European Journal of Human Genetics (2000) 8, 381–384

www.nature.com/ejhg

SHORT REPORT

Positional cloning and characterisation of the human *DLGAP2* gene and its exclusion in progressive epilepsy with mental retardation

Susanna Ranta[1,2], Yonghui Zhang[3], Barbara Ross[3], Elina Takkunen[1,2], Aune Hirvasniemi[4], Albert de la Chapelle[1,5], T Conrad Gilliam[3] and Anna-Elina Lehesjoki[1,2]

[1]*Folkhälsan Institute of Genetics, Helsinki;* [2]*Department of Medical Genetics, University of Helsinki, Finland;* [3]*Columbia Genome Center, Columbia University, New York, USA;* [4]*Department of Pediatrics, Kainuu Central Hospital, Kajaani, Finland;* [5]*Comprehensive Cancer Center, Ohio State University, Columbus, OH, USA*

In search of the gene for progressive epilepsy with mental retardation (EPMR) we identified *DLGAP2*, the human homolog of the gene encoding the rat PSD-95/SAP90-associated protein-2 (*Dlgap2*). We extended the transcript in both the 5' and 3' directions and characterised the genomic structure of the approximately 10 kb gene. Sequence comparisons of human *DLGAP2* cDNA sequences obtained from human testis and brain cDNA libraries with homologous rat genes suggest alternative splicing in the 5' end of the gene. The 5' coding sequence of the testis cDNA is complete, whereas based on homology with the rat gene 103 bp of coding sequence may still be missing in the 5' end of the *DLGAP2* brain transcript. *DLGAP2* was excluded as the gene responsible for EPMR. *European Journal of Human Genetics* (2000) **8**, 381–384.

Keywords: human *DLGAP2* gene; exon–intron structure; progressive epilepsy with mental retardation; neuronal ceroid lipofuscinosis

Progressive epilepsy with mental retardation (EPMR) is an autosomal recessive disorder characterised by normal early development and childhood onset epilepsy followed by mental retardation.[1] Recently, EPMR was recognised as a new subtype of neuronal ceroid lipofuscinoses (NCLs).[2]

We previously assigned the *EPMR* locus to an approximately 700 kb interval on chromosome 8p (Figure 1),[3–5] and constructed a 3412 bp partial cDNA sequence (Genbank AF009204) mapping to this region.[5] Between nucleotides 1259 and 2956 our cDNA sequence was identical to a partial cDNA sequence encoding a human discs large (hDLG) and a postsynaptic density protein 95 kDa (PSD-95)-associated protein-2, DAP2[6] (GenBank AB000275), except for a G to A difference at position 1741 and 42 nucleotides downstream of position 1867 which was present in DAP2 but absent in our sequence. Our cDNA also showed homology to a rat gene (GenBank U67138) encoding the PSD-95/synapse associated protein 90 (SAP90)-associated mRNA-2, *Sapap2*.[7] According to the HUGO Nomenclature Committee, the DAP/SAPAP/GKAP family of discs large associated proteins is denoted as DLGAP. In this paper we refer to the human homolog of the rat *Sapap2* (*Dlgap2*) gene as *DLGAP2*. The DLGAP proteins interact with the PSD-95/SAP90 protein, which in turn interacts with shaker type potassium channels and other molecules and clusters with these at synaptic junctions.[6–8] PSD-95/SAP90 may play an important role in the organisation of ion channels and signalling molecules in the synaptic junction. Furthermore, as mutations in two potassium channel genes were recently identified underlying human epilepsy,[9–11] we considered *DLGAP2* a good positional candidate for EPMR.

The 3412 bp *DLGAP2* sequence was first extended at the 3' end by 1.4 kb using rapid extension of cDNA ends (RACE) (primer 5'-GTG GCC TGG CTC ACA CTT GGC TCT-3'; Clontech Marathon Ready Human Fetal Brain cDNA Library). Secondly, a brain specific EST[12] (A004F35) was found to be part of the 3' untranslated region of *DLGAP2*. This EST was

Correspondence: Susanna Ranta, Folkhälsan Institute of Genetics, Mannerheimintie 97, 00280 Helsinki, Finland. Tel: +358 9 61585558; Fax: +358 9 61585632; E-mail: susanna.ranta@helsinki.fi
Received 4 June 1999; revised 12 October 1999; accepted 19 October 1999

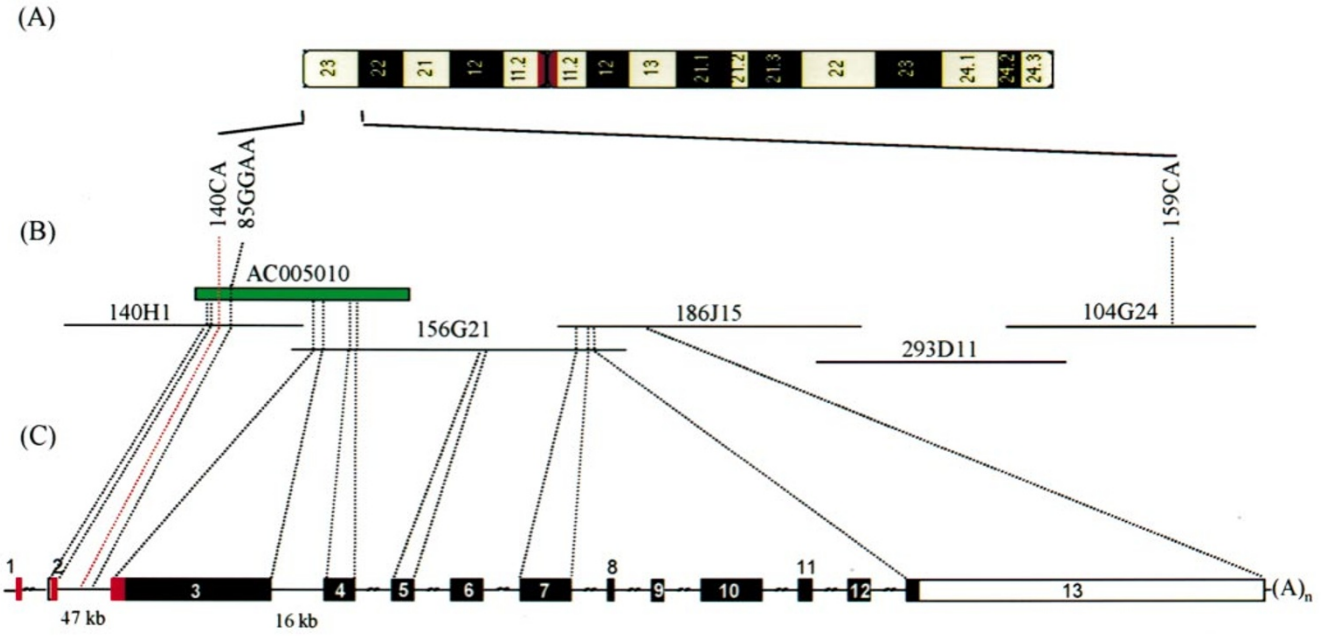

Figure 1 Positional cloning of *DLGAP2* (**A**) The *EPMR* locus is located on chromosome 8p23 close to the telomere. (**B**) The *EPMR* critical region flanked by markers 140CA and 159CA is covered by five overlapping BAC clones. The green rectangle represents the sequence derived from GenBank (AC005010). (**C**) Genomic structure of *DLGAP2*. In brain, exons 2–12 as well as the region of exon 13 marked with black is contained in the open reading frame. The cDNA sequence 5' to exon 2 is not known in brain, but resides outside the critical region. The coding region of the testis cDNA is marked in black. The 5' untranslated region of the testic cDNA is marked in red. Dotted lines indicate the location of the exons in BACs. Broken intronic lines indicate unknown intron sizes. The exon sizes are drawn approximately to scale except for the first and last exons. Exon 8 is not present in the cDNA clones identified by us.

extended by screening a brain cDNA library (Clontech cat. no. H13002a), by 5' RACE (primer 5'-TGG AAG GTC GGG GGA TGA GTT G-3'; Clontech Marathon Ready Human Brain cDNA Library), and by incorporating sequences derived from three genomic clones from the EPMR region yielding a 3.2 kb sequence around this EST.

Telomeric to the 3412 bp sequence one exon obtained by exon amplification from BAC 140H1 (Figure 1)[5] was found homologous to the 5' coding region of the rat *Dlgap2* (GenBank U67138). To confirm that this exon belonged to *DLGAP2* we screened the Marathon Ready Brain cDNA Library with primers designed from the exon and *DLGAP2* (5'-AGA AGC TGG AGA CTT GGT CCA G-3' and 5'CTC AGC AGG TAG GGC GGC CG-3', nested primers 5'-AGC CGG GCA TCA GCT TTC-3' and 5'-TCG GGC CCG TGG TGC AGG T-3'). A 214 bp fragment bridging the exon to the known cDNA sequence was amplified. Further sequence comparison with the rat *Dlgap2* sequence suggested that 103 bp of the 5' coding sequence was still missing from *DLGAP2*. Of note, this putative missing region was found to reside outside the critical EPMR region (see below and Figure 1).

We attempted to isolate the remaining 5' end by screening various RACE and cDNA libraries. Amplification of the Marathon Ready Human Testis cDNA library (Clontech; primer 5'AGT AGC GCT CCT CAT TGA AGT GCT GC-3'; nested primer 5'-TAC TGC GGG TCT AGA TCC TCC TCT G-3') resulted in a cDNA fragment containing a putative translation initiator ATG codon and an in-frame stop codon 5' to it. This ATG observed in the testis cDNA corresponds to the suggested start codon in the four known rat *Dlgap* genes. However, in the rat *Dlgap2* this codon is the second ATG in the open reading frame.[7] As in the human *DLGAP2* the sequence derived from a brain library the sequence also remains open 5' to the suggested initiator codon, and is homologous to rat *Dlgap2*, it remains possible that in brain the translation of *DLGAP2* is initiated at the earlier ATG than in other tissues. However, despite several attempts we were not able to isolate this longer brain *DLGAP2* transcript.

The total length of *DLGAP2* cDNA sequence assembled is approximately 10 kb, 2928 bp of which is open reading frame in testis and 3062 bp in brain transcripts (Genbank AF119817 and AF119818). By northern analysis (MTN-1; Clontech) *DLGAP2* was found to be expressed mainly in brain (Figure 2).

We established the genomic structure of *DLGAP2* by sequencing genomic clones obtained from a plasmid library of BAC 156G21 using part of *DLGAP2* as a probe, and from

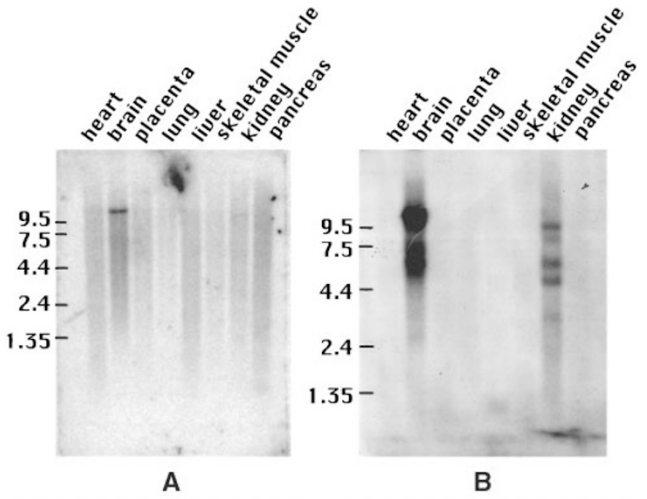

Figure 2 The tissue expression pattern of *DLGAP2* **A** Only a 10 kb brain-specific transcript is detected using a probe from the 3' untranslated region (nt 8059–8626; GenBank AF119817). **B** Using a probe from the coding region (nt 753–1790; GenBank AF119817) additional smaller brain transcripts, as well kidney transcripts of relatively weak intensity are detected suggesting either alternatively spliced forms of *DLGAP2* or hybridisation to homologous *DLGAP* genes.

sequence data derived from the sequencing effort of the EPMR region (Columbia University Genome Center, New York) as well as from a 120 kb genomic sequence present in the database (GenBank AC005010). The sequences of the exon–intron boundaries are shown in Table 1. We found 12 coding exons that corresponded to the brain transcript (exons 2–13, Figure 1). The putative 103 bp of coding sequence as well as the 5' untranslated sequence are still missing from the brain cDNA sequence (see above). Exons 2 and 3 are alternatively spliced in testis, where the last 29 bp of exon 2 and the first 68 bp of exon 3 form the 5' untranslated region in addition to the 22 bp of sequence encoded by exon 1 (Figure 1).

We then proceeded to screen the known coding regions and their immediately flanking intronic sequences in *DLGAP2* for mutations by sequencing amplified genomic DNA from two patients and one control individual. The primers used for sequencing are shown in Table 2. A homozygous silent substitution 1928G > A (both GA<u>G</u> and GA<u>A</u> coding for Glu) was found in exon 6 in the two patients. As an EPMR carrier was also found homozygous for this change it was considered a polymorphism. No other sequence differences between patients and the control were identified.

Based on sequence information obtained from GenBank (AC005010), marker 140CA is located between exons 2 and 3 of *DLGAP2* (see Figure 1). The region telomeric to 140CA has been previously excluded from harbouring the EPMR mutation on the basis of a recombination event that has occurred with 140CA and markers telomeric to it in one EPMR family.[5] Using a novel tetranucleotide marker, 85GGAA (GenBank

Table 2 Primers used to amplify the coding regions for mutation detection. Primers for exon 2 amplify both exon 2a and 2b sequences (see Table 1)

No.	*Forward primer 5'–3'*	*Reverse primer 5'–3'*
exon2	TGAAGATGTGCAGGGAATGA	GCTAACGTGTGTTTGTGGGA
exon3	CCAGCCTGGGCGACAGAG	CGACAATACCACCTGCTGTG
exon4	TCCAAAAAGGAGCTGATGCT	CGCAGGCAGTGGGAAAGT
exon5	GTGTGGGTTGGATGGTCATT	AGACTGGACCCCAGGGAG
exon6	GTAACGTGATGGTGACCCTG	TATGCCTCTAGAGTCCCCGC
exon7	AGTAGACCACAGGCTGACGG	TGCTGGGGTTACAGTCAGTG
exon8	CAAAATAGTCCCTTGCCCAG	AAGGACAGGCATGATTGAGG
exon9	TCCTCTCAGAAGGGCTACCA	CACTTGAATACCAAAGGGAGGA
exon10	GCTCCCTTGGTGTGATGTTT	AATGATGCCCATTAGCTTGG
exon11	CTCTCCTGGCTCTGAACACC	AGTGAGAAAGGGGGTTTGCT
exon12	AGCTCGCGAGCTGCTGG	GCCACAACCGTCCCCTCG

Table 1 Exon–intron boundaries of the known *DLGAP2* gene sequence. The exon and intron sequences are shown in upper case and lower case letters, respectively. The last two columns indicate the nucleotide positions of the exons in the *DLGAP2* brain and testis cDNA sequences

No.	*Intron–Exon boundary sequence*	*Size (bp)*	*Position in brain cDNA*	*Position in testis cDNA*
1	TTCAGCGGCCTGAATTCTAGAA[a]	22		1–22
2a	gtctttgc<u>ag</u>AGGAAGAAGC....GAGGATCTAG<u>gt</u>agagtaca	66	1–66	
2b	cgggcatc<u>ag</u>CTTTCCGGGG....GAGGATCTAG<u>gt</u>agagtaca	29		23–51
3	ttccccac<u>ag</u>ACCCGCAGTA....CTACCTCCAG<u>gt</u>aagcaggc	1058	67–1124	52–1109
4	ctgctccc<u>ag</u>GTACCTCAGG....AGCACCAGAC<u>gt</u>aagtgaga	212	1125–1336	1110–1321
5	tttcctgt<u>ag</u>CCAGACCTAC....CGTGAGCCAG<u>gt</u>cagggtcc	148	1137–1484	1322–1469
6	atgattgc<u>ag</u>GTGAGCGAGG....TCAACAGCAG<u>gt</u>aaggggac	220	1485–1704	1470–1689
7	tttcttac<u>ag</u>CTGTCTCATA....CGGGATTCAG<u>gt</u>agctgctc	350	1705–2054	1690–2039
8	ttttgttt<u>ag</u>GATTCTGAAT....AAGGTCAGAT<u>gt</u>aagtaccg	42	2055–2096	2040–2081
9	gttgaatt<u>ag</u>GTGGAAACGG....ATGAGAAGCG<u>gt</u>aactcagc	86	2097–2182	2082–2167
10	ccttttgc<u>ag</u>ACACGGACGT....TCGGAGGAAA<u>gt</u>aagagctc	416	2183–2598	2168–2583
11	ttttaaac<u>ag</u>TTCTCGGTAA....ACAGAATATG<u>gt</u>aagtgaat	92	2599–2690	2584–2675
12	gtgtcccc<u>ag</u>GACCCCAGCG....GGAAAGAAAG<u>gt</u>aagggcat	153	2691–2843	2676–2828
13	tgcttttc<u>ag</u>GAAGAAAGAA....TATTTAGTTM–(A)n	7290	2844–10133	2829–10118

[a]The exon–intron boundaries of this non-coding exon were not determined.

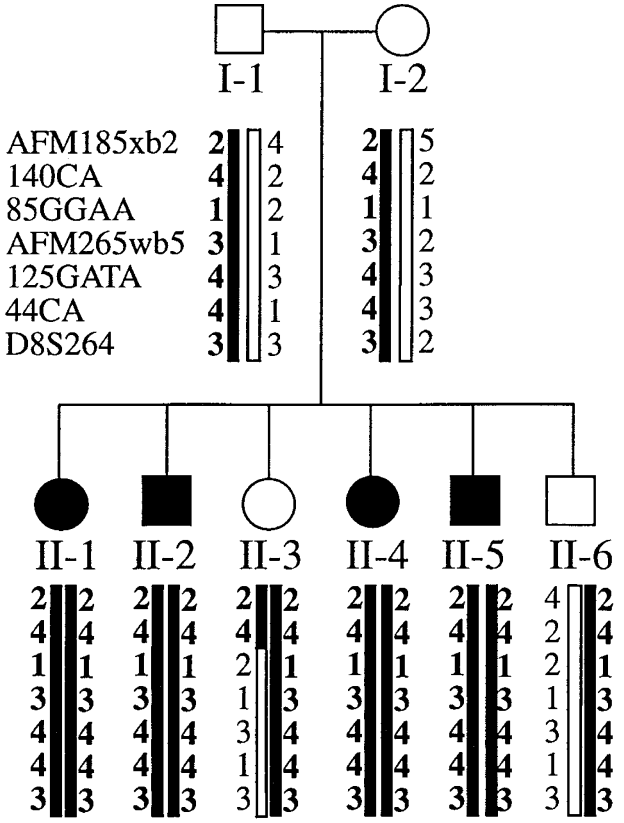

Figure 3 Pedigree of an EPMR family showing a recombination event refining the *EPMR* critical region telomeric to 140CA. The disease-associated haplotypes are shown with filled boxes. The recombination breakpoint is seen between markers 140CA and 85GGAA in the paternal meiosis in individual II-3.

acc. no AC005010; primers 5'-GGATGC AAG GAT GGA AG-G AT-3' and 5'-CCA TCC ATC TTT AAG ACC CA-3') within the same intron as 140CA, some 7 kb centromeric to it (Figure 1), this recombination event was further refined between markers 140CA and 85GGAA (Figure 3). Thus, based on the localisation of markers 140CA and 85GGAA in relation to *DLGAP2* (Figure 1), the region 5' to exon 3 of the *DLGAP2* cDNA was excluded from containing the EPMR mutation.

In this study we describe the characterisation of a novel gene, *DLGAP2* localised in the EPMR critical region. Our results provide evidence that *DLGAP2* is not the gene underlying EPMR, although as a brain gene interacting with ion channels through the PSD-95 protein *DLGAP2* was initially considered a reasonable positional and functional candidate.

Acknowledgements

SR and YZ contributed equally to this work. Supported by the Academy of Finland, the Sigrid Juselius Foundation, the Ulla Hjelt Foundation, and the Finnish State Grant TYH8310.

References

1 Hirvasniemi A, Lang H, Lehesjoki A-E, Leisti J: Northern epilepsy syndrome: an inherited childhood onset epilepsy with associated mental deterioration. *J Med Genet* 1994; **31**: 177–182.
2 Haltia M, Tyynelä J, Hirvasniemi A, Herva R, Ranta S, Lehesjoki A-E: CLN8: northern epilepsy. In: Goebel H, Mole SE, Lake BD (eds). *The Neuronal Ceroid Lipofuscinoses (Batten's Disease)*. IOS Press: Amsterdam, 1999; pp117–124.
3 Tahvanainen E, Ranta S, Hirvasniemi A *et al*: The gene for a recessively inherited human childhood progressive epilepsy with mental retardation maps to the distal short arm of chromosome 8. *Proc Natl Acad Sci USA* 1994; **91**: 7267–7270.
4 Ranta S, Lehesjoki A-E, Hirvasniemi A *et al*: Genetic and physical mapping of the progressive epilepsy with mental retardation (EPMR) locus on chromosome 8p. *Genome Res* 1996; **6**: 351–360.
5 Ranta S, Lehesjoki A-E, Bonaldo MF *et al*: High-resolution mapping and transcript identification at the progressive epilepsy with mental retardation locus on chromosome 8p. *Genome Res* 1997; **7**: 887–896.
6 Satoh K, Yanai H, Senda T *et al*: DAP-1, a novel protein that interacts with the guanylate kinase-like domains of hDLG and PSD-95. *Genes Cells* 1997; **2**: 415–424.
7 Takeuchi M, Hata Y, Hirao K, Toyoda A, Irie M, Takai Y: SAPAPS. A family of PSD-95/SAP90-associated proteins localized at post-synaptic density. *J Biol Chem* 1997; **272**: 11943–11951.
8 Kim E, Naisbitt S, Hsueh Y-P *et al*: GKAP, a novel synaptic protein that interacts with the gyanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. *J Cell Biol* 1997; **136**: 669–678.
9 Bievert C, Schroeder BC, Kubisch C *et al*: A potassium channel mutation in neonatal human epilepsy. *Science* 1998; **279**: 403–406.
10 Charlier C, Singh NA, Ryan SG *et al*: A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. *Nat Genet* 1998; **18**: 53–55.
11 Singh NA, Charlier C, Stauffer D *et al*: A novel potassium channel gene, *KCNQ2*, is mutated in an inherited epilepsy of newborns. *Nat Genet* 1998; **18**: 25–29.
12 Schuler GD, Boguski MS, Stewart EA *et al*: A gene map of the human genome. *Science* 1996; **274**: 540–546.