Genomic structure and assessment of the retinally expressed RFamide-related peptide gene in dominant cystoid macular dystrophy

Heidi L. Schulz,¹ Heidi Stoehr,¹ Karen White,¹ Marc A. van Driel,² Carel B. Hoyng,³ Frans Cremers,² Bernhard H. F. Weber¹

¹Institut für Humangenetik, Universität Würzburg, D-97074 Würzburg, Germany; Departments of ²Human Genetics and ³Ophthal-mology, University Medical Center Nijmegen, 6500 Nijmegen HB, The Netherlands

Purpose: Computer-assisted sampling of EST data contained within the UniGene human sequences collection is being used to establish a catalog of novel genes that are expressed exclusively or predominantly in the human retina. This provides a valuable source for candidate genes possibly involved in retinal degeneration. In this report we present the characterization of the C7orf9 gene locus encoding RFamide-related peptides (RFRPs) and its evaluation in dominant cystoid macular dystrophy (CYMD).

Methods: Bioinformatics and cDNA library screening were used to isolate the full-length cDNA sequence and to determine the genomic organization of C7orf9. Expression profiling was done by RT-PCR and Northern blot analysis. C7orf9 was evaluated as a candidate gene for CYMD by DNA sequencing and Southern blot analysis in two affected individuals from an extended Dutch CYMD family.

Results: The C7orf9 cDNA transcript consists of 1190 bp and is organized into 3 exons on the short arm of chromosome 7 within the critical region for CYMD. The transcript is specifically expressed in the retina but not in a large range of other human tissues. No disease-causing mutations or larger gene rearrangements were found.

Conclusions: We provide the genomic organization of the RFamide-related peptide gene, C7orf9, which encodes a precursor protein for at least two small neuropeptides, referred to as NPSF (alias RFRP-1) and NPVF (alias RFRP-3) and show that it is abundantly expressed in the human retina. Results of our comprehensive mutation analysis suggests that C7orf9 is not the CYMD gene.

The human retina is a complex, multi-layered structure of neuroectoderm-derived cells that are highly differentiated to perform specific tasks in the detection of light and its integration and processing into visual signals. The specialized functions of these cells together with the various cell types providing their metabolic and structural support is dependent on the activity of a sophisticated network of retinally expressed genes. In an ongoing project, we are aiming to identify, characterize, and catalog the genes that are expressed exclusively or predominantly in the retinal tissue.

The retina is affected by a host of hereditary retinopathies, the underlying genes for many of which have not yet been identified (RetNet). A comprehensive gene catalog can be used as a source for candidate genes, selected by their chromosomal locations and potential functional properties. These genes can also be systematically evaluated as potential predisposing factors for complex retinopathies such as age-related macular degeneration (AMD). AMD is an incurable cause of visual impairment that is expected to affect more than 16 million people worldwide by the year 2030 [1]. Ultimately, a full characterization of the retinal transcriptome will serve as

Correspondence to: Bernhard H. F. Weber, Institut für Humangenetik, Biozentrum, Am-Hubland, D-97074 Würzburg, Germany; Phone: (49)931-888-4062; FAX: (49)931-888-4069; email: bweb@biozentrum.uni-wuerzburg.de

a valuable reference for basic research into retinal physiology, facilitating the study of proteins that are integral to retinal structure and function.

Our approach involves the expression profiling of retina ESTs retrieved from the UniGene Human Sequences Collection [2]. Its application has led to the identification of several retinal genes encoding for example a G protein-coupled receptor [3] or a novel member of the MAGUK protein family [4]. Here, we report the further characterization of a transcript, identified as a retina-specific EST cluster, Hs. 60473, which, by RT-PCR and Northern blot analysis, appears to be uniquely transcribed in the human retina. The gene maps to the same region of chromosome 7p as the gene for dominant cystoid macular dystrophy (CYMD). CYMD is an autosomal dominantly inherited condition causing the early onset of a cystic edema in the posterior pole of the eye [5]. We have performed comprehensive mutation analysis in CYMD patients and demonstrate that this gene is not responsible for the retinal phenotype in these individuals.

METHODS

Isolation of the full-length cDNA and assignment of chromosomal location: EST cluster Hs.60473 was retrieved from the UniGene Human Sequences Collection (release number 113, June 2000). A retina cDNA library constructed in the TriplEx2

vector was screened using an α^{32} P-dCTP-labeled 199 bp DNA fragment obtained by PCR of genomic DNA with primers A129F (5'-TCT GAG CCT AGA GGA TAC C-3') and A129R (5'-GAT CTC AGA GGC AGG TTG-3') designed from 5' EST sequences of cluster Hs.60473. Filters were hybridized overnight in 0.5 mM sodium phosphate buffer, pH 7.2, 7% SDS, 1 mM EDTA at 58 °C and washed for 20 min. in 2X SSPE/0.1% SDS, 1X SSPE/0.1% SDS and 0.5X SSPE/0.1% SDS at 58 °C. Plasmid DNA was isolated and sequenced with primer walking technology using an ABI 310 automated sequencer and the ABI PRISM Ready Reaction Sequencing Kit (Perkin Elmer, Norwalk, USA).

To extend the transcript to the initiation site, 5'-RACE was applied to total human retina RNA according to the supplier's instructions (Life Technologies, Rockville, USA). First strand cDNA synthesis was performed using the genespecific oligonucleotide primer A129R. A first round of PCR was performed with the gene-specific reverse primer A129R5 (5'-TGC TGT GAA GAT TGG AGA TC-3') followed by a nested PCR reaction with primer A129R4 (5'-AGC TTG AAG TGG CTA AAG TC-3'). The PCR products were directly sequenced as desribed above.

The chromosomal location of C7orf9 was determined by PCR-based screening of the NIGMS Human/Rodent Somatic Cell Hybrid Panel No. 2 (Coriell Cell Repositories, Camden, USA). Primer pair A129E1F (5'-ACA TTG GGC TGC ACA TAG-3') and A129E1R (5'-ATG CTA CTC ACA TTA GAG AGA TT-3') were used on the panel and amplified a 216 bp product.

Expression studies: RT-PCR analysis was performed as described previously [6] using gene specific primers A129F3 (5'-TGATCT CCA ATC TTC ACA GC-3') and A129R amplifying a 244 bp cDNA fragment. For reverse transcription, commercially available total RNAs from brain, fetal brain, spinal

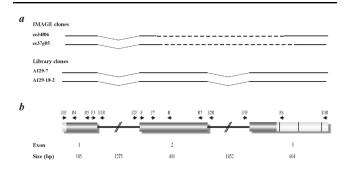


Figure 1. cDNA cloning and genomic structure of C7orf9. **A**: A consensus cDNA sequence was assembled from I.M.A.G.E. retina clones ze34f06 and ze37g05 and cDNA clones retrieved from cDNA library screenings. The dotted line indicates unknown sequence. **B**: C7orf9 comprises three coding exons which are represented by shaded boxes. The exon and intron sizes are derived from alignment of the cDNA to the corresponding human draft sequence. Light gray boxes represent the 5'-and 3'-untranslated regions (UTRs). Three potential poly(A)adenylation sites are depicted as vertical black bars in the 3'-UTR. The name and relative positions of oligonucleotide primers utilized in the study are shown.

cord, bone marrow, retina, kidney, liver, fetal liver, lung, trachea, spleen, thymus, colon, skeleton muscle, prostate, testis, uterus, placenta, adrenal gland, and salivary gland were used (Clontech, Heidelberg, Germany). To test for RNA integrity, primers G3PDH-ex8F (5'-ACC ACA GTC CAT GCC ATC AC-3') and G3PDH-ex9R (5'-TCC ACC ACC CTG TTG CTG TA-3') were used to amplify 452 bp of the ubiquitous gene glyceraldehyde 3-phosphate dehydrogenase. For Northern blot analysis, the 244 bp cDNA fragment was labeled with α^{32} PdCTP (3000 Ci/mmol) and hybridized to a Northern blot membrane containing 10 µg total RNA isolated from retina, heart, liver and cerebellum with the RNA-Clean kit (Angewandte Gentechnologie Systeme GmbH, Heidelberg, Germany). Hybridizations were carried out overnight in 0.5 mM sodium phosphate buffer, pH 7.2, 7% SDS, 1 mM EDTA at 65 °C. Washings were performed for 20 min. each in 2X SSPE/0.1% SDS, 1X SSPE/0.1% SDS and 0.5X SSPE/0.1% SDS at 65 °C. Autoradiography was done for 4 days.

Mutation analysis: Genomic DNA was obtained from

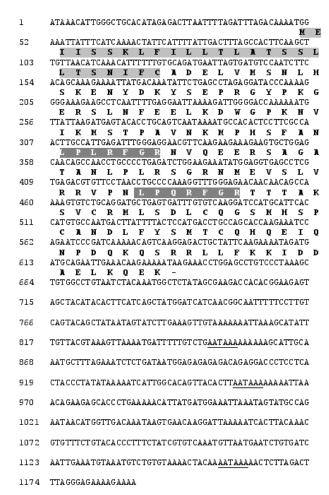
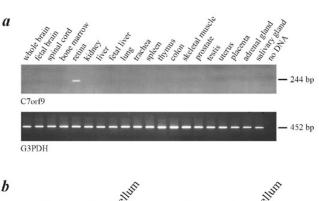


Figure 2. Full-length cDNA and hypothetical peptide sequence of C7orf9. A putative 26 amino acid signal peptide is highlighted in gray. The two characteristic motifs for RFamide peptides are depicted in dark gray. Three putative polyadenylation signals, AATAAA, are underlined.

two affected members (numbers 4716 and 10084) of a large Dutch CYMD family previously described by Pinckers et al. [7]. To analyse the coding region of C7orf9, the three exons were PCR-amplified using the following oligonucleotide primers: exon 1, A129E1F and A129E1R; exon 2, 129E2F (5'-TTA AGT TAA TTG GGG GTT TA-3') and 129E2R (5'-TAT AAC TGC CCA TGC ACT T-3'); exon 3, A129E3F (5'-GTT GTT CTG ATG GGT GAC-3') and A129R6 (5'-AAA TTG CCG TTG ATG ATC C-3'). The resulting PCR products were directly sequenced as described above.

For Southern blot analysis, $10 \mu g$ of DNA each was digested with restriction enzymes EcoRI, HincII, and PstI and electrophoretically fractionated on a 0.6% agarose-TBE gel. DNA was transferred to a HybondN+ membrane (Amersham Pharmacia Biotech, Freiburg, Germany) and hybridized overnight with a labeled probe obtained by amplification of retina cDNA with A129E1F and A129E3R (5'-TTT TCT TTT CTC CCT AAA GTC-3').

Assessment of the allele frequency of three single nucle-



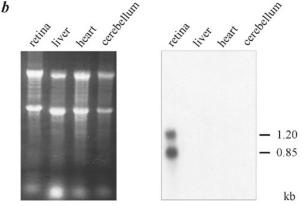


Figure 3. Expression analysis of C7orf9. **A**: RT-PCR analysis of C7orf9 was done using RNA from 20 different tissues. Primer pair A129F3 and A129R amplified a 244 bp product from only retina, as seen on the ethidium bromide-stained agarose gel. As a positive control, primers G3PDH-ex8F/G3PDH-ex9R amplified a 452 bp product from the ubiquitous gene glyceraldehyde-3-phosphate dehydrogenase from all tissues. **B**: RNA from different tissues was separated on a formaldehyde agarose gel and stained with ethidium bromide (left). The resulting Northern blot was probed with an α^{32} P-dCTP-labeled PCR product amplified by A129F3 and A129R. Two signals at approximately 1.2 kb and 0.85 kb were detected in retina, but not in the other tissues tested.

otide polymorphisms: Single-stranded conformational polymorphism analysis (SSCP) with primer pair A129E1F/A129E1R was performed to establish the allele frequency of 143G->C in 46 control subjects. The allele frequency of the alteration 172A->G was determined in 22 unaffected controls by direct DNA sequencing. The 431C->T alteration was analyzed by MnII restriction enzyme analysis of a 297 bp PCR fragment amplified from genomic DNA of 79 unrelated controls using primers A129F7 (5'-GTT ATT AAG ATG AGT ACA CCT GCA GTC AAT A-3') and A129R7 (5'-GTG CTG GCA GGT CAT GGA G-3'). All controls were of Caucasian ancestry.

Computational analysis: Nucleotide and protein BLAST programs at NCBI were used for sequence homology searches in public databases. Pattern and profile searches were performed with the Simple Modular Architecture Research Tool (SMART) at EMBL and the SignalP program.

RESULTS & DISCUSSION

Identification of the human RFamide-related peptide cDNA C7orf9: EST cluster Hs.60473 contains the non-overlapping 5' and 3' end sequences of two cDNA clones (ze34f06 and ze37g05) isolated from the Soares N2b4HR retina library (Figure 1A). Fourteen additional cDNA clones with inserts ranging from 0.5 to 1.2 kb were identified by screening a retina (TriplEx2 cDNA library; Figure 1A). Sequence assembly yielded a 1190 bp transcript, termed C7orf9 (GenBank accession number AF440392), with an open reading frame (ORF) of 591 bp and a potential start codon (ATG) located 48 bp downstream of the 5' end of the cDNA (Figure 2). Repeated applications of 5'-RACE and sequencing of the resulting cDNA fragments failed to reveal further upstream sequences suggesting a comprehensive coverage of the most 5' region of the transcript. Three putative polyadenylation signals were identified at positions 848 bp, 955 bp, and 1156 bp (Figure 2). The ORF of C7orf9 encodes a putative 196 amino acid peptide containing a 26 amino acid signal peptide and two LPLRFGR motifs characteristic of a large family of secreted neuropeptides (Figure 2) [8,9].

BLASTN searches with the full-length cDNA sequence revealed significant identity to two independently isolated human cDNA transcripts (GenBank accession numbers AB040290 and AF330057). These genes have previously been reported to encode a precursor protein for two [10] or possi-

TABLE 1. SINGLE NUCLEOTIDE POLYMORPHISM FREQUENCY OF THREE VARIANTS OBSERVED IN THE C7ORF9 GENE SEQUENCE

Nucleotide EXCHANGE*	LOCATION	Amino acid change	Allele Frequency
143G->C	Exon 1 Exon 1 Exon 2	M32I	G: 0.93 (N=92)
172A->G		D42G	A: 0.73 (N=22)
431C->T		P128P	C: 0.73 (N=158)

^{*}Numbering starts at the first base of the cDNA

bly three [11] RFamide-related peptides (RFRPs), referred to as NPSF, NPVF and RFRP-1 to -3, respectively. In a wide range of species, the RFRPs constitute a large family of neuropeptides and are known to exert a variety of functions such as neurotransmission, neuromodulation, cardioexcitation or control of muscle contraction [12-14]. Hinuma et al. [11] demonstrated that synthetic RFRP-1 and RFRP-3 are specific agonists of an orphan G protein-coupled receptor OT7T022 possibly mediating regulatory functions in prolactin secretion. Furthermore, endogenous RFRP-1 was purified from bovine hypothalamus and found to consist of 35 amino acid residues suggesting that the mature neuropeptide is generated by cleavage of the RFRP preprotein between residues Arg56 and Ser57 corresponding to residues Arg55 and Ser56 in human [15] (Figure 2). The existence of endogenous RFRP-2 and RFRP-3 remains to be shown.

Genomic structure and chromosomal localization of C7orf9: To determine the genomic structure of C7orf9, the cDNA sequence was aligned to the working draft sequence of BAC clone CTB-136N17 (GenBank accession number AC004129). This clone contains the entire cDNA sequence revealing three exons (GenBank accession numbers AF440393, AF440394, and AF440395) separated by two intervening sequences of 1275 bp and 1452 bp (Figure 1B). All

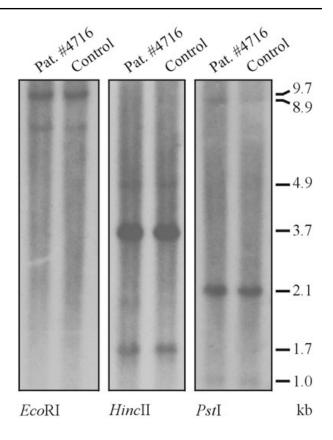


Figure 4. Southern blot analysis of C7orf9 locus. DNAs from patient number 4716 and an unrelated control were digested with restriction enzymes EcoRI, HincII, and PstI and probed with a cDNA fragment representing the full-length transcript including the 5' and 3' UTRs. No gross intragenic rearrangements are observed.

identified donor and acceptor splice site sequences conform to the GT-AG rule [16]. The three exons of the gene span approximately 4 kb of genomic DNA.

The chromosomal locus of C7orf9 was assigned by PCR-based screening of a monochromosomal human/rodent somatic cell hybrid DNA panel to human chromosome 7 (data not shown). Furthermore, BAC clone CTB-136N17 as well as overlapping PAC clone RP11-1151M13 (GenBank accession number AC083852) have previously been mapped to the short arm of chromosome 7 refining the genomic locus of C7orf9 to 7p15-p21. In addition, alignment of C7orf9 cDNA to genomic contig NT_007902 demonstrates that the C7orf9 locus maps approximately 64 kb proximal to cytochrom c (HCS) and 207 kb proximal to oxysterol binding protein-like 3 (OSBPL3).

Expression analysis: To examine the expression of C7orf9, RT-PCR analysis was performed with primers A129F3 and A129R amplifying a 244 bp transcript exclusively from human retina but not from the other 19 tissues tested (Figure 3A). Northern blot hybridization confirmed this pattern of expression (Figure 3B). Two bands of approximately 0.8 kb and 1.2 kb were observed, suggesting that two of the three hypothetical polyadenylation sites, those at 848 bp and 1156 bp, are used. In agreement with our findings, expression of RFRP mRNA was shown in rat tissues to be restricted to hypothalamus and eye [11].

Analysis of the C7orf9 gene as a candidate for CYMD: The retinal expression of C7orf9 as well as the localization of its genomic locus within the critical region for CYMD which is limited by D7S493 distally and D7S2444 proximally ([5] and unpublished data), makes this gene an excellent candidate for the CYMD gene. We therefore proceeded to analyze C7orf9 in two affected individuals (numbers 10084 and 4716) from the same large Dutch CYMD pedigree that was used to establish linkage [5].

We sequenced the complete coding region of the gene in both patients and an unrelated unaffected individual. The unaffected individual is heterozygous for a 172A->G alteration while patient number 10084 revealed heterozygosity for a 143G->C nucleotide substitution and individual number 4716 heterozygosity for a 431C->T alteration. The 143G->C and the 172A->G changes result in non-conservative amino acid substitutions from methionine at codon 32 to isoleucine and from aspartate at codon 42 to glycine, respectively. In contrast, the 431C->T variant does not alter the amino acid sequence of the putative protein. We conclude that neither of the nucleotide changes 143G->C and 431C->T is likely to be disease causing as both affected individuals belong to the same extended pedigree and would be expected to carry the same disease causing mutation. Moreover, in a group of unaffected controls, we have determined the minor allele frequency for 143G->C, 172A->G and 431C->T to be 0.07, 0.27 and 0.27, respectively, demonstrating their polymorphic nature (Table 1). Heterozygosity of both patients at the 143G->C and 431C->T sites confirms that both carry two C7orf9 alleles. This rules out the possibility of a heterozygous deletion involving the entire gene, a situation that would otherwise have remained undetected by PCR amplification. Further investigation of the C7orf9 gene locus in patient number 4716 by Southern blot analysis using restriction enzymes EcoRI, HincII and PstI also showed no evidence of gross intragenic rearrangements (Figure 4).

This work provides the characterization of the C7orf9 gene and its evaluation as a candidate for CYMD. The lack of mutations in the coding exons and lack of evidence for gross rearrangements argues against a causative role for C7orf9 mutations in the etiology of CYMD. It should be noted, however, that small mutations in the promoter sequences or within the intronic regions are difficult to assess and have not been excluded in this study. The knowledge of the gene structure of C7orf9 will facilitate further mutational analyses in other hereditary retinal dystrophies with as yet unknown genetic causes.

ACKNOWLEDGEMENTS

The authors would like to thank Andrea Gehrig for the construction of the retina TriplEx2 cDNA library and Andrea Rivera for her valuable technical assistance. This work was supported by a grant from the Bundesministerium für Bildung und Forschung (BMBF) under 01KW9921 and the following Dutch blindness foundations: Algemene Nederlandse Vereniging ter Voorkoming van Blindheid, Gelderse Blindenvereniging, Landelijke Stichting Blinden en Slechtzienden, Rotterdamse Vereninging Blindenbelangen, Stichting Blindenhulp, and Stichting voor Ooglijders.

REFERENCES

- Evans J, Wormald R. Is the incidence of registrable age-related macular degeneration increasing? Br J Ophthalmol 1996; 80:9-14.
- Stoehr H, Mah N, Schulz HL, Gehrig A, Frohlich S, Weber BH. EST mining of the UniGene dataset to identify retina-specific genes. Cytogenet Cell Genet 2000; 91:267-77.
- Sauer CG, White K, Stoehr H, Grimm T, Hutchinson A, Bernstein PS, Lewis RA, Simonelli F, Pauleikhoff D, Allikmets R, Weber BH. Evaluation of the G protein coupled receptor-75 (GPR75) in age related macular degeneration. Br J Ophthalmol 2001; 85:969-75.

- Stoehr H, Weber BH. Cloning and characterization of the human retina-specific gene MPP4, a novel member of the p55 subfamily of MAGUK proteins. Genomics 2001; 74:377-84.
- Kremer H, Pinckers A, van den Helm B, Deutman AF, Ropers HH, Mariman EC. Localization of the gene for dominant cystoid macular dystrophy on chromosome 7p. Hum Mol Genet 1994; 3:299-302.
- Mah N, Stoehr H, Schulz HL, White K, Weber BH. Identification
 of a novel retina-specific gene located in a subtelomeric region
 with polymorphic distribution among multiple human chromosomes. Biochim Biophys Acta 2001; 1522:167-74.
- Pinckers A, Deutman AF, Lion F, Aan de Kerk AL. Dominant cystoid macular dystrophy (CYMD). Ophthalmic Paediatr Genet 1983; 3:157-67.
- 8. Price DA, Greenberg MJ. Structure of a molluscan cardioexcitatory neuropeptide. Science 1977; 197:670-1.
- Dockray GJ, Sault C, Holmes S. Antibodies to FMRF amide, and the related pentapeptide LPLRF amide, reveal two groups of immunoreactive peptides in chicken brain. Regul Pept 1986; 16:27-37
- 10. Liu Q, Guan XM, Martin WJ, McDonald TP, Clements MK, Jiang Q, Zeng Z, Jacobson M, Williams DL Jr, Yu H, Bomford D, Figueroa D, Mallee J, Wang R, Evans J, Gould R, Austin CP. Identification and characterization of novel mammalian neuropeptide FF-like peptides that attenuate morphine-induced antinociception. J Biol Chem 2001; 276:36961-9.
- 11. Hinuma S, Shintani Y, Fukusumi S, Iijima N, Matsumoto Y, Hosoya M, Fujii R, Watanabe T, Kikuchi K, Terao Y, Yano T, Yamamoto T, Kawamata Y, Habata Y, Asada M, Kitada C, Kurokawa T, Onda H, Nishimura O, Tanaka M, Ibata Y, Fujino M. New neuropeptides containing carboxy-terminal RFamide and their receptor in mammals. Nat Cell Biol 2000; 2:703-8.
- 12. Raffa RB. The action of FMRFamide (Phe-Met-Arg-Phe-NH2) and related peptides on mammals. Peptides 1988; 9:915-22.
- 13. Greenberg MJ, Price DA. Relationships among the FMRFamide-like peptides. Prog Brain Res 1992; 92:25-37.
- Li C, Nelson LS, Kim K, Nathoo A, Hart AC. Neuropeptide gene families in the nematode Caenorhabditis elegans. Ann N Y Acad Sci 1999; 897:239-52
- 15. Fukusumi S, Habata Y, Yoshida H, Iijima N, Kawamata Y, Hosoya M, Fujii R, Hinuma S, Kitada C, Shintani Y, Suenaga M, Onda H, Nishimura O, Tanaka M, Ibata Y, Fujino M. Characteristics and distribution of endogenous RFamide-related peptide-1. Biochim Biophys Acta 2001; 1540:221-32.
- 16. Penotti FE. Human pre-mRNA splicing signals. J Theor Biol 1991; 150:385-420.

The print version of this article was created on 9 April 2002. 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.