Hum Genet (1998) 103:355–364

ORIGINAL INVESTIGATION

Ming-Hui Wei · Irina Karavanova · Sergey V. Ivanov
Nicolae C. Popescu · Catherine L. Keck · Svetlana Pack
Jonathan A. Eisen · Michael I. Lerman

In silico-initiated cloning and molecular characterization of a novel human member of the L1 gene family of neural cell adhesion molecules

Received: 14 April 1998 / Accepted: 18 June 1998

Abstract To discover genes contributing to mental retardation in 3p$^-$ syndrome patients we have used *in silico* searches for neural genes in NCBI databases (dbEST and UniGene). An EST with strong homology to the rat CAM L1 gene subsequently mapped to 3p26 was used to isolate a full-length cDNA. Molecular analysis of this cDNA, referred to as CALL (**c**ell **a**dhesion **L**1-**l**ike), showed that it is encoded by a chromosome 3p26 locus and is a novel member of the L1 gene family of neural cell adhesion molecules. Multiple lines of evidence suggest CALL is likely the human ortholog of the murine gene CHL1: it is 84% identical on the protein level, has the same domain structure, same membrane topology, and a similar expression pattern. The orthology of CALL and CHL1 was confirmed by phylogenetic analysis. By in situ hybridization, CALL is shown to be expressed regionally in a timely fashion in the central nervous system, spinal cord, and peripheral nervous system during rat development. Northern analysis and EST representation reveal that it is expressed in the brain and also outside the nervous system in some adult human tissues and tumor cell lines. The cytoplasmic domain of CALL is conserved among other members of the L1 subfamily and features sequence motifs that may involve CALL in signal transduction pathways.

M.-H. Wei · I. Karavanova · S.V. Ivanov
Intramural Research Support Program, SAIC Frederick, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, MD 21702-1201, USA

N.C. Popescu · C.L. Keck
Laboratory of Experimental Carcinogenesis, Molecular Cytogenetics Section, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

S. Pack
Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

J.A. Eisen
Department of Biological Sciences, Stanford University, Stanford, CA 94305–5020, USA

M.L. Lerman (✉)
Laboratory of Immunobiology, National Cancer Institute–Frederick Cancer Research and Development Center, Frederick, MD 21702-1201, USA
e-mail: lerman@mail.ncifcrf.gov, Tel.: +1-301-8461288, Fax: +1-301-8466145

Introduction

Cell migration, neurite (axon) outgrowth, and adhesion are fundamental events in neurogenesis and development of the nervous system (Goodman and Shatz 1993; Reichert and Boyan 1997; Schachner 1997). Cells and axon growth cones migrate toward precise targets guided by specific local and distant cues to establish the unique brain architecture based on adhesive contacts that create cell-to-cell and cell-to-matrix connections. This connectivity is fundamental in both building the brain and later in its function. The formation of these contacts and neural networks involves active dialogues between cells themselves and their immediate environment and is primarily mediated by a multitude of membrane-associated cell adhesion molecules (CAMs). The major CAMs involved in brain building and function belong to the immunoglobulin-like gene superfamily (Brummendorf and Rathjen 1995). These CAMs can be categorized into different subfamilies on the basis of their overall domain arrangement, membrane topology, and sequence similarity. Among these, the L-CAM gene family, defined by the L1 gene (Brummendorf and Rathjen 1995; Hortsch 1996), has attracted considerable interest since a range of neurological disorders, inherited and acquired, are caused by mutations in the L1 gene (Bateman et al. 1996; Fransen et al. 1996). Members of this family that include L1 and a number of highly related L1-like proteins (NgCAM, NrCAM, CHL1, E587, neurofascin) share a basic structural plan of six extracellular C2-type immunoglobulin (Ig) domains followed by five fibronectin type III (FNIII) domains linked by a single membrane-spanning region to a short (85–147 amino acids) cytoplasmic domain. The extracellular portion of these proteins is highly glycosylated and involves them in homophilic and heterop-

hilic interactions that provide a basis for a number of functions. These include cell migration, axonal growth, fasciculation, synaptogenesis, and synaptic remodeling events underlying the processes of learning and memory acquisition. The downstream events that involve these surface molecules in signal transduction are not yet elucidated. However, inspection of the cytoplasmic domains that are relatively conserved and share several highly conserved sequence motifs suggests exciting clues to the mechanisms that might be involved (Davis et al. 1993; Dubreuil et al. 1996).

Genetic research on complex behaviors such as mental retardation is striving to identify the genes responsible for its high heritability and to generate an understanding of the underlying molecular mechanisms (Plomin and Craig 1997; Plomin et al. 1994b; Wahlstrom 1990). Many single gene disorders having mental retardation among their symptoms are each responsible for a distinct subtype and gravity of the disorder. Subtle mutations in these genes are necessary and sufficient to cause the development of distinct forms of mental retardation. Mutations in the human L1 gene cause a range of clinically related diseases, collectively referred to as CRASH (**c**orpus callosum hypoplasia, mental **r**etardation, **a**dducted thumbs, **s**pastic paraplegia, and **h**ydrocephalus) that always include various degrees of mental retardation, ranging from slight learning disabilities to severe mental dysfunctions, leading to IQs in the range between 20 and 50 (Bateman et al. 1996; Fransen et al. 1996). In some patients with CRASH syndrome, mental retardation is the only clinical manifestation of the disorder.

Mental retardation is also a common feature in patients with $3p^-$ and ring chromosome 3 syndromes, with breakpoints at 3p25.3 and 3p26.1 (Asai et al. 1992; Wilson et al. 1982), respectively, suggesting that a gene on 3p26 located close to the telomere may contribute to general intelligence and, when mutated, lead to mental retardation. In addition, an allelic association strategy to discover markers associated with high versus low IQ has produced a marker on 3p that represents a brain-specific mRNA and is significantly associated with high IQ (Plomin et al. 1994a).

Therefore, in this study we set out to identify the putative gene located in 3p25.3–p26.1 by searching for ESTs representing neural-related genes in NCBI public databases (UniGene and dbEST). We report the isolation, molecular characterization, and phylogenetic analysis of CALL (**c**ell **a**dhesion **L**1-**l**ike), a novel human member of the L1 gene family which is most likely the human ortholog of the close murine homolog of L1, the CHL1 gene (Holm et al. 1996). We further demonstrate the expression of CALL in the developing rat nervous system (central and peripheral), in adult human tissues, and in some tumor cell lines.

Materials and methods

Molecular biology techniques

All molecular manipulations (screening cosmid and cDNA libraries, northern blot analysis, PCR) were performed using standard methods (Sambrook et al. 1989).

Cosmid and cDNA libraries

For cosmid isolation, the chromosome 3-specific library in the pWE15 cosmid vector described previously (Lerman et al. 1991) was screened by colony hybridization using cDNA probes representing the CALL gene. A commercial fetal brain cDNA library in the pSPORT expression vector (BRL, Cockeysville, Md.) was screened with a human EST probe (GenBank accession number R21470) representing the CALL gene.

DNA sequence determination

cDNA clones and cosmid fragments were sequenced on an Applied Biosystems 373 DNA sequencer (Stretch) using Taq Dyedeoxy Terminator Cycle Sequence kits (Applied Biosystems, Foster City, Calif.) with either vector- or clone-specific walking primers.

mRNA expression analyses

Northern blot hybridization was performed with the 7.6-kb CALL cDNA or a 122-bp fragment from the 3′ UTR using commercial MTN polyA RNA blots (Clontech, Palo Alto, Calif.) from a variety of adult human tissues and tumor cell lines. In addition, the presence of CALL transcripts was monitored by BLAST homology searches (Altschul et al. 1997) of public EST databases (http://www.ncbi.nlm.nih.gov/dbEST/index.html).

Rat fetal brain cDNA from 10- to 19-day post-coitus (dpc) embryos was used to obtain a rat cDNA probe 94% identical to human CALL ORF (residues 324–468) by PCR and subcloning (GenBank accession number AF069775). In situ hybridization on tissue sections was done according to Wilkinson and Green (1990). Whole embryos from 10-to 15-dpc rats were fixed in buffered 4% paraformaldehyde overnight and embedded in paraffin. RNA antisense probe was labeled by incorporation of ^{35}[S]-UTP (Amersham, Arlington Heights, Ill.) by in vitro transcription from the rat CALL cDNA fragment of 432 bp (GenBank accession number AF069775) cloned into a pBlueScript vector.

Fluorescence in situ hybridization (FISH)

Metaphase spreads derived from normal peripheral lymphocytes, synchronized with 5-bromodeoxyuridine, were used as a template. Probe containing the 7.6-kb CALL cDNA was labeled with digoxigenin 11-dUTP by nick translation and hybridization signals were detected with rhodamine-conjugated anti-digoxigenin antibodies (Boehringer Mannheim, Indianapolis, Ind.). The conditions of hybridization, detection of hybridization signals, and digital-image acquisition, processing, and analysis were performed as previously described (Pack et al. 1997). Chromosomes were identified by converting DAPI banding into G-simulated banding using the IP Lab Image Software (Scan Analytics, Vienna, Va.). Rehybridization with alpha-satellite centromeric probes (Oncor, Gaithersburg, Md.) was performed to confirm chromosomal localization. In addition, a cosmid probe containing part of CALL's ORF was labeled with biotin or digoxigenin using a Random Primed DNA Labeling kit (Boehringer Mannheim) and used for FISH on human chromosomes derived from methotrexate-synchronized normal peripheral lymphocyte cultures as previously described (Popescu et al. 1994).

Sequence analyses

World Wide Web-based servers were used to analyze the CALL cDNA and protein sequences. Global sequence alignments were done using BLAST and Advanced BLAST programs as provided by NCBI (http://www.ncbi.nlm.nih.gov), and BLAST 2/WU BLAST, provided by EMBL (http://www.bork.embl-heidelberg.de:8080/Blast2/). Global and local multiple sequence alignments were done using the

CLUSTAL version W program as provided by EMBL (http://www.bork.embl-heidelberg.de/Alignment/), Baylor Computing Center (http://dot.imgen.bcm.tmc.edu:9331/seq-search/alignment.html), and the Wisconsin Genetics Computer Group, package 8, program (http://www.gcg.com/). Protein domains were discovered on the Pfam server (http://www.sanger.ac.uk/Pfam/), and membrane topology on the pSORT, PHD, and Tmspred servers (http://expasy.hcuge.ch/www/expasy-top.html). Some protein motifs were found visually by inspecting local alignments or using the protein motif server (http://www.mips.biochem.mpg.de).

Phylogenetic analysis

Phylogenetic trees of CALL and closely related proteins were generated by molecular phylogenetic analysis using methods as previously described (Eisen 1998). First, proteins with significant amino acid similarity to CALL were identified in the NCBI non-redundant database using the BLAST 2 program (Altschul et al. 1997). Since this search identified hundreds of proteins, we restricted the phylogenetic analysis to only those proteins particularly closely related to CALL. The sequences of these proteins were aligned using the CLUSTALW (Thompson et al. 1994) program. Phylogenetic trees were generated from the sequence alignments using the PAUP* program (Swofford 1993) on a PowerBook 3400/180. Regions of low sequence conservation and of ambiguous alignment were excluded from the phylogenetic analysis. Trees were generated using both parsimony and distance-based methods and the robustness of the results was assessed using bootstrap resampling.

Results and discussion

Isolation of the near full-length CALL cDNA

The vast EST/UniGene databases (http://www.ncbi.nlm.nih.gov/dbEST/index.html; http://www.ncbi.nlm.nih.gov/UniGene/index.html) and chromosome transcription maps (http://www.ncbi.nlm.nih.gov/SCIENCE96/) provide a rapid access to genes whose location is defined either by genetic linkage or other means including deletion mapping in human malignancies. The human chromosome 3p25–p26 region is involved in a number of disease conditions, notably the $3p^-$ and ring (chromosome 3) syndromes (Asai et al. 1992; Wilson et al. 1982) and in several malignancies including breast cancer (Deng et al. 1996) and familial childhood neuroblastoma (Altura et al. 1997). To discover genes contributing to these diseases we have been scanning the UniGene and dbEST databases for ESTs with homologies to neural and cancer genes mapped to chromosome 3 or specifically to the 3p25–p26 region. We first found an EST (accession number R21470) with 66% homology to the rat L1 gene; this EST and several overlapping ones were subsequently mapped to 3p26 (http://www.ncbi.nlm.nih.gov/SCIENCE96/) and shown to have over 90% homology with the mouse CHL1 gene (Holm et al. 1996), a new member of the L1 gene family. We then set out to isolate a full-length cDNA for this gene, precisely map the gene by FISH, and characterize the cDNA by sequence and expression analyses.

The ESTs were completely sequenced and aligned to each other and the mouse CHL1 cDNA. The overlapping ESTs R21470 and M86085, representing the 5′ end of the human homolog of CHL1, were used to estimate the size of the transcript and to screen human brain cDNA libraries. Six cDNA clones, named BB1, BB2, BB5, BB7, BB12–8, and BB13, were isolated and completely sequenced. The longest clone, BB12–8 (about 6.8 kb in size), representing most of the ORF, contained a long 3′ UTR sequence with several canonical (AATAAA) polyadenylation signals at nucleotides 5064, 6320, 6923, and 7622, ending with a stretch of polyA; one non-canonical signal (ATTAAA) is also present at nucleotide 4425, all indicating sites of potential alternative polyadenylation. A nearly full-length sequence of 7642 bp (accession number AF002246) was then compiled using the 5′ end represented by the EST M86085 and clone BB12–8; two additional isoforms designated II and III could be discerned since clone BB5 lacked 16 amino acids, and clone BB13 was missing 53 amino acids. The nearly full-length cDNA sequence of 7642 bp (accession number AF002246), designated isoform I, corresponds to the roughly estimated 8-kb mRNA and is composed of 271 bp of the 5′ UTR, followed by a 3675-bp ORF, followed by a 3696-bp 3′ UTR. The 5′ UTR contains several in-frame termination codons prior to the strong Kozak consensus initiation sequence and is probably lacking some sequence from the 5′ UTR of the mRNA. The ORF predicts a 136-kDa protein of 1224 amino acids. The predicted CALL protein is of high complexity and is not biased against any amino acids. The total number of negatively charged residues (Asp+Glu) is 160, exceeding the number, 127, of positively charged (Arg+Lys) residues and thus predicting a theoretical pI of 5.54.

Bioinformatics analysis of CALL cDNA and protein

We employed a number of bioinformatics web-based servers to analyze the CALL cDNA and protein sequences. A global BLAST of CALL sequences employing NCBI (http://www.ncbi.nlm.nih.gov) and EMBL (http://www.bork.embl-heidelberg) servers has revealed high degrees of similarity to members of the L1 family of neural CAMs, especially to mouse CHL1, a recently discovered novel member of the L1 family (Holm et al. 1996), suggesting that CALL is a member of this important family. The protein alignments of CALL isoform I and CHL1 depicted in Fig. 1 show that they share 84% identity overall and 97% identity in the last 105 amino acids at the COOH end. This high level of identity (also seen on the cDNA level) suggests that the CALL and CHL1 genes and corresponding proteins are functionally analogous and may be orthologous genes. To better characterize the evolutionary relationship of CALL and CHL1, we performed molecular phylogenetic analysis of CALL and its close relatives (Fig. 2). Phylogenetic trees generated with multiple methods showed the same results, indicating that the relationships shown in Fig. 2 are highly robust. The high bootstrap values of the patterns in the tree also suggest that the results are robust. This analysis shows that the L1 subfamily and related proteins can be divided into distinct subgroups (labeled in Fig. 2). In particular, the patterns in the tree suggest that

Fig. 1 Global amino acid sequence alignment of mouse CHL1 and human isoform 1 CALL (**c**ell **a**dhesion **L**1-**l**ike) proteins obtained using the Wisconsin Genetics Group, package 8, program (http://www.gcg.com/). The amino acid identity is 82.5% [for CALL isoform 2 the identity is higher (84%) due to splicing out of 16 amino acids not present in CHL1]. *Vertical dashes* between human and mouse sequences indicate identity, *double dots* indicate similarity, and *single dots* indicate remote similarity between amino acids

multiple gene duplications occurred in the L1 subfamily sometime during the evolutionary history of vertebrates resulting in four distinct orthologous subgroups (BRAVO, NEUROFASCIN, CALL, and NrCAML1). The CALL subgroup is composed of the CALL and CHL1 proteins, confirming that these proteins are orthologous. Therefore, we next investigated whether the overall protein domain topography and sequence motifs of CALL conform to the basic structural plan ("L family cassette"; Holm et al. 1996) of six extracellular C2-type Ig repeats followed by four to five FNIII repeats and a highly conserved short cytoplasmic domain characteristic of the L1 family of neural CAMs. Indeed, analysis of CALL protein isoform I sequence using the multiple alignment tool CLUSTALW (http://dot.imgen.bcm.tmc.edu:9331/seq-search/alignment.html), Pfam (http://www.sanger.ac.uk/Pfam/), pSORT (http://expasy.hcuge.ch/www/expasy-top.html), TM (http://expasy.hcuge.ch/www/expasy-top.html), and SP (http://www.cbs.dtu.dk/services/SignalP) web servers has revealed a signal peptide (cleavage site between residues 25 and 26), six Ig domains (residues 50–111; 146–206; 271–328; 361–419; 454–512; 545–609) followed by four FIII domains (residues 628–714; 727–813; 825–920; 932–1021), a single-pass transmembrane peptide (residues 1103–1119), and a short cytoplasmic domain of 105 amino acids (residues 1119–1224) (Fig. 3). These results strongly support the suggestion that CALL is a novel member of the L1 family of neural CAMs.

The diverse biological functions of the L1 family proteins are mediated by homophilic and heterophilic (cell–cell or cell–matrix) bindings involving the extended extracellular portions of these CAMs (Vaughn and Bjorkman 1996). These interactions trigger the activation of neuronal RTKs, namely, FGFR (Doherty and Walsh 1996), or EPH (Zisch et al. 1997) resulting in downstream signal transduction events (Doherty and Walsh 1996; Kamiguchi and Lemmon 1997; Zisch et al. 1997). The emerging models of these interactions assume that stable binding involves multidomain interactions and is mediated by anti-parallel alignment of apposing molecules and correct arrangement

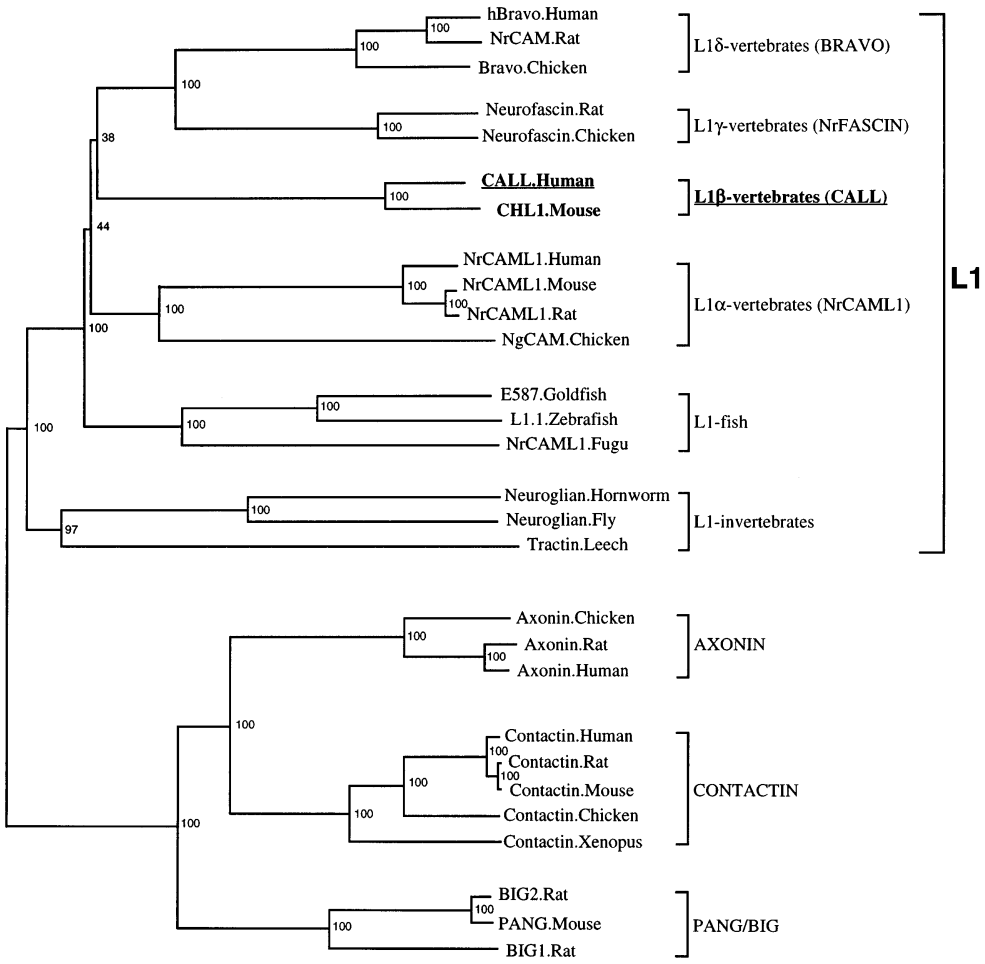

Fig. 2 Neighbor-joining phylogenetic tree of CALL and closely related proteins. The phylogenetic tree was generated from a CLUSTALW multiple sequence alignment using the neighbor-joining algorithm. Regions of low sequence conservation or ambiguous alignment were excluded from the phylogenetic analysis. Numbers corresponding to bootstrap values are indicated on the tree. Horizontal distances between proteins (as traced along the branches) correspond to estimated evolutionary distance; vertical distances between proteins have no meaning. Proposed subgroups are *bracketed*. We call particular attention to the four L1 subgroups from vertebrates (L1α–δ). Each of these are likely distinct orthologous groups suggesting that CALL is an ortholog of the mouse CHL1

of corresponding "high-affinity" sequence motifs in close proximity to one another (Ranheim et al. 1996; Vaughn and Bjorkman 1996). However, the multiplicity of L1 mutations (Fransen et al. 1996) scattered throughout the whole extracellular domain and conservation of several tyrosines and tryptophans in the FNIII domains (Fig. 3) underscore the importance of the overall structure of the extracellular segment in binding (Fransen et al. 1996; Vaughn and Bjorkman 1996). The short highly conserved motifs that include the FGFR homology sequences: APY, APYW (Doherty and Walsh 1996), and several potential integrin recognition sequences: SFT, ETA, RGDG, RGDS, and NGR (Haas and Plow 1994), are found at different locations among the L1 family members. CALL contains only one APY motif (residues 795–797) in the second FNIII domain and only one SFT integrin recognition sequence (residues 403–405) in the fourth Ig domain.

The cytoplasmic domains of the transmembrane members of the L1 family are similar in length (100–105 residues), highly conserved, and contain perfectly conserved sequence motifs (Davis et al. 1993; Dubreuil et al. 1996). Two of these are of particular importance since they provide binding sites for interaction with the intracellular cytoskeleton. The membrane-proximal sequence KGGK (Dahlin-Huppe et al. 1997) interacts with actin fibers via an unknown adaptor, and the NEDGSFIGQY sequence in the C-terminal half of the cytoplasmic domain mediates ankyrin binding (Davis and Bennett 1994) which might be involved in axonal pathfinding as suggested by mutations of the unc-44 gene in *Caenorhabditis elegans* (Otsuka et al. 1995). The cytoplasmic domain of CALL is overall 91% identical to CHL1 and 52–59% identical to other members of the family and contains both motifs involved in intracellular cytoskeleton interactions (Fig. 4). Neither CALL nor CHL1 contain the mini-exon RSLE which might be involved in cell migration (Kamiguchi and Lemmon 1997), however, they share a similar sequence, RSLN (residues 1165–1168; Fig. 4).

The complexity of interactions involving the L1 family proteins is further influenced by distinct biochemical modifications that include *O*-glycosylation, amidation, myristoylation, and, most importantly, phosphorylation of specific highly conserved tyrosine and serine residues. The ScanProsite server (http://expasy.hcuge.ch/sprot/scnpsite.html) revealed in CALL isoform 1 21 potential *N*-glycosylation sites, more than 50 *O*-glycosylation sites, 22 *N*-myristoylation sites, two amidation sites (residues 483–486, 682–685), a single cAMP/cGMP-dependent protein kinase phosphorylation site (residues 684–687), a single tyrosine kinase phosphorylation site (residues 480–487) in the Ig5 domain, 19 PKC phosphorylation sites, and 21 CK2 phosphorylation sites (data not shown). These poten-

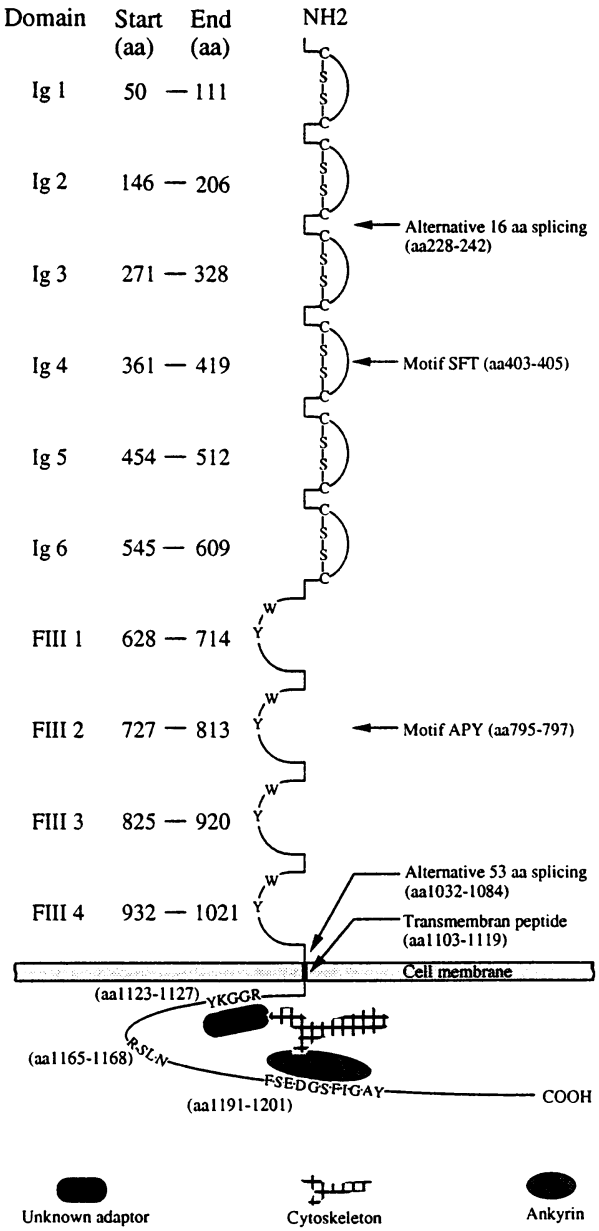

Fig. 3 Schematic representation of the modular structure of the protein backbone of isoform 1 CALL (based on analysis using Pfam (http://www.sanger.ac.uk/Pfam/), TM (http://expasy.hcuge.ch/www/expasy-top.html), SP (http://www.cbs.dtu.dk/services/SignalP), and CLUSTALW (http://dot.imgen.bcm.tmc.edu:9331/seq-search/alignment.html) programs). CALL is a transmembrane glycoprotein with six immunoglobulin-like (Ig 1–6) and four fibronectin-type III-like (FIII 1–4) domains in its extracellular segment. The intracellular tail contains conserved sequence motifs (RGGKY and FSEDGSFIGAY) that could anchor CALL to the cytoskeleton via a neural form of ankyrin. The positions of the integrin recognition sequence (SFT), the FGFR homology sequence (APY) and the positions of cysteine (C) residues inside the Ig 1–6 and the conserved tyrosine (Y) and tryptophan (W) residues in the FIII 1–4 domains are indicated

tial posttranslational modifications of CALL and, in general, other members of the family may add more complexity and flexibility to the variety of functions attributed to these proteins. However, it remains to be determined when and to what extent these modifications take place and how they influence the associated signaling pathways.

In summary, the detailed bioinformatics analysis of CALL has identified this gene as a novel human member of the L1 family of neuronal CAMs and as a likely ortholog of the mouse gene CHL1. Probably CALL, like L1 and other related proteins, could play a role in the generation of neural networks in brain building and development of the nervous system. It is, therefore, quite reasonable to speculate that CALL´s haploinsufficiency, as may occur in the 3p$^-$ and ring (chromosome 3) syndromes, could be responsible for the mental retardation phenotype observed in these patients. Interestingly, CHL1$^{-/-}$ mice were shown to be neurotic and deficient in learning abilities (M. Schachner, personal communication), while L1$^-$ mice showed gross malformations of their nervous system (Dahme et al. 1997), comparable to changes seen in the CRASH syndrome (Bateman et al. 1996; Fransen et al. 1996).

Fig. 4 Multiple amino acid sequence alignment of the cytoplasmic domains of CALL and other members of the L1 family starting with the last four amino acids of the membrane-spanning peptide: mouse CHL1, human KIAA0343, rat NrCAM, chicken NrCAM, human hBRAVO, human L1, rat L1, chicken neurofascin. A consensus sequence is included at the bottom. Highly conserved residues are indicated by *asterisks.* The sequence motifs (see Fig. 2) likely involved in anchoring CALL to the cytoskeleton (RGGK and FSEDGSFIGAY) are within these conserved residues

```
CALL_huma(1116-1224)   VCFVKRNRGGKYSVKEKEDLHPDPEIQSVK..DETFGEYSD.SDEKPLK.GSLRSLNRDM
CHL1_mous(1101-1209)   ICFVKRNRGGKYSVKEKEDLHPDPEVQSAK..DETFGEYSD.SDEKPLK.GSLRSLNRNM
KIAA0343_(1063-1175)   VCFIRRNKGGKYPVKEKEDAHADPEIQPMKEDDGTFGEYSDAEDHKPLKKGSRTPSDRTV
NrCAM_rat(1098-1210)   VCFIRRNKGGKYPVKEKEDAHADPEIQPMKEDDGTFGEYSDAEDHKPLKKGSRTPSDRTV
NrCAM_chi(1151-1263)   VCFIRRNKGGKYPVKEKEDAHADPEIQPMKEDDGTFGEYSDAEDHKPLKKGSRTPSDRTV
hBRAVO_hu(1182-1294)   VCFIRRNKGGKYPVKEKEDAHADPEIQPMKEDDGTFGEYSDAEDHKPLKKGSRTPSDRTV
CAML1_hum(1140-1246)   LCFIKRSKGGKYSVKDKEDTQVDSEARPMK..DETFGEYSDNEEKAFG..SSQPSLNGDI
CAML1_rat(1142-1248)   LCFIKRSKGGKYSVKDKEDTQVDSEARPMK..DETFGEYSDNEEKAFG..SSQPSLNGDI
Neurofasc(1054-1161)   VCFIKRSRGGKYPVREKKDVPLGPEDPKEE..DGSFDYSD..EDNKPLQ.GSQTSLDGTI
                       .**..* .**** *..* *     *       * .*       .       *       .

CALL_huma(1116-1224)   QPTESADSLVEYGEGDHGLFSEDGSFIGAYAGSKEKGSVESNGSSTATFPLRA
CHL1_mous(1101-1209)   QPTESADSLVEYGEGDQSIFNEDGSFIGAYTGAKEKGSVESNGSSTATFPLRA
KIAA0343_(1063-1175)   KKEDSDDSLVDYGEGVNGQFNEDGSFIGQYSGKKEKEPAEGNESSEAPSPVNA
NrCAM_rat(1098-1210)   KKEDSDDSLVDYGEGVNGQFNEDGSFIGQYSGKKEKEPAEGNESSEAPSPVNA
NrCAM_chi(1151-1263)   KKEDSDDSLVDYGEGVNGQFNEDGSFIGQYSGKKEKEPAEGNESSEAPSPVNA
hBRAVO_hu(1182-1294)   KKEDSDDSLVDYGEGVNGQFNEDGSFIGQYSGKKEKEPAEGNESSEAPSPVNA
CAML1_hum(1140-1246)   KPLGSDDSLADYGGSVDVQFNEDGSFIGQYSGKKEKEAAGGNDSSGATSPI..
CAML1_rat(1142-1248)   KPLGSDDSLADYGGSVDVQFNEDGSFIGQYSGKKEKEAAGGNDSSGATSPI..
Neurofasc(1054-1161)   KQQESDDSLVDYGEGGEGQFNEDGSFIGQYTVRKDKEETEGNESSEATSPVNA
                       .   * *** .**      * ******* *.  *.*     * ** *  *.
```

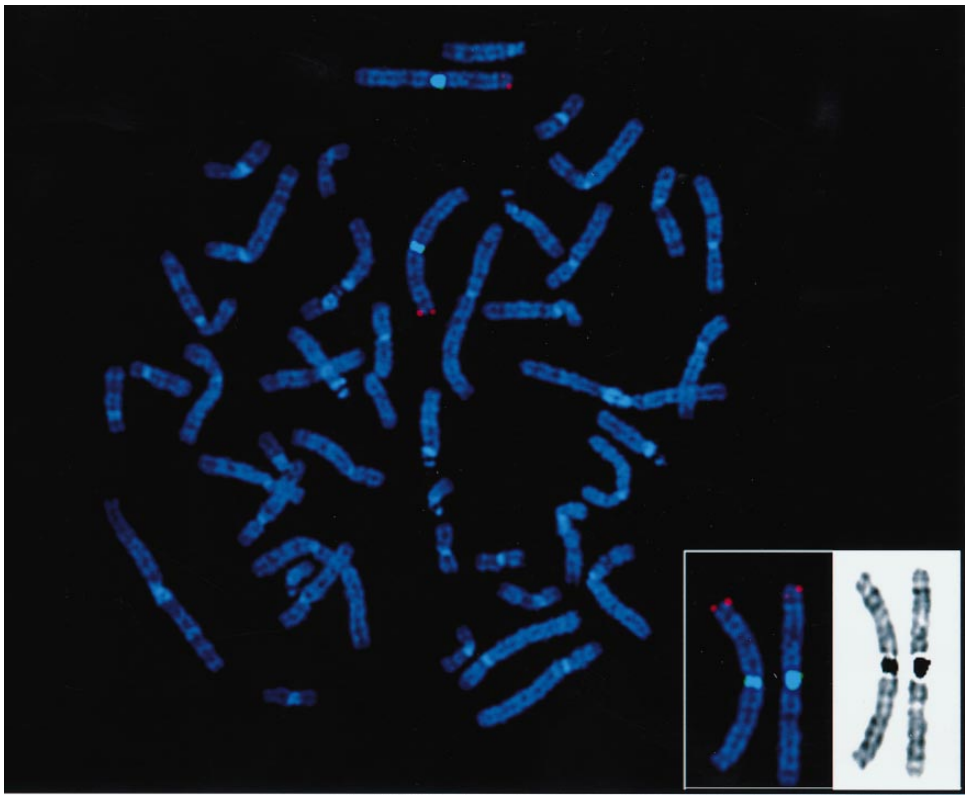

Fig. 5 Subchromosomal localization of the human CALL gene. Metaphase after fluorescence in situ hybridization using the cDNA probe showing a single location of the CALL gene on the long arm of chromosome 3p26.1. *Inset* demonstrates the position of CALL on DAPI-banded human chromosomes, 3p26.1

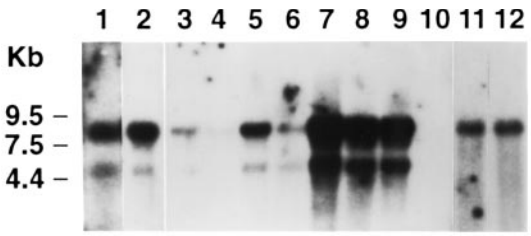

Fig. 6 Northern blot analysis using CALL isoform 1 cDNA or a 3′ UTR fragment as hybridization probes. The RNA filters are from Clontech (#7759, #7760) and contain 2 µg of poly(A)$^+$ mRNA per tissue indicated: *1* heart, *2* brain, *3* spleen, *4* thymus, *5* prostate, *6* testis, *7* ovary, *8* small intestine, *9* colon, *10* peripheral blood leukocytes. *Lanes 11* (heart) and *12* (brain) were probed with a 142-bp PCR fragment amplified from the 3′ UTR of the CALL cDNA (nucleotides 5375–5517, GenBank accession number AF002246) downstream of a non-canonical (ATTAAA) polyadenylation signal at nucleotides 4425–4431

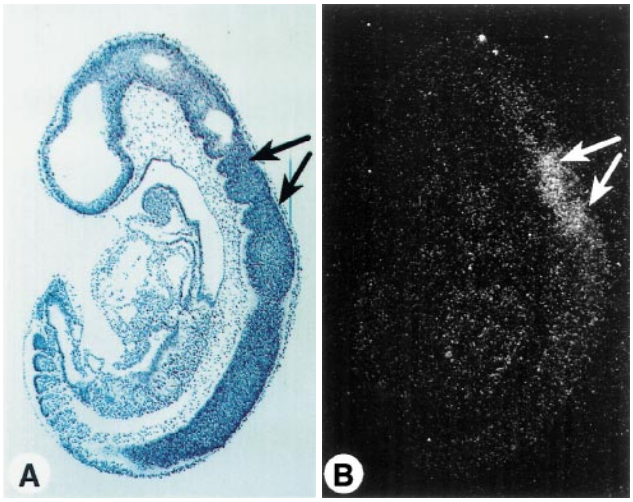

Fig. 7 Bright (**A**) and dark field (**B**) images of a sagittal section of an 11-day post-coitus (dpc) rat embryo. Expression of CALL is present only in the rhombomeres of the myelencephalon

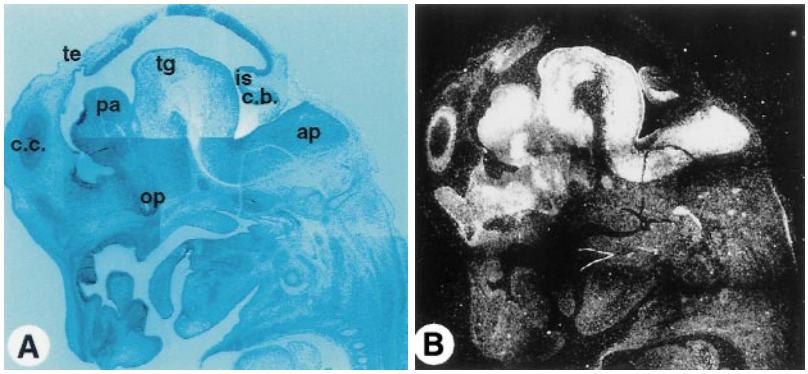

Fig. 8 Bright (**A**) and dark field (**B**) images of a sagittal section of a 15-dpc rat head. Only neural tissues express CALL. Strong expression is seen in the cortex, pallidum, tegmentum, cerebellum, pons area, and medulla oblongata (*ap* alar plate, *cb* cerebellum, *cc* cerebral cortex, *is* isthmus, *pa* pallidum, *tg* tegmentum, *te* telencephalon)

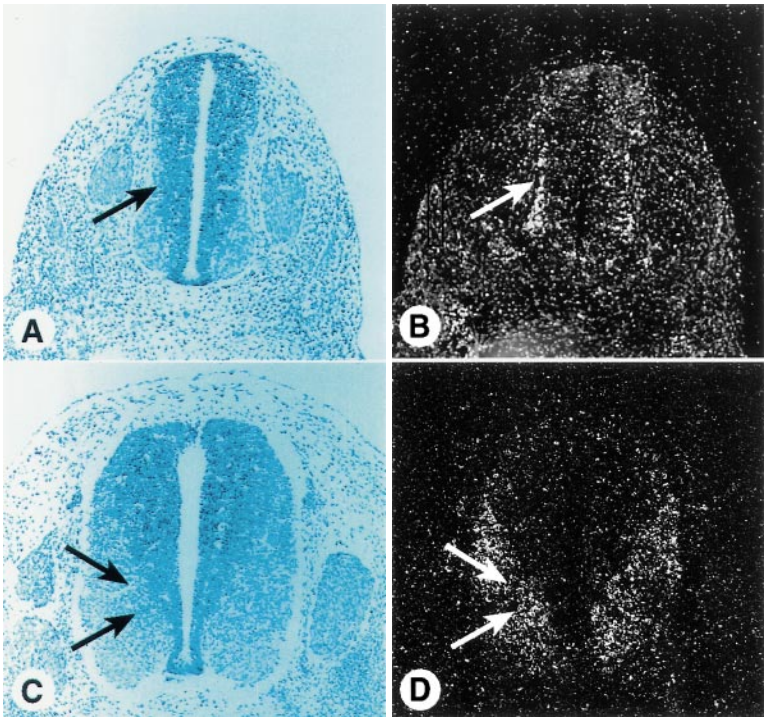

Fig. 9 Bright (**A, C**) and dark field (**B, D**) images of a transverse (coronal) sections of 12-dpc (**A, B**) and 14-dpc (**C, D**) rat spinal cord. Expression is at the boundary of the intermediate and marginal zones (**A, B**) and ventrolateral mantle layer (**C, D**)

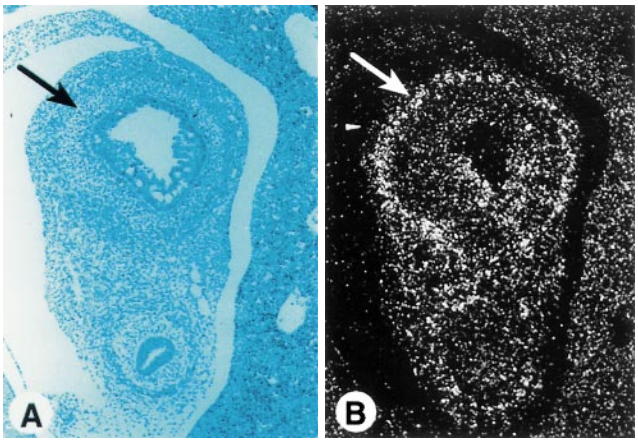

Fig. 10 Bright (**A**) and dark field (**B**) images of a sagittal section of 15-dpc rat stomach. Expression is in the enteric ganglia

Chromosomal mapping of the CALL gene locus

The chromosomal location of the CALL gene locus was determined by FISH using a 7.6-kb cDNA probe or a cosmid probe containing part of the cDNA. CALL was localized by both probes as a single locus on 3p26.1 (Fig. 5) confirming the mapping of ESTs representing CALL by RH hybridization as reported previously (UniGene entry: Hs.21226).

Expression analysis of CALL

We analyzed the expression patterns of the CALL gene by Northern blot hybridization of human adult tissues and tumor cell lines, monitoring EST databases, and by in situ hybridization during rat brain development. Northern blot hybridization was performed with commercial multiple-tissue polyA RNA blots (Clontech). CALL is expressed in many adult human tissues, predominantly as an 8-kb mRNA; a smaller sized transcript of about 4.5 kb resulting from alternative polyadenylation is observed in some tissues and cell lines at different levels compared to the predominant 8-kb band (Fig. 6). The gene is highly expressed in brain, heart, prostate, ovary, small intestine, and colon and moderately expressed in lung, kidney, pancreas, spleen, and testis. Very low expression was detected in placenta, liver, skeletal muscle, thymus, and peripheral blood leukocytes (data not shown; Fig. 6). Of the tumor cell lines tested (RNA filter from Clontech, #7757–1), melanoma G361 cells showed high levels of expression, HeLa S3 moderate expression, no expression was detected in several leukemia, colorectal (SW480 cells), and lung (A549 cells) adenocarcinomas (data not shown). The EST matches were mostly in the brain libraries (75%), testis, ovary, prostate epithelium, uterus, and intestine. Several ESTs were detected in libraries made from tumor tissues, namely, ovarian carcinoma, melanomas, and liposarcomas (NCBI, dBEST, NCI CGAP database, http://www.ncbi.nlm.nih.gov/ncicgap/). These combined results suggest that CALL is expressed outside the nervous system, however, it remains to be seen whether this expression does not come from nervous ganglia residing in the somatic tissues.

To gain more information on CALL expression in the developing central nervous system we performed (albeit limited) in situ hybridization studies with an antisense rat CALL probe which is 92% identical to murine CHL1 and 86% identical to CALL on the RNA level and 94–95% on

the protein level (GenBank accession number AF069775). Here we present an initial assessment of sites of CALL expression in rat nervous system development at embryonic days 11–15 (Figs. 7–10). At 11 dpc CALL expression was detected only in the rhombomeres of the myelencephalon (Fig. 7). At 13 dpc, expression was localized in the most ventral neurons of the mes- and metencephalon corresponding to zones of postmitotic neurons that start to differentiate in the rat brain at this stage of embryonic development (data not shown). As development proceeds the expression of CALL becomes more widespread throughout the brain: at 15 dpc (Fig. 8) CALL mRNA was detected in the forebrain, midbrain, and hindbrain, remaining restricted to the postmitotic neurons of differentiating fields in the forebrain, midbrain, and hindbrain (Fig. 8). In the peripheral nervous system the expression was also localized in the postmitotic neurons of the spinal cord, starting at 12 dpc in a thin stripe of ventrolateral cells and was readily detected throughout the mantle layer at 14–15 dpc (Fig. 9). At 15 dpc, CALL-positive cells were detected in the stomach, most probably corresponding to the enteric ganglia (Fig. 10).

Overall, we observed that CALL mRNA is strongly expressed in all parts of the developing brain, spinal cord, and some peripheral ganglia. It is expressed in a regionalized and timely fashion, reflecting the numerous putative functions served by the CALL protein in the control of differentiation, determination, and maturation of neurons and neural networks. The continuous expression of CALL throughout adulthood (see Northern analysis of expression in adult human tissues) implies that the functions served by the product(s) of the CALL gene seem to play a role(s) in the function of the adult nervous system and some other tissues.

Conclusions

We have presented *in silico*-initiated cloning, bioinformatics analyses, and molecular characterization of CALL, a novel human member of the L1 family of neuronal CAMs. On the basis of the data presented we may draw the following conclusions. First, our work demonstrates the power of *in silico* approaches in gene discovery and subsequent characterization of the structure and function of the gene by bioinformatic computations. Second, these methods showed the extent of structural and functional relatedness of CALL to the L1 family including the molecular processes involving these proteins. Third, on the basis of these findings we may assume that the CALL gene plays a unique role in brain development and in the function of the adult nervous system. In addition, given the location of the gene at 3p26 and its deletion in $3p^-$ syndrome patients manifesting mental retardation, we may speculate that the mechanistic function(s) of CALL may involve the gene in cognitive intelligence. And, finally, since the gene is located in a region involved in carcinogenesis, CALL, if expressed in epithelial cells, may play a role in tumor development.

Acknowledgements We wish to thank Jerrold Ward for helping to interpret the rat brain in situ hybridization patterns in development and Alex Bateman and Evan Birney for helping with Pfam analysis of the domain structure of CALL. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-56000. The content of the publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Altura RA, Maris JM, Li H, Boyett JM, Brodeur GM, Look T (1997) Novel regions of chromosomal loss in familial neuroblastoma by comparative genomic hybridization. Genes Chromosom Cancer 19:176–184

Asai M, Ito Y, Iguchi T, Ito J, Okada N, Oishi H (1992) Terminal deletion of the short arm of chromosome 3. Jpn J Hum Genet 37:163–168

Bateman A, Jouet M, MacFarlane J, Du J-S, Kenwrick S, Clothia C (1996) Outline structure of the human L1 cell adhesion molecule and the sites where mutations cause neurological disorders. EMBO J 15:6050–6059

Brummendorf T, Rathjen FG (1995) Cell adhesion molecules 1: immunoglobulin superfamily. Protein Profile 2:963–1108

Dahlin-Huppe K, Berglund EO, Ranscht B, Stallcup WB (1997) Mutational analysis of the L1 neuronal cell adhesion molecule identifies membrane-proximal amino acids of the cytoplasmic domain that are required for cytoskeletal anchorage. Mol Cell Neurosci 9:144–156

Dahme M, Bartsch U, Anliker B, Schachner M, Mantei N (1997) Disruption of the mouse L1 gene leads to malformations of the nervous system. Nat Genet 17:346–349

Davis JQ, Bennett V (1994) Ankirin binding activity shared by the neurofascin/L1/NrCAM family of nervous system cell adhesion molecules. J Biol Chem 269:27163–27166

Davis JQ, McLaughlin T, Bennett V (1993) Ankirin-binding proteins related to nervous system cell adhesion molecules: candidates to provide transmembrane and intercellular connections in adult brain. J Cell Biol 121:121–133

Deng G, Lu Y, Zlotnikov G, Thor AD, Smith HS (1996) Loss of heterozygosity in normal tissue adjacent to breast carcinomas. Science 274:2057–2059

Doherty P, Walsh FS (1996) CAM–FGF receptor interactions: a model for axonal growth. Mol Cell Neurosci 8:99–111

Dubreuil RR, MacVicar G, Dissanayake S, Liu C, Homer D, Hortsch M (1996) Neuroglian-mediated cell adhesion induces assembly on the membrane skeleton at cell contact sites. J Cell Biol 133:647–655

Eisen JA (1998) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res 8:163–167

Fransen E, Vits L, Van Gamp G, Willems PJ (1996) CRASH syndrome: the clinical spectrum of mutations in L1, a neuronal cell adhesion molecule. Am J Med Genet 64:73–77

Goodman CS, Shatz CJ (1993) Developmental mechanisms that generate presize patterns of neuronal connectivity. Cell 72:77–98

Haas TA, Plow EF (1994) Integrin–ligand interactions: a year in review. Curr Opin Cell Biol 6:656–662

Holm J, Hillenbrand R, Steuber V, Bartsch U, Moos M, Lubbert H, Montag D, Schachner M (1996) Structural features of a close homologue of L1 (CHL1) in the mouse: a new member of the L1 family of neural recognition molecules. Eur J Neurosci 8:1613–1629

Hortsch M (1996) The L1 family of neural cell adhesion molecules: old proteins performing new tricks. Neuron 17:587–593

Kamiguchi H, Lemmon V (1997) Neural cell adhesion molecule L1: signaling pathways and growth cone motility. J Neurosci Res 49:1–8

Lerman MI, Latif F, Glenn GM, Daniel L, Brauch H, Hosoe S, Hampsch K, Delisio J, Orcutt ML, McBride OW, Grzeschik KH, Takahashi T, Minna J, Anglard P, Linehan WM, Zbar B (1991) Isolation and regional localization of a large collection (2000) of single copy DNA fragments on human chromosome 3 for mapping and cloning tumor suppressor genes. Hum Genet 86:567–577

Otsuka AJ, Franco R, Yang B, Shim KH, Tang LZ, Zhang YY, Boontrakulpoontawee P, Jeyaprakash A, Hedgecock E, Wheaton VI (1995) An ankirin-related gene (unc-44) is necessary for proper axonal guidance in *Caenorhabditis elegans*. J Cell Biol 129:1081–1092

Pack SD, Tanigami A, Ledbetter DH, Sato T, Fukuda MN (1997) Assignment of trophoblast/endometrial epithelium cell adhesion molecule trophinin gene TRO to human chromosome bands-Xp11.22óp11.21 by in situ hybridization. Cytogenet Cell Genet 79:123–124

Plomin R, Craig I (1997) Human behavioral genetics of cognitive abilities and disabilities. Bioessays 19:1117–1124

Plomin R, McClearn GE, Smith DL, Vignetti S, Chorney MJ, Chorney K, Venditti CP, Kasarda S, Thompson LA, Detterman DK, Daniels J, Owen M, McGuffin P (1994a) DNA markers associated with high versus low IQ: the IQ quantitative trait loci (QTL) project. Behav Genet 24:107–118

Plomin R, Owen MJ, McGuffin P (1994b) The genetic basis of complex human behaviors. Science 264:1733–1739

Popescu NC, Zimonjic D, Hatch C, Bonner W (1994) Chromosomal mapping of the human histone gene H2AZ to 4q24 by fluorescence in situ hybridization. Genomics 20:333–335

Ranheim TS, Edelman GM, Cunningham BA (1996) Homophilic adhesion by the neural cell adhesion molecule involves multiple immunoglobulin domains. Proc Natl Acad Sci USA 93:4071–4075

Reichert H, Boyan G (1997) Building a brain: developmental insights in insects. Trends Neurosci 20:258–264

Sambrook Y, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schachner M (1997) Neural recognition molecules and synaptic plasticity. Curr Opin Cell Biol 9:627–634

Swofford DL (1993) Paup: a computer program for phylogenetic inference using maximum parsimony. J Gen Physiol 102:9A

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Vaughn DE, Bjorkman PJ (1996) The (Greek) key to structures of neural adhesion molecules. Neuron 16:261–273

Wahlstrom J (1990) Gene map of mental retardation. J Ment Defic Res 34:11–27

Wilkinson D, Green J (1990) In situ hybridization and the three-dimensional reconstruction of serial sections. In: Copp AJ, Cokroft (eds) Postimplantation mammalian embryos. Oxford University Press, London, pp 155–171

Wilson GN, Pooley J, Rarker J (1982) The phenotype of ring chromosome 3. J Med Genet 19:471–473

Zisch AH, Stallcup WB, Chong LD, Dahlin-Huppe K, Voshol J, Schachner M, Pasquale EB (1997) Tyrosine phosphorylation of L1 family adhesion molecules: implication of the Eph kinase CEK5. J Neurosci Res 47:655–665