential score[16] from 76.4 to 62.0, suggesting that the splice site is inactivated. The same patient and patient 25540 carry a 3′ splice-site mutation (2978G→A) in exon 9, which has a mild effect on the splice potential score (82.2 compared with the normal 85.6). This alteration introduces a Tyr→Cys change in the fourteenth EGF-like

Table 2 • *CRB1* mutations in RP patients with preserved para-arteriolar RPE

Patient	Allele 1		Allele 2	
	Mutation	Effect	Mutation	Effect
25983	617C→T	Ala161Val	617C→T	Ala161Val
RP112	885T→G	Cys250Trp	885T→G	Cys250Trp
24228	1343C→G	Ser403Stop	2425C→T	Arg764Cys
25977	2320insAlu	truncation	2320insAlu	truncation
24868	2369C→T	Thr745Met	2369C→T	Thr745Met
25540	2369C→T	Thr745Met	2978G→A	Cys948Tyr
26023	2425C→T	Arg764Cys	3118G→T	Glu995Stop
25710	2978+5G→A	splice defect	2978G→A	Cys948Tyr
22147	3257T→C	Met1041Thr	3257T→C	Met1041Thr
RP0136	3347T→C	Leu1071Pro	3347T→C	Leu1071Pro

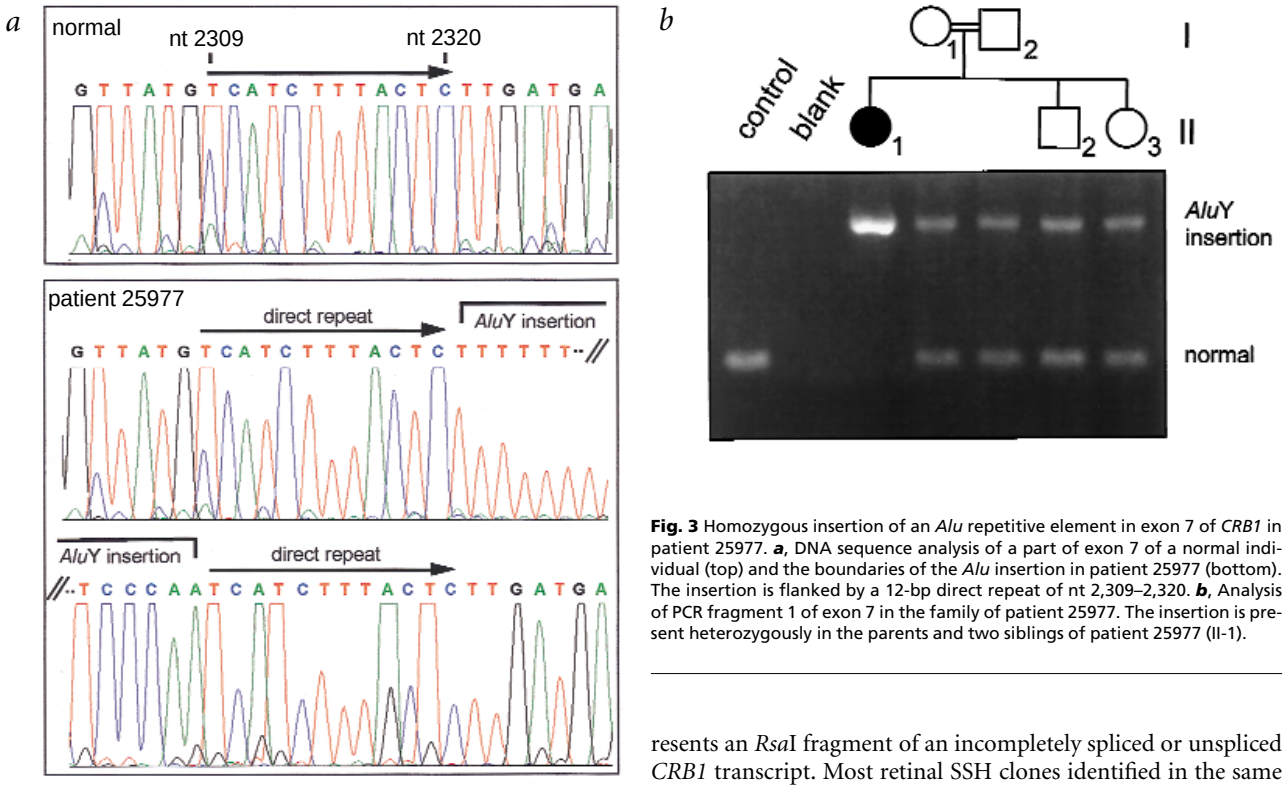

Fig. 3 Homozygous insertion of an *Alu* repetitive element in exon 7 of *CRB1* in patient 25977. ***a***, DNA sequence analysis of a part of exon 7 of a normal individual (top) and the boundaries of the *Alu* insertion in patient 25977 (bottom). The insertion is flanked by a 12-bp direct repeat of nt 2,309–2,320. ***b***, Analysis of PCR fragment 1 of exon 7 in the family of patient 25977. The insertion is present heterozygously in the parents and two siblings of patient 25977 (II-1).

domain of CRB1. In the EGF-like motif, Cys948 is a conserved residue and is probably used in protein folding by the formation of a disulfide bridge with Cys933 (ref. 17). Similarly, a conserved Cys in the sixth EGF-like domain is substituted homozygously by a Trp in a British Asian patient (RP112). Analysis of this mutation in the family of this patient revealed that two affected siblings are also homozygous for the mutation; the parents and three unaffected siblings are heterozygous (data not shown). The mutation was not found in 100 racially matched controls. The Thr745Met mutation found in patients 24868 and 25540 alters a threonine residue that is conserved in the last three laminin A G-like domains of *Drosophila* CRB and in all three G-like domains of CRB1. Of the remaining *CRB1* missense mutations, Arg764Cys is a non-conservative substitution, whereas Ala161Val and Leu1071Pro are conservative substitutions.

On SSCP or sequence analysis of the entire *CRB1* ORF, we identified 20 mutant alleles in 10 of 15 patients. SSCP analysis in 25 control individuals did not reveal any polymorphisms in the *CRB1* ORF. Except for the Met1041Thr mutation, none of the mutations were found in 100 healthy individuals, providing convincing evidence that *CRB1* is involved in RP12. No mutations have been detected so far in five patients. Sequence analysis of the *CRB1* promotor and of larger parts of the introns could identify additional mutations. It is also possible that these patients have been misdiagnosed or that this specific form of RP is genetically heterogenous.

To our knowledge, we are the first to identify a human disease gene using SSH. In contrast to conventional subtractive hybridization techniques previously used to isolate retina-specific genes, SSH allows the isolation of tissue-specific transcripts expressed at moderate or low levels in the tissue of interest by combining normalization and subtraction[1,2]. Screening of an amplified retina cDNA library with *RET3C11* revealed 16 cDNA clones of 1.2 million, suggesting that *CRB1* expression is relatively low (<0.01%). Comparison of the *RET3C11* subtraction clone with the *CRB1* cDNA and gene suggests that *RET3C11* represents an *Rsa*I fragment of an incompletely spliced or unspliced *CRB1* transcript. Most retinal SSH clones identified in the same procedure correspond to properly processed mRNA molecules (ref. 1, and A.I.d.H. and F.P.M.C., unpublished data). Most *Rsa*I fragments encountered in the SSH libraries are small *Rsa*I fragments (<500 bp) from the 3′ region of genes[1]. As SSH is a PCR-based technique, small *Rsa*I fragments will preferentially be amplified over large *Rsa*I fragments[1]. *CRB1* cDNA contains only 4 *Rsa*I sites at positions 439, 1,365, 1,470 and 3,429. The scarcity of *Rsa*I sites in *CRB1* cDNA has probably resulted in the enrichment of an incompletely spliced RNA carrying suitable *Rsa*I sites at its 3′ end.

The biochemical function of CRB1 and its role in RP pathogenesis remain to be elucidated. *Drosophila* CRB is involved in the organization of ectodermally derived epithelia and the establishment of polarity of epithelial cells, but mutations in *crb* also lead to abnormalities in the development of neural tissues[11,18]. CRB co-localizes with the zonula adherens, suggesting that it is involved in cell-cell interactions[19]. EGF-like domains, laminin A G-like domains and CTL domains function in protein-protein interactions. Many proteins essential for neuronal development, such as reelin, agrin, tenascin and proteins of the notch family, contain multiple EGF-like domains[20]. Besides CRB and CRB1, G-like domains are also associated with EGF-like domains in several other proteins, including neurexins[21], heparan sulfate proteoglycan core protein[22] and *Drosophila* perlecan, slit, agrin and fat proteins[8,11,23,24]. The combination of EGF-like domains and G domain-like repeats in CRB1 and its preferential expression in the retina suggest that CRB1 may be involved in neuronal development of the retina, presumably through protein-protein interactions. Its highest similarity to CRB suggests that it may have a role in the organization or polarity of retinal cells.

We have identified *CRB1* and demonstrated its involvement in RP12, a clinically distinct form of autosomal recessive RP. The processes underlying RP12 pathogenesis are distinct from those operating in other forms of autosomal recessive RP, for which most defects have been found in components of the phototransduction cascade or retinol metabolism[25,26], suggesting that CRB1 mutations trigger a novel mechanism of photoreceptor

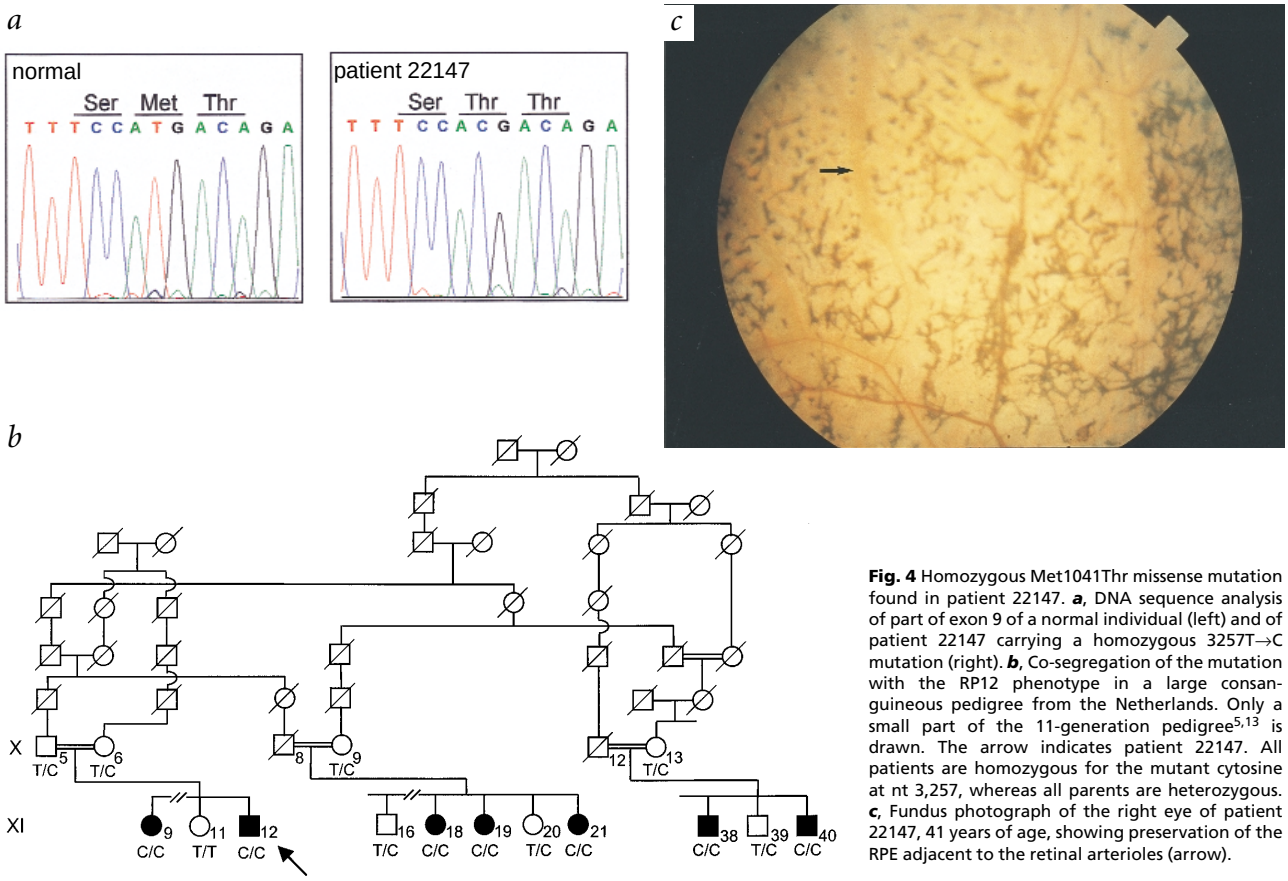

Fig. 4 Homozygous Met1041Thr missense mutation found in patient 22147. ***a***, DNA sequence analysis of part of exon 9 of a normal individual (left) and of patient 22147 carrying a homozygous 3257T→C mutation (right). ***b***, Co-segregation of the mutation with the RP12 phenotype in a large consanguineous pedigree from the Netherlands. Only a small part of the 11-generation pedigree[5,13] is drawn. The arrow indicates patient 22147. All patients are homozygous for the mutant cytosine at nt 3,257, whereas all parents are heterozygous. ***c***, Fundus photograph of the right eye of patient 22147, 41 years of age, showing preservation of the RPE adjacent to the retinal arterioles (arrow).

degeneration. The (sub)cellular localization of CRB1 and the identification of interacting proteins may shed light on the pathogenesis of RP12 and possibly on the aetiology of other forms of retinal degeneration.

Methods

cDNA cloning and sequencing. For cDNA library screening, we amplified a 155-bp PCR product from the *RET3C11* subtraction clone and labelled it with α-^{32}P-dCTP by random primer extension. We used the probe to screen 1.2×10^6 plaques of a human retina 5′-STRETCH cDNA library (Clontech) in λgt10 at 80,000 plaques/150-mm dish by standard protocols using Hybond-N$^+$ membranes (Amersham). Hybridizations were performed at 65 °C in 7% (w/v) SDS, PO_4 (0.5 M) and EDTA (1 mM). We purified 5 of 16 positive clones by secondary and tertiary screening using nitrocellulose filters and NC hybridization buffer (6×SSC, 5×Denhardt's, 0.2% (w/v) SDS, 10% (w/v) dextrane-sulphate). Phage DNA was isolated with the Lambda Miniprep Kit (Qiagen). Inserts were cloned in pBluescript (Stratagene) and sequenced with the Thermo Sequenase Dye Terminator Cycle Sequencing Pre-mix kit (Amersham) using an ABI 370A automated sequencer. To identify the 5′ end of the gene, we performed reiterated 5′-RACE using human fetal brain Marathon-Ready cDNA (Clontech). The first 5′-RACE was performed by a primary PCR using primer AP1 and a gene-specific primer at position 2,936–2,956, and a nested PCR using primer AP2 and a primer at position 2,896–2,918. We cloned the resulting PCR products, ranging from 400 to 1,500 bp, in a T/A vector (Invitrogen) and sequenced them. A second 5′-RACE was performed using primer AP1 and a primer at postion 1,689–1,712, and in a nested PCR primer AP2 and a primer at position 1,648–1,671. The resulting PCR products ranged from 300 to 1,700 bp. We confirmed the 5′ end of the cDNA by a third 5′-RACE using primer AP1 and a primer at position 382–404, and in a nested PCR primer AP2 and a primer at position 351–373. We carried out nucleic acid and protein database searches using the FASTA, TFASTA and BLAST programs[27,28] against the GenBank, EMBL and SWISS-PROT databases.

Northern-blot and RT-PCR analysis. For northern-blot analysis, we isolated total RNA from human retina by RNAzol B (Campro Scientific). We isolated retina poly(A)$^+$ RNA from total RNA with an mRNA purification kit (Pharmacia Biotech). Retina poly(A)$^+$ RNA (5 μg) was separated on a 1% agarose gel containing 1×MOPS and 18% (w/v) formaldehyde and transferred to a Hybond-N$^+$ membrane (Amersham). A cDNA fragment containing *CRB1* nt 1,604–1,838 was labelled with α-^{32}P-dCTP by random primer extension and hybridized to a multiple-tissue northern blot (Clontech) and the retina northern blot in ExpressHyb hybridization solution (Clontech) at 65 °C. Blots were washed twice in 0.1×SSC, 0.1% SDS at 65 °C for 20 min. We performed RT-PCR for 35 cycles on RNA from 10 human adult tissues, 2 fetal tissues and 2 RPE cell lines (ARPE-19, ref. 29; D407, ref. 30) as described[1] with primers amplifying *CRB1* nt 2,941–3,233.

Intron/exon structure. The intron/exon boundaries of exons 1–5 were determined by ligation-mediated (LM) PCR (ref. 12). Genomic or YAC DNA was digested with *Alu*I, *Bal*I, *Dra*I, *Eco*RV, *Hae*III, *Pvu*II, *Rsa*I, *Sca*I or *Stu*I before ligation of adaptors. We performed LM-PCR with an exon-specific primer and an adaptor primer. To determine the intron/exon boundaries of exons 6–11, we designed primers from the cDNA sequence and performed exon-exon PCR reactions on genomic DNA or on DNA from YAC 772C10 or 755E11 (ref. 3) using *Taq* DNA polymerase (BRL) or Ex-Taq (TaKaRa). LM-PCR products and exon-exon PCR products were separated on agarose gel, purified with the Qiaquick Gel Extraction kit (Qiagen) and (partially) sequenced as described above.

RP12 patients. The 15 unrelated RP patients with preserved para-arteriolar RPE were identified by L.I.v.d.B. (patients 22147 (ref. 5), 24228, 24868 and 25710), J.R.H. (patients 25976, 25977, 25978, 25979, 25980, 25982 and 25983 (ref. 4)), U.K. (patient 25540), M. van Schooneveld (patient 26023) or D. Bessant (patient RP112). We isolated genomic DNA of these patients from blood samples by standard procedures.

Mutation analysis. For mutation analysis of the *CRB1* ORF and splice junctions, we designed 25 primer pairs from exonic and intronic sequences that amplified genomic DNA fragments of 150–350 bp. Primer sequences and PCR conditions are available (Table 1, see http://genetics.nature.com/supplementary_info/). We performed mutation analysis for 15 unrelated RP patients by SSCP analysis and sequencing of PCR products. Exons 1, 2.1, 4 and 5 were screened by SSCP analysis and the remaining amplicons by sequence analysis. PCR reactions for SSCP were performed with α-^{32}P end-labelled primers or in the presence of α-^{32}P-dCTP. PCR products were denatured and electrophoresed on 5% polyacrylamide gels. Sequencing was performed on PCR products as above or using a SequiTherm cycle sequencing kit (Epicentre) and α-^{32}P-dCTP.

GenBank accession number. cDNA sequence of *CRB1*, AF154671.

Acknowledgements

We thank R. Roepman and H. Yntema for their help with northern-blot analysis and cDNA library screening; M. van Schooneveld and D. Bessant for ascertaining patients with RP12; and P. de Jong for his continuous interest in this project. This work was supported by The Foundation Fighting Blindness, Inc., USA and The British Retinitis Pigmentosa Society.

Received 7 June; accepted 10 August 1999.

1. den Hollander, A.I. *et al.* Isolation and mapping of novel candidate genes for retinal disorders using suppression subtractive hybridization. *Genomics* **58**, 240–249 (1999).
2. Diatchenko, L. *et al.* Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl Acad. Sci. USA* **93**, 6025–6030 (1996).
3. van Soest, S. *et al.* Integrated genetic and physical map of the 1q31–q32.1 region, encompassing the RP12 locus, the F13B and HF1 genes and the EEF1AL11 and RPL30 pseudogenes. *Cytogenet. Cell Genet.* **84**, 22–27 (1999).
4. Heckenlively, J.R. Preserved para-arteriole retinal pigment epithelium (PPRPE) in retinitis pigmentosa. *Br. J. Ophthalmol.* **66**, 26–30 (1982).
5. van den Born, L.I. *et al.* Autosomal recessive retinitis pigmentosa with preserved para-arteriolar retinal pigment epithelium. *Am. J. Ophthalmol.* **118**, 430–439 (1994).
6. von Heijne, G. Signal sequences. The limits of variation. *J. Mol. Biol.* **184**, 99–105 (1985).
7. Doolittle, R.F., Feng, D.F. & Johnson, M.S. Computer-based characterization of epidermal growth factor precursor. *Nature* **307**, 558–560 (1984).
8. Rothberg, J.M. & Artavanis-Tsakonas, S. Characterization of a conserved carboxy-terminal sequence in secreted proteins and a motif implicated in extracellular protein interactions. *J. Mol. Biol.* **227**, 367–370 (1992).
9. Sasaki, M., Kleinman, H.K., Huber, H., Deutzmann, R. & Yamada, Y. Laminin, a multidomain protein. *J. Biol. Chem.* **263**, 16536–16544 (1988).
10. Drickamer, K. Two distinct classes of carbohydrate-recognition domains in animal lectins. *J. Biol. Chem.* **263**, 9557–9560 (1988).
11. Tepass, U., Theres, C. & Knust, E. *crumbs* encodes an EGF-like protein expressed on apical membranes of Drosophila epithelial cells and required for organization of epithelia. *Cell* **61**, 787–799 (1990).
12. Kere, J. *et al*. Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts. *Genomics* **14**, 241–248 (1992).
13. van Soest, S. *et al.* Assignment of a gene for autosomal recessive retinitis pigmentosa (RP12) to chromosome 1q31-q32.1 in a inbred and genetically heterogeneous disease population. *Genomics* **22**, 499–504 (1994).
14. Batzer, M.A. *et al.* Standardized nomenclature for Alu repeats. *J. Mol. Evol.* **42**, 3–6 (1996).
15. Vidaud, D. *et al.* Haemophilia B due to a de novo insertion of a human-specific *Alu* subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* **1**, 30–36 (1993).
16. Shapiro, M.B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
17. Appella, E., Weber, I.T. & Blasi, F. Structure and function of epidermal growth factor-like regions in proteins. *FEBS Lett.* **231**, 1–4 (1988).
18. Knust, E. *et al.* EGF homologous sequences encoded in the genome of *Drosophila melanogaster*, and their relation to neurogenic genes. *EMBO J.* **6**, 761–766 (1987).
19. Wodarz, A., Hinz, U., Engelbert, M. & Knust, E. Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of Drosophila. *Cell* **82**, 67–76 (1995).
20. Nakayama, M. *et al*. Identification of high-molecular weight proteins with multiple EGF-like motifs by motif-trap screening. *Genomics* **51**, 27–34 (1998).
21. Ushkaryov, Y.A., Petrenko, A.G., Geppert, M. & Südhof, T.C. Neurexins: synaptic cell surface proteins related to the α-latrotoxin receptor and laminin. *Science* **257**, 50–56 (1992).
22. Kallunki, P. & Tryggvason, K. Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD protein containing multiple domains resembling elements of the low density lipoprotein receptor, laminin, neural cell adhesion molecules, and epidermal growth factor. *J. Cell Biol.* **116**, 559–571 (1992).
23. Joseph, D.R. & Baker, M.E. Sex hormone-binding globulin, androgen-binding protein, and vitamin K-dependent protein S are homologous to laminin A, merosin, and *Drosophila crumbs* protein. *FASEB J.* **6**, 2477–2481 (1992).
24. Patthy, L. A family of laminin-related proteins controlling ectodermal differentiation in *Drosophila*. *FEBS Lett.* **298**, 182–184 (1992).
25. Dryja, T.P. & Li, T. Molecular genetics of retinitis pigmentosa. *Hum. Mol. Genet.* **4**, 1739–1743 (1995).
26. van Soest, S., Westerveld, A., de Jong, P.T.V.M., Bleeker-Wagemakers, E.M. & Bergen, A.A.B. Retinitis pigmentosa: defined from a molecular point of view. *Surv. Ophthalmol.* **43**, 321–334 (1999).
27. Pearson, W.R. & Lipman, D.J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
28. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
29. Dunn, K.C., Aotaki-Keen, A.E., Putkey, F.R. & Hjelmeland, L.M. ARPE-19, a human retinal pigment epithelial cell line with differentiated properties. *Exp. Eye Res.* **62**, 155–169 (1996).
30. Davis, A.A. *et al*. A human retinal pigment epithelial cell line that retains epithelial characteristics after prolonged culture. *Invest. Ophthalmol. Vis. Sci.* **36**, 955–964 (1995).