



HHS Public Access

Author manuscript

Hum Genet. Author manuscript; available in PMC 2016 June 06.

Published in final edited form as:

Hum Genet. 2004 July ; 115(2): 139–148. doi:10.1007/s00439-004-1134-6.

A giant novel gene undergoing extensive alternative splicing is severed by a Cornelia de Lange-associated translocation breakpoint at 3q26.3

Emma T. Tonkin,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

Melanie Smith,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

Piet Eichhorn,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

Sandie Jones,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

Burhan Imamwerdi,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

Susan Lindsay,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

Mike Jackson,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

Tzu-Jou Wang,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

Correspondence to: Tom Strachan, tom.strachan@ncl.ac.uk.

Present address: M. Smith, Serono International, 1211 Geneva 20, Switzerland

Present address: P. Eichhorn, The Netherlands Cancer Institute, Antoni van Leeuwenhoek Ziekenhuis, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Present address: M. Ireland, Postgraduate Institute for Medicine and Dentistry, University of Newcastle, Newcastle upon Tyne, NE2 4AB, UK

Note added in proof After completion of the work described in this paper, investigation of a new gene, *NIPBL* on chromosome 5p13 has led to the identification of associated mutations in at least a proportion of CdLS patients.

Tonkin E, Wang T-J, Lisgo S, Bamshad M, Strachan T (2004) *NIPBL*, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nature Genetics* DOI 10.1038/ng1363, and Krantz et al. (2004; full reference not available). Both papers scheduled for online publication on 16 May 2004.

Maggie Ireland,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

John Burn,

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

Ian D. Krantz,

Division of Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia and the University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

Philippa Carr, and

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK

Tom Strachan

Institute of Human Genetics, International Centre for Life, University of Newcastle, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK, Tel.: +44-191-2418616 Fax: +44-191-2418666

Tom Strachan: tom.strachan@ncl.ac.uk

Abstract

Cornelia de Lange syndrome (CdLS) is a rare developmental malformation syndrome characterised by mental handicap, growth retardation, distinctive facial features and limb reduction defects. The vast majority of CdLS cases are sporadic. We carried out a high density bacterial artificial chromosome (BAC) microarray comparative genome hybridisation screen but no evidence was found for a consistent pattern of microdeletion/micro-duplication. As an alternative, we focused on identifying chromosomal regions spanning associated translocation breakpoints. We prioritised the distal 3q region because of the occurrence, in a classical CdLS patient, of a de novo balanced translocation with a breakpoint at 3q26.3 and of reports of phenotypic overlap between cases of mild CdLS and individuals trisomic for the 3q26-q27 region. We show that the 3q26.3 breakpoint severs a previously uncharacterised giant gene, *NAALADL2*, containing at least 32 exons spanning 1.37 Mb. Northern blot analysis identified up to six different transcripts in the 1–10 kb range with strongest expression in kidney and placenta; embryonic expression was largely confined to duodenal and stomach endoderm, mesonephros, metanephros and pancreas. Transcript analysis identified extensive alternative splicing leading to multiple 5' and 3' untranslated regions and variable coding sequences. Multiple protein isoforms were defined by different N-terminal regions (with at least four alternative initiating methionine codons), and by differential protein truncation/use of alternative C-terminal sequences attributable to alternative splicing/polyadenylation. Outside the N-terminal regions, the predicted proteins showed significant homology to N-acetylated alpha-linked acidic dipeptidase and transferrin receptors. Mutation screening of *NAALADL2* in a panel of CdLS patient DNA samples failed to identify patient-specific mutations. We discuss the possibility that the 3q26.3 translocation could nevertheless contribute to pathogenesis.

Introduction

Cornelia de Lange Syndrome (CdLS; MIM 122470), which is also known as Brachmann-de Lange syndrome, is a developmental malformation syndrome characterised by mental handicap, growth retardation, distinctive facial features and limb reduction defects (Jackson et al. 1993; Ireland et al. 1993). Classical cases often have severe limb abnormalities and mental retardation but a clinically mild subtype is also recognised in which affected children are less severely growth retarded and developmentally delayed (Van Allen et al. 1993; Allanson et al. 1997). The disorder is rare, with an approximate incidence of 1:40,000.

Affected individuals typically occur as sporadic cases but some reports have documented families with multiple affected members. The low estimates for the general recurrence risk for sibs (Jackson et al. 1993) argues against recessive inheritance and, in some of the multi-case pedigrees, there is strong evidence for autosomal dominant inheritance (Krajewska-Walasek et al. 1998; Kozma 1996; McKenney et al. 1996; Russell et al. 2001; McConnell et al. 2003). In other multi-case pedigrees, the birth of more than one affected child to unaffected parents can be explained by gonadal mosaicism.

Because of the predominantly sporadic nature of the disorder, it has been difficult to obtain suitably informative linkage data. Clues to subchromosomal location have, however, been suggested from associated chromosomal abnormalities. A variety of abnormalities have been reported (for a review, see Kousseff et al. 1994) but much attention has been focused on distal 3q as a result of (1) phenotypic overlap between the duplication 3q syndrome and mild CdLS (Steinbach et al. 1981; Wilson et al. 1985) and (2) the identification of a balanced de novo translocation t(3;17)(q26.3;q23.1) in a patient with classical CdLS, some time after the original reports of phenotypic similarity between mild CdLS and the duplication 3q syndrome (Ireland et al. 1991). The significance of the 3q26.3 breakpoint appeared to be enhanced when the 3q26.3-q27 interval was suggested as the critical region for the dup 3q syndrome (Aqua et al. 1995; Ireland et al. 1995; Rizzu et al. 1997); individuals trisomic for regions immediately proximal or distal to this region were reported not to resemble CdLS (Rizzu et al. 1997; Lopez-Rangel et al. 1993).

In the present report, we show that the CdLS-associated translocation breakpoint at 3q26.3 severs a giant (1.37 Mb) previously uncharacterised gene, *NAALADL2* without any accompanying significant deletion or DNA rearrangement. *NAALADL2* is subject to extensive alternative processing, resulting in at least four protein isoforms, some with a strongly predicted transmembrane region near the N-terminus and others appearing to lack such a region. The N-terminal sequence preceding the strongly predicted transmembrane region is not significantly related to other protein sequences but the remaining protein sequence shows significant, but low, homology to the N-acetylated alpha-linked acidic dipeptidase and to transferrin receptors. Mutation screening of *NAALADL2* exons in a panel of CdLS patient DNA and control DNA samples failed to identify patient-specific mutations. We review the question of whether the translocation breakpoint at 3q26.3 could, nevertheless, contribute to pathogenesis.

Materials and methods

Bacterial artificial chromosome microarray comparative genome hybridisation (CGH) screen

DNA samples from 11 individuals with classical CdLS (including eight subjects with limb reduction defects) were hybridised in CGH screens against genomic microarrays containing over 3,000 bacterial artificial chromosome (BAC) clones at approximately 1 Mb intervals. Construction of the microarrays and the labelling and hybridisation of the patient-specific probes were as previously described (Fiegler et al. 2003).

Contig assembly as an aid to defining the 3q26 translocation breakpoint

Approximately 25 ng single-copy 3q26 region marker probes were used to screen gridded yeast artificial chromosome (YAC) libraries (Chumakov et al. 1992, Anand et al. 1990, Larin et al. 1991), the RPCI-1 P1 artificial chromosome (PAC) library (Ioannou et al. 1994), the RPCI-11 BAC library (Osoegawa et al. 2001), and the Lawrence Livermore National Laboratory chromosome-3-specific cosmid library, constructed by using the method of Gingrich et al. (1996). Filters were hybridised in 0.5 M phosphate buffer with 7% SDS and 1% bovine serum albumin overnight at 65°C and washed in 2× SSC/0.1%SDS (1× SSC = 150 mM NaCl, 15 mM sodium citrate, pH 7.0) at room temperature before exposure to film. Following clone isolation, the predicted marker content was confirmed by the polymerase chain reaction (PCR) and Southern hybridisation. As required, internal and end clone sequences were generated by using inverse-PCR (Ochman et al. 1988), vector-hexamer PCR (Herring et al. 1998) or Alu-PCR (Nelson et al. 1989) or by direct sequencing of BAC and cosmid ends as described previously (Tonkin et al. 2002). Derived single-copy end sequences were used as probes for re-screening the libraries to develop a contiguous contig spanning the 3q26 breakpoint region, as determined by fluorescent in situ hybridisation (FISH; see below). BACs spanning the region were sequenced by dedicated chromosome 3 sequencing centres, notably the Baylor College of Medicine sequencing centre.

Mapping the translocation breakpoints in the t(3;17) translocation cell line

A lymphoblastoid cell line was generated from a blood sample from the t(3;17)(q26.3;q23.1) case previously described by Ireland et al. (1991). Chromosome-specific centromere probes obtained from Appligene Oncor were used to identify chromosomes 3 and 17. Genomic clones (a progressive hierarchy of YAC, BAC and cosmid clones) from 3q26 or 17q23 (identified in the same manner as 3q26 clones) were labelled by nick translation incorporating either SpectrumGreen or SpectrumRed (Vysis), according to the manufacturer's instructions, and hybridised to metaphase chromosomes derived from the CdLS t(3;17) lymphoblastoid cell line. Breakpoint-spanning clones were identified by the presence of three signals, representing the two derivative translocation chromosomes and a normal chromosome 3 or chromosome 17 homologue.

cDNA sequence assembly and determination of exon-intron organization

At the outset, expressed sequence tags (ESTs) were identified by using publically available sources and placed on the physical map by PCR/Southern hybridisation-based typing or,

more recently, BLAST analysis (Altschul et al. 1990). Full-length *NAALADL2* cDNA sequences were assembled by the extension of partial cDNA sequences by in-house sequencing and the determination of sequence overlap by using *in silico* analysis and by 5' and 3' rapid amplification of cDNA ends (RACE). RACE was conducted as described by the manufacturer (Clontech) by amplifying human adult kidney cDNA with nested pairs of exon-specific primers in combination with linker-specific primers and in the presence of 1 M betaine. Products were cloned into pGEM-T Easy (Promega) before sequencing by Amersham ET chemistry for MegaBACE (Amersham Biosciences). The *NAALADL2* exon-intron organisation was determined by referencing cDNA sequences against cognate BAC sequences. Where gaps existed in the genomic sequence, specific regions of BAC clones were sequenced in-house as described previously (Tonkin et al. 2002).

Gene expression analyses

Gene expression analyses involved labelling a 581-bp cDNA probe (spanning the sequence in exons 14 and 17). For Northern blot analyses, the probe was labelled to high-specific activity with [α - 32 P]-dCTP by the Rediprime II random prime labelling system (Amersham) and hybridised against both adult and fetal blots (Clontech) containing approximately 2 μ g mRNA per lane at 68°C for 1 h. Blots were washed in 0.1 \times SSC/0.1 SDS at room temperature before exposure to film. For tissue *in situ* hybridisation, linearised plasmid containing the 581-bp fragment was transcribed with T7 and SP6 RNA polymerases to generate [35 S]-UTP-labelled sense and antisense riboprobes, respectively, which were hybridised to sections of human embryonic tissue. The collection and use of human embryos was carried out with ethical permission from the joint Ethics Committee of the Newcastle Health Authority and with appropriate signed consent. Embryos were collected following medically induced termination of pregnancy (Bullen et al. 1998) and were staged by microscopic examination. Fixation and processing of tissue and *in situ* hybridisation of labelled probes was carried out as previously described (Lako et al. 1998).

Mutation screening

Point mutations were screened by single-strand conformation polymorphism (SSCP)/heteroduplex analysis by means of standard protocols. Amplified exon products (150–250 bp including 50–100 bp flanking intron) or overlapping internal exon sequences were denatured and size-fractionated in 1 \times MDE gels (BioWhittaker), containing 5% glycerol and 0.6 \times TBE buffer (1 \times TBE buffer = 0.09 M TRIS-borate, 0.002 M EDTA, pH 8.3) at 300 V and 15°C, for approximately 20 h (depending on fragment size), and visualised by silver staining. Any samples demonstrating band differences when compared with an unaffected control were sequenced. Large-scale deletions were screened by fluorescent multiplex amplifiable probe hybridisation (MAPH; White et al. 2002). Amplified exon sequences were cloned into the vector pGEM-T Easy (Promega). Additional control probes were generated for a number of autosomal loci not on chromosome 3 and to SRY (Y chromosome) and DMD (X chromosome). Following hybridisation and stringent washes, the amplified fluorescently labelled products (1.5 μ l/MAPH amplification) were size-fractionated on a 5% LongRanger gel (BioWhittaker) on an ABI 377 automated DNA sequencer and the data were analysed.

Sequence motif and comparative sequence homology analysis

The NIX suite of programs (<http://www.hgmp.mrc.ac.uk/Registered/>) was used in the analysis of large regions of genomic sequence. It included the use of the Eponine program for determining transcriptional start sites in mammalian DNA (Down and Hubbard 2002). Additional analysis of the 5' region of the gene, including the putative promoter and transcriptional start site, made use of publicly available software at the Zhang lab (<http://rulai.cshl.org/software/index1.htm>). Transmembrane region prediction was achieved by using the TMPred program, which is available at http://www.ch.embnet.org/software/TMPRED_form.html. Human–rodent comparative sequence homology was carried out by using VISTA (Dubchak et al. 2000; Mayor et al. 2000; <http://www-gsd.lbl.gov/vista/>) and PipMaker (Schwartz et al. 2000; <http://bio.cse.psu.edu/pipmaker/>). The output data from VISTA were also used in rVISTA (Loots et al. 2002) as a primary prediction tool for conserved transcription factor-binding sites.

DNA sequence accession numbers

DNA sequences were deposited with accession numbers as follows: AJ607395—*NAALADL2* gene, IMAGE clone 2095161; AJ607396—*NAALADL2* gene, IMAGE clone 2969158; AJ607397—*NAA-LADL2* gene, IMAGE clone 1847027; AJ607398—*NAALADL2* gene, 5' untranslated region (UTR) variant 1; AJ607399—*NAALADL2* gene, 5'UTR variant 2; AJ607400—*NAALADL2* gene, 5'UTR variant 3; AJ607401—*NAALADL2* gene, 5'UTR variant 4.

Results

A CdLS-associated translocation breakpoint at 3q26.3 severs a novel gene, *NAALADL2*

In parallel with the early Whitehead/MIT physical mapping efforts, we developed and extended a YAC map of the 3q26.3 region (approximately 5.6 Mb) and then assembled a PAC- and BAC-based contig of approximately 3.2 Mb (data not shown). We subsequently identified a BAC (RPCI-11 827e14; accession no. AC063954) that spanned the 3q26.3 translocation breakpoint (Fig. 1a). We derived cosmid clones (AC110-B10 and AC216-H7) spanning the breakpoint and defining a minimal breakpoint region of about 70 kb.

Seeking cognate transcripts across the BAC contig, we identified the IMAGE 2095161 clone through a positive BLAST hit against a publically available partial insert sequence (accession no. AI420905). After the sequencing of the full-length insert of the IMAGE 2095161 clone, it became apparent that this transcript mapped to the breakpoint-spanning clone RPCI-11 827e14 (AC0639854) and also to a proximal flanking BAC RPCI-11 393n4 (AC069218) and a distal flanking BAC RPCI-11 5 11 (AC021681). We established that the latter two BACs were approximately 700 kb apart, suggesting that the gene, which gave rise to the 2095161 clone sequence, was severed by the translocation in this patient. We used a combination of 5' and 3'RACE, database searching and in-house cDNA clone sequencing to define extended transcripts for the gene disrupted by the 3q26.3 breakpoint. The recent availability of complete genomic sequences for this region of 3q allowed us to define precisely the full exon/intron organization of the *NAA-LADL2* breakpoint gene.

NAALADL2 is a giant gene and undergoes extensive alternative processing in both the untranslated regions and in the coding sequence

As illustrated in Fig. 2b, the *NAALADL2* gene spans 1.37 Mb. There is evidence for at least 32 exons, of which 29 are validated by cognate sequences in cDNA clones. A further three exons (exons 1, 5 and 10, specifying 5'UTR sequences) have so far been supported by 5'RACE experiments only. *NAALADL2* is located in a genomic environment that is generally exon-poor (the six exons of the neighbouring neuroligin I gene, *NLGN1*, are distributed over 0.7 Mb) and some *NAALADL2* introns are hundreds of kilobases long, with the longest, intron 13, spanning over 230kb. *NAALADL2* is subject to extensive alternative processing, leading to many different 5'UTR and 3'UTR sequences (Fig. 2c, d) plus a considerable variety of alternative predicted protein products.

The 5'UTR can be specified by the sequences present in the first 16 exons but there is evidence for extensive alternative splicing (Fig. 2c, d). Early 5'RACE experiments with exon 11-specific primers identified five upstream exons (now labelled as exons 1, 2, 5, 7, 9 and 10) in four types of product. Another four exons (exons 3, 4, 6 and 8) are evident from cDNA sequencing (IMAGE clone 1845618; accession no. AI218427). Whereas exon 1 sequences are not apparent in cDNA clone sequences, exon 2 is represented in different cDNA clones and is located within a CpG island sequence for which high confidence transcriptional start sites are predicted (Eponine program; score =0.9991; see above).

Additional alternative RNA splicing and polyadenylation leads to *NAALADL2* isoforms with alternative coding sequences and 3'UTRs (Fig. 2c, d). For full-length isoforms, the terminal exon is a variant of exon 32 with either a 2,433-bp 3'UTR (e.g. clone DKFZp686E21111) or a 1,034-bp 3'UTR (e.g. clone DKFZp666C1410). Orthologous transcripts in mouse, cow, chick and pig have sequences corresponding to human exon 32 with precisely the same length of coding sequence as that specified by human exon 32.

The longest open reading frame (795 amino acids) is available in transcripts where the predicted AUG initiation codon is specified by exon 12. This codon is located within a highly conserved strong Kozak consensus sequence (A at -3 position; G at +4; see Kozak, 1996) that is preceded, within exon 12, by termination codons in all three reading frames. In some transcripts, exons 12 and 13 are skipped, leading to the predicted use of an alternative AUG initiation codon specified by exon 14, with loss of the 17 N-terminal amino acids found in the predicted full-length isoform. This alternative sequence is also located within a strong and highly conserved Kozak consensus sequence. Two other alternative AUG translation initiation codons are predicted to be utilised in other transcripts as a result of employing exon 13 where there is an adequate Kozak consensus (A at -3 position and T at +4) or exons 15 plus 16. In the latter case, the first available AUG does not provide an adequate Kozak consensus sequence, raising the possibility of leaky scanning and alternative utilisation of a secondary AUG in a strong Kozak consensus sequence, located over 100 codons downstream within exon 18 (data not shown; see also below).

5'RACE extension with an exon-12-specific primer consistently failed and cDNA clones containing exon 12 sequence (or its counterpart in mouse and pig) lack sequences derived from upstream exons. Transcripts containing exon 12 sequence may, therefore, be primed by

an alternative promoter located in the upstream sequence neighbouring exon 12. The same pattern of alternative translational start sites is apparent in orthologous mouse transcripts (e.g. RIKEN E330010P06 and IMAGE 4913962 cDNA clones—accession nos. BB666308 and BG865829 respectively; see Fig. 2e). The absence of any evidence for linking exons 13 and 15 to upstream exons may mean that transcripts using exons 13 and exons 15+16 are primed from alternative promoters upstream of exons 13 and 15, respectively.

Truncated transcripts include those where different variants of exon 18 constitute a terminal exon, rather than the normal spliced one, as in certain human, cow, pig and chick cDNA clones (Fig. 2d, e). In one variant, designated exon 18B, the exon-18-coding sequence is extended by 24 amino acids (as in IMAGE clone 2095161 and 4593084), but the additional coding sequence is poorly conserved. Because exon 12 is skipped in image 209516 the predicted protein lacks the 17 N-terminal amino acids of full-length isoforms, being translated from the alternative AUG initiation codon specified by exon 14. Another exon 18 variant, 18C, is represented in the IMAGE cDNA 2969158 clone and by pig MI-P-CPI-nwg-b-08-0-UI and cow BP250016B20B10 (BF043307) clones (Fig. 2e). This variant is truncated at the 3' end, but not as a result of using an alternative splice acceptor and without incurring a stop codon. A final transcript variation involves the skipping of exon 19 and inclusion of untranslated exons 21 and 25 (as in IMAGE clone 1847 027). The loss of exon 19 results in a frameshift, permitting 14 novel codons before a premature stop codon is encountered within exon 20, marking the start of a long 3'UTR transcribed from exon 20 through to exon 25 (Fig. 2d).

As a result of the alternative processing events, multiple protein isoforms are predicted in addition to the 795-amino-acid isoform and some are predicted to lack a sequence spanning amino acids 128 and 146 in the full-length isoform, which is highly hydrophobic in available orthologue sequences (see below) and strongly suggested to be a transmembrane sequence (e.g. by the TMpred prediction program; see above). The additional isoforms include: (1) a 320-amino-acid isoform predicted from Image clone 2095161 (lacks the first 17 N-terminal amino acids; residues 313–795 are replaced by a novel 24-aminoacid sequence, viz. VGPVNVIQWFGQYFALFCW-NYMLL); (2) a 595-amino-acid isoform predicted from the AK126272 sequence (the first 15 N-terminal amino acids are replaced by MSRHEIQG; the C-terminal residues 601–795 are replaced by the dipeptide AE); (3) isoforms of 406 or 302 amino acids predicted from the AK131044 sequence (as described above, depending on the location of the initiating methionine, the N-terminal sequence spanning residues 1–182 is replaced by the tetrapeptide MKLR or the sequence from 1–282 is deleted; thereafter, the sequence spanning residues 363–411 is deleted, as is the C-terminal sequence from 633–795); and (4) an isoform predicted from the sequence of Image clone 2969158 and spanning at least 295 amino acids (replacement of C-terminal residues 292–795 by the tetrapeptide LNRL followed possibly by polylysine following the read through into the poly (A) tail). As expected, Northern blot hybridisation confirms substantial heterogeneity in transcript sizes. At least six different transcripts can be discerned in the 1–kb to 10-kb size range (Fig. 3). Of the tissues surveyed, expression was most prominent in placenta and in kidney (fetal and adult). Moderate expression was observed in liver and skeletal muscle.

NAALADL2 appeared during vertebrate evolution and encodes a protein distantly related to N-acetylated alpha-linked acidic dipeptidase and to transferrin receptors

The recent availability of the virtually complete mouse and rat genome sequences has permitted comprehensive comparison of human and rodent *NAALADL2* orthologues. With the exception of some of the 5'UTR exons, the exon-intron organisation has been largely conserved and, like the human gene, the rodent genes span large (ca. 1.3 Mb) genomic regions. Sequence alignment of the coding exons reveals moderately high sequence conservation (see also below) but there is markedly reduced evolutionary conservation of the untranslated exons. Long-range sequence alignments (PipMaker and VISTA; see above) have been particularly helpful in defining orthologous 5' UTR exons. Whereas no rodent counterpart for various of the 5'UTR *NAALADL2* exons are obvious, rodent equivalents can be discerned for human exons 2, 7, 8 and 11 (data not shown). Comparison of the *NAALADL2* transcription unit in human, mouse and rat has also identified a large number of highly conserved regions outside the exons. BLAST analysis suggests that the majority are single-copy sequences.

TBlastN-based analyses reveal that all but the first 100 or so amino acids of the 795-amino-acid isoform showed significant, but low, homology to certain members of the M28 metalloproteinase family (see the MEROPS database at <http://merops.sanger.co.uk/>). The most significant homology (25%–26% sequence identity/45% sequence similarity) was obtained with the protein encoded by *FOLH1*, the folate hydrolase gene (also called N-acetylated alpha-linked acidic dipeptidase, glutamate carboxypeptidase II or prostate-specific membrane antigen), and the *NAALAD2*-encoded N-acetylated alpha-linked acidic dipeptidase 2 (data not shown). In recognition of these homologies and because of the existence of another N-acetylated alpha-linked acidic dipeptidase-like gene, *NAALADL1* (whose predicted protein product shows 25% sequence identity/44% sequence similarity to the region spanning amino acids 115–734 of the 795-amino-acid query), the gene in the current study has been provisionally accorded the symbol *NAALADL2*. Additional significant homologies were found with the *TRF2*-encoded transferrin receptor 2 (26% sequence identity/46% sequence similarity over amino acids 241–650 of the 795-amino-acid query) and with the *TRFC*-encoded transferrin receptor (27% sequence identity/46% sequence similarity over amino acids 263–593 of the *NAALADL2* protein).

The evolutionary origins appear to be comparatively recent. *NAALADL2* orthologues are not easily identified in invertebrates. Distantly related proteins can be found in fish but database screens do not provide compelling evidence for an orthologue in zebrafish (*Brachydanio rerio*), although some small-length scaffold sequences in the pufferfish (*Takifugu rubripes*) are consistent with a reasonably closely related sequence (data not shown). There is clear evidence, however, for a chick orthologue: the inferred protein sequences of the chick EST sequences shown in Fig. 2d show respectively 64% sequence identity over a region spanning amino acids 191–313 of the 795-amino-acid isoform of human *NAALADL2*, and 75% sequence identity over the C-terminal region from amino acid number 640–795. The full-length (795 amino acid) human and mouse *NAALADL2* orthologues show 74% sequence identity (data not shown).

Tissue expression and mutation screening analyses do not support a role for *NAALADL2* in the pathogenesis of CdLS

Analysis of the presumed promoter region of *NAALADL2* by means of the rVISTA program identifies a number of transcription-factor-binding sites conserved in multiple arrays between human and rodents, viz. target motifs for an assortment of developmentally regulated transcription factors including various PAX proteins (data not shown). In order to examine *NAALADL2* developmental expression, we hybridised a *NAALADL2*-specific antisense RNA probe against human embryonic tissue sections representing Carnegie stages 16–21 (CS16–CS21). A weak expression signal was observed in the mesonephros at CS16 (Fig. 4a). At later stages, moderate expression was observed in the metanephros and the endodermal lining of the stomach (Fig. 4b) and in the pancreas (Fig. 4c); a potential signal, albeit weak, was present in the midgut-loops (bottom of Fig. 4c). Strong expression was evident in the endodermal lining of the duodenum (Fig. 4c). No expression was observed in the developing brain or limbs of the stages examined.

Although the associated embryonic expression pattern does not make *NAALADL2* a compelling CdLS candidate gene, we sought evidence for patient-specific mutations. Genomic DNA samples from a panel of 80 CdLS individuals covering the broad clinical phenotypic range and from some unaffected individuals were SSCP/heteroduplex-screened for small nucleotide changes in all *NAALADL2* exons. Additionally, a fluorescence-based MAPH screen surveyed large mutations by assaying changes in exon copy number (see Introduction). Both analyses failed to reveal evidence for patient-specific mutations.

Discussion

The molecular basis of the pathogenesis seen in CdLS has been an enduring enigma. The extreme rarity of families with multiple affected individuals has made it difficult to establish a subchromosomal localisation for the gene(s) underlying this disorder. As a result, much attention has focused on associated chromosomal breakpoints. A considerable variety of such breakpoints has been found, together with a lack of consistency in subchromosomal localization (see Kousseff et al. 1994). Whereas many of the breakpoints may be coincidental occurrences that do not contribute to pathogenesis, the de novo 3q26.3 translocation breakpoint was considered to be potentially significant because of the apparent phenotypic overlap with partial trisomy 3q cases involving the 3q26-q27 region (see Introduction).

The study reported here, however, now shows that the 3q26.3 breakpoint severs a giant gene that appears to have arisen relatively recently during evolution. The comparative evolutionary novelty of this gene and its tissue expression pattern do not support its direct involvement in the pathogenesis of CdLS and affected individuals do not appear to have deleterious mutations in this gene. These findings raise two new questions: what is the functional significance of *NAALADL2*, and could the 3q26.3 breakpoint still be associated with pathogenesis?

The predicted *NAALADL2* protein is significantly related to glutamate carboxypeptidase II (also known as N-acetylated alpha-linked acidic dipeptidase, folate hydrolase and prostate-

specific membrane antigen), a membrane-bound enzyme with a large extracellular region that contains a catalytically active domain (Rawlings and Barrett 1997). Like this enzyme, the NAALADL2 protein has what appears to be a transmembrane domain close to its N-terminus suggesting that the full-length NAALADL2 isoform would have a large C-terminal extracellular region showing homology to N-acetylated alpha-linked acidic dipeptidase and transferrin receptors. The strong gut expression pattern is also consistent with a role in the processing of dietary components. However, alignment of the NAALADL2 protein sequence with other members of the M28B metallopeptidase family reveals that it often does not possess favoured amino acids at certain key positions that are highly conserved and that are envisaged to be critically important for metallopeptidase function (because of a role in the active site or in metal binding; for sequence alignments, see Goodyer et al. 2001; Merops database, <http://merops.sanger.co.uk/>). Although significantly related to N-acetylated alpha-linked acidic dipeptidase, the NAALADL2 protein is likely to have developed a somewhat different function. As noted above, additional functional diversity may result from the variable presence in isoforms of the transmembrane region.

The functional significance of the extraordinary diversity in the 5'UTR is not clear, although parallels have been seen in genes encoding other transmembrane proteins, e.g. growth hormone receptor, neuropeptide YY₅ receptor (Goodyer et al. 2001; Parker and Xia 1999). Alternative 5' UTR splicing in such systems has often been viewed as regulating gene expression but evolutionary conservation of the 5'UTR exons of *NAALADL2* appears to be limited, as has often been found in interspecies comparative analyses (Nurtdinov et al. 2003).

Although the current study strongly suggests that *NAALADL2* is not involved in the pathogenesis of CdLS, the question of whether the 3q26.3 translocation breakpoint could nevertheless contribute to pathogenesis remains unresolved. Because of position effects, genes underlying some disorders have sometimes been found many hundreds of kilobases from the translocation breakpoint (for a recent review, see Kleinjan and van Heyningen 1998). Given the huge extent of many of its introns, the possibility of unidentified genes being embedded within *NAALADL2* has to be considered. Certainly, comparative sequence alignments of orthologues show a large number of highly conserved sequences within the *NAALADL2* transcriptional unit in addition to the known exons. Whereas several such sequences may not produce a functional RNA or protein product, database analysis shows that many are associated with cDNAs/ESTs and evidence exists for several spliced transcripts being located within the *NAALADL2* transcription unit, all lacking an obvious translational reading frame (data not shown). There is increasing appreciation of the functional significance of non-coding RNAs, many of which are now known to play important roles in development. Although none of the additional spliced transcripts is severed by the breakpoint, a position effect on gene expression is possible and so the involvement of these transcripts in pathogenesis cannot be formally excluded at this stage. During the course of these analyses, it has become apparent that marker segregation does not support a location for the underlying gene in at least some of the rare families showing vertical transmission for CdLS (Krantz et al. 2001). Whereas the highly specific features of the syndrome have often been taken to suggest locus homogeneity, the possibility of locus heterogeneity cannot, however, be discounted.

For the sake of completeness, we have also mapped the 17q23.1 breakpoint in the present study (see Fig. 1b) and shown that it is located at the 5' end of a novel gene of unknown function (data not shown). However, there have been various reports of individuals who are monosomic for the 17q23.1 region (e.g. the 17q21.3-q24.3 region; see Dallapiccola et al. 1993) but who do not show CdLS; on this basis, the 17q breakpoint may be viewed as having a low probability of contributing to pathogenesis. Accordingly, attention also needs to be paid to other chromosomal regions that have been associated with CdLS. Prioritisation of regions for study is not helped, however, by the lack of any consistency in associated chromosome regions and thus we have also carried out parallel studies involving the mapping of other chromosome breakpoints in additional affected individuals with apparently balanced de novo translocations.

Should these and independent linkage studies fail to identify a specific chromosomal location, then position-independent strategies will need to be employed. One such approach, comparative genomic hybridisation (CGH), is suggested by the possibility that the mental retardation and multiple tissue defects in CdLS may arise as a result of microdeletion or microduplication, instead of a single gene defect. In collaboration with Dr. Nigel Carter and colleagues at the Wellcome Trust Sanger Institute, we carried out a genome-wide microarray-based CGH screen of a panel of 3,000 BACs on samples from CdLS patients and controls but no evidence was found for a consistent pattern of microdeletion/microduplication (data not shown).

Acknowledgments

We thank the UK Community Fund and the UK Medical Research Council for providing funding for this project. We are indebted to Heike Figler and Nigel Carter for supervising the BAC microarray CGH screen and to the MRC-Wellcome Human Developmental Biology Resource for the supply of human embryonic tissue samples. We are particularly grateful to many CdLS individuals, their families and additional referring clinicians for their generosity in providing access to CdLS blood samples and associated phenotype data, and to Alan Peaford and colleagues at the UK and Ireland Cornelia de Lange Foundation for their support.

References

- Allanson JE, Hennekam RC, Ireland M. de Lange syndrome: subjective and objective comparison of the classical and mild phenotypes. *J Med Genet.* 1997; 34:645–650. [PubMed: 9279756]
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215:403–410. [PubMed: 2231712]
- Anand R, Riley JH, Butler R, Smith JC, Markham AF. A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage. *Nucleic Acids Res.* 1990; 18:1951–1956. [PubMed: 2186372]
- Aqua MS, Rizzu P, Lindsay EA, Shaffer LG, Zackai EH, Overhauser J, Baldini A. Duplication 3q syndrome: molecular delineation of the critical region. *Am J Med Genet.* 1995; 55:33–37. [PubMed: 7702094]
- Bullen PJ, Robson SC, Strachan T. Human post-implantation embryo collection: medical and surgical techniques. *Early Hum Dev.* 1998; 51:213–221. [PubMed: 9692791]
- Chumakov I, Rigault P, Guillou S, Ougen P, Billaut A, Guasconi G, Gervy P, LeGall I, Soularue P, Grinas L, Bougueleret L, Bellanné-Chantelot C, Lacroix B, Barillote E, Gesnouin P, Pook S, Vaysseix G, Frelat G, Schmitz A, Sambucy J-L, Bosch A, Estivill X, Weissenbach J, Vignal A, Riethman H, Cox D, Patterson D, Gardiner K, Hattori M, Sakaki Y, Ichikawa H, Ohki M, Le Paslier D, Heilig R, Antonarakis S, Cohen D. Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature.* 1992; 359:380–387. [PubMed: 1406950]

- Dallapiccola B, Mingarelli R, Digiglio C, et al. Interstitial deletion del (17)(q21.3q23 or 24.2) syndrome. *Clin Genet.* 1993; 43:54–55. [PubMed: 8462199]
- Down T, Hubbard T. Computational detection and location of transcription start sites in mammalian genomic DNA. *Genome Res.* 2002; 12:458–461. [PubMed: 11875034]
- Dubchak I, Brudno M, Loots GG, Pachter L, Mayor C, Rubin EM, Frazer KA. Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res.* 2000; 10:1304–1306. [PubMed: 10984448]
- Fiegler H, Carr P, Douglas EJ, Burford DC, Hunt S, Scott CE, Smith J, Vetrie D, Gorman P, Tomlinson IP, Carter NP. DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosom Cancer.* 2003; 36:361–374. [PubMed: 12619160]
- Gingrich JC, Boehrer DM, Garnes JA, Johnson W, Wong BS, Bergmann A, Eveleth GG, Langlois RG, Carrano AV. Construction and characterization of human chromosome 2-specific cosmid, fosmid, and PAC clone libraries. *Genomics.* 1996; 32:65–74. [PubMed: 8786122]
- Goodyer CG, Zogopoulos G, Schwartzbauer G, Zheng H, Hendy GN, Menon RK. Organization and evolution of the human growth hormone receptor gene 5' flanking region. *Endocrinology.* 2001; 142:1923–1934. [PubMed: 11316758]
- Herring CD, Chevillard C, Johnston SL, Wettstein PJ, Riblet R. Vector-hexamer PCR isolation of all insert ends from a YAC contig of the mouse *Igh* locus. *Genome Res.* 1998; 8:673–681. [PubMed: 9647641]
- Ioannou P, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* 1994; 6:84–89. [PubMed: 8136839]
- Ireland M, English C, Cross I, Houlby WT, Burn J. A de novo translocation t(3;17)(q26.3;q23.1) in a child with Cornelia de Lange syndrome. *J Med Genet.* 1991; 28:639–640. [PubMed: 1956066]
- Ireland M, Donnai D, Burn J. Brachmann-de Lange syndrome. Delineation of the clinical phenotype. *Am J Med Genet.* 1993; 47:959–964. [PubMed: 8291539]
- Ireland M, English C, Cross I, Lindsay S, Strachan T. Partial trisomy 3q and the mild Cornelia de Lange syndrome phenotype. *J Med Genet.* 1995; 32:837–838. [PubMed: 8558571]
- Jackson L, Kline AD, Barr MA, Koch S. de Lange syndrome: a clinical review of 310 individuals. *Am J Med Genet.* 1993; 47:940–946. [PubMed: 8291537]
- Kleinjan DJ, van Heyningen V. Position effect in human genetic disease. *Hum Mol Genet.* 1998; 7:1611–1618. [PubMed: 9735382]
- Kousseff BG, Newkirk P, Root AW. Brachmann-de Lange syndrome. 1994 update. *Arch Pediatr Adolesc Med.* 1994; 148:749–755. [PubMed: 8019632]
- Kozak M. Interpreting cDNA sequences: some insights from studies on translation. *Mamm Genome.* 1996; 7:563–574. [PubMed: 8679005]
- Kozma C. Autosomal dominant inheritance of Brachmann-de Lange syndrome. *Am J Med Genet.* 1996; 66:445–448. [PubMed: 8989465]
- Krajewska-Walasek M, Chrzanowska K, Tylki-Szymanska A, Bialecka M. A further report of Brachmann-de Lange syndrome in two sibs with normal parents. *Clin Genet.* 1998; 47:324–327. [PubMed: 7554368]
- Krantz ID, Tonkin E, Smith M, Devoto M, Bottani A, Simpson C, Hofreiter M, Abraham V, Jukofsky L, Conti BP, Strachan T, Jackson L. Exclusion of linkage to the *CDL1* gene region on chromosome 3q26.3 in some familial cases of Cornelia de Lange syndrome. *Am J Med Genet.* 2001; 101:120–129. [PubMed: 11391654]
- Lako M, Lindsay S, Bullen P, Wilson DI, Robson SC, Strachan T. A novel mammalian *Wnt* gene, *WNT8B*, shows brain-restricted expression in early development, with sharply delimited expression boundaries in the developing forebrain. *Hum Mol Genet.* 1998; 7:813–822. [PubMed: 9536085]
- Larin Z, Monaco AP, Lehrach H. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc Natl Acad Sci USA.* 1991; 88:4123–4127. [PubMed: 2034658]

- Loots GG, Ovcharenko I, Pachter L, Dubchak I, Rubin EM. rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res.* 2002; 12:832–839. [PubMed: 11997350]
- Lopez-Rangel E, Dill FJ, Hrynchak MA, Van Allen MI. Partial duplication of 3q (q25.1 q26.1) without the Brachmann-de Lange phenotype. *Am J Med Genet.* 1993; 47:1068–1071. [PubMed: 8291525]
- Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I. VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics.* 2000; 16:1046–1047. [PubMed: 11159318]
- McConnell V, Brown T, Morrison PJ. An Irish three-generation family of Cornelia de Lange syndrome displaying autosomal dominant inheritance. *Clin Dysmorphol.* 2003; 12:241–244. [PubMed: 14564211]
- McKenney RR, Elder FF, Garcia J, Northrup H. Brachmann-de Lange syndrome: autosomal dominant inheritance and male-to-male transmission. *Am J Med Genet.* 1996; 66:449–452. [PubMed: 8989466]
- Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc Natl Acad Sci USA.* 1989; 86:6686–6690. [PubMed: 2771952]
- Nurtdinov RN, Artamonova II, Mironov AA, Gelfand MS. Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum Mol Genet.* 2003; 12:1313–1320. [PubMed: 12761046]
- Ochman H, Gerber AS, Hartl DL. Genetic applications of an inverse polymerase chain reaction. *Genetics.* 1988; 120:621–623. [PubMed: 2852134]
- Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, de Jong PJ. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.* 2001; 11:483–496. [PubMed: 11230172]
- Parker EM, Xia L. Extensive alternative splicing in the 5′-untranslated region of the rat and human neuropeptide YY₅ receptor genes regulates receptor expression. *J Neurochem.* 1999; 73:913–920. [PubMed: 10461880]
- Rawlings ND, Barrett AJ. Structure of membrane glutamate carboxypeptidase. *Biochim Biophys Acta.* 1997; 1339:247–252. [PubMed: 9187245]
- Rizzu P, Haddad BR, Vallcorba I, Alonso A, Ferro MT, Garcia-Sagredo JM, Baldini A. Delineation of a duplication map of chromosome 3q: a new case confirms the exclusion of 3q25-q26.2 from the duplication 3q syndrome critical region. *Am J Med Genet.* 1997; 68:428–432. [PubMed: 9021016]
- Russell KL, Ming JE, Patel K, Jukofsky L, Magnusson M, Krantz ID. Dominant paternal transmission of Cornelia de Lange syndrome: a new case and review of 25 previously reported familial recurrences. *Am J Med Genet.* 2001; 104:267–276. [PubMed: 11754058]
- Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.* 2000; 10:577–586. [PubMed: 10779500]
- Steinbach P, Adkins WN Jr, Caspar H, Dumars KW, Gebauer J, Gilbert EF, Grimm T, Habedank M, Hansmann I, Herrmann J, Kaveggia EG, Langenbeck U, Meisner LF, Najafzadeh TM, Opitz JM, Palmer CG, Peters HH, Scholz W, Tavares AS, Wiedeking C. The dup(3q) syndrome: report of eight cases and review of the literature. *Am J Med Genet.* 1981; 10:159–177. [PubMed: 7315873]
- Tonkin E, Hagan DM, Li W, Strachan T. Identification and characterisation of novel mammalian homologues of *Drosophila* polyhomeotic permits new insights into relationships between members of the polyhomeotic family. *Hum Genet.* 2002; 111:435–442. [PubMed: 12384788]
- Van Allen MI, Filippi G, Siegel-Bartelt J, Yong SL, McGillivray B, Zuker RM, Smith CR, Magee JF, Ritchie S, Toi A, Reynolds JF. Clinical variability within Brachmann-de Lange syndrome: a proposed classification system. *Am J Med Genet.* 1993; 47:947–958. [PubMed: 8291538]
- White S, Kalf M, Liu Q, Villerius M, Engelsma D, Kriek M, Vollebregt E, Bakker B, van Ommen GJ, Breuning MH, den Dunnen JT. Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. *Am J Hum Genet.* 2002; 71:365–374. [PubMed: 12111668]

Wilson GN, Dasouki M, Barr M Jr. Further delineation of the dup(3q) syndrome. *Am J Med Genet.* 1985; 22:117–123. [PubMed: 4050847]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

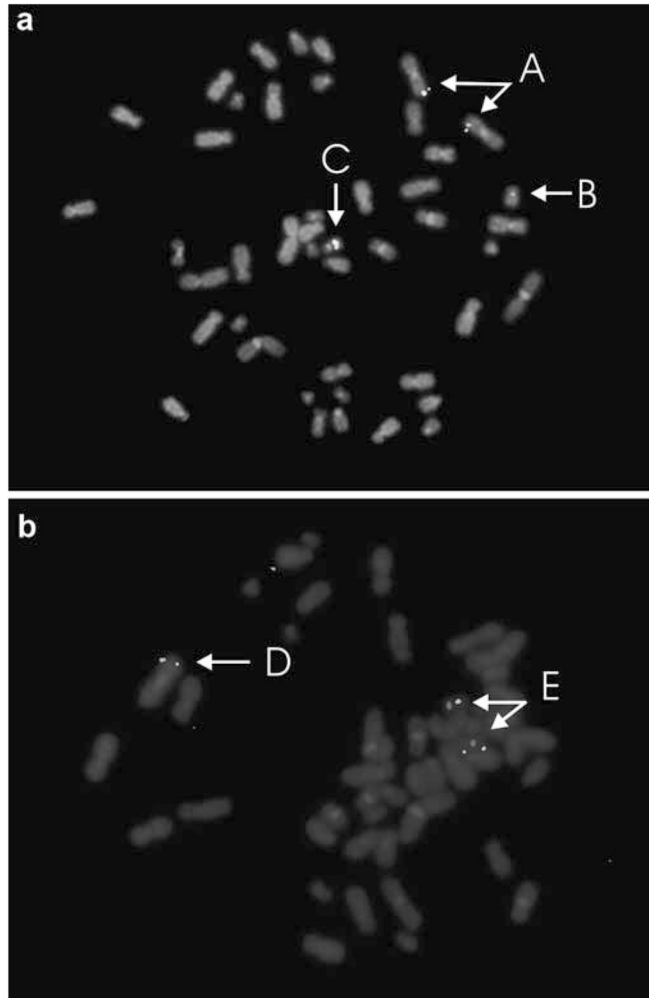


Fig. 1.

a, b FISH mapping of chromosome breakpoints in the t(3;17) cell line. **a** FISH identifies BAC RP11-827e14 (*green*) as spanning the 3q26.3 breakpoint. *Red* A chromosome 17 centromere-specific probe. Indicated chromosomes are as follows: *A* chromosomes 3 and der(3) t(3;17)(q26.3;q23.1), *B* chromosome 17, *C* der(17) t(3;17) (q26.3;q23.1). **b** FISH identifies cosmid LA173B3 (*green*) as spanning the 17q23.1 breakpoint. *Red* A chromosome 17 centromere-specific probe. Indicated chromosomes are as follows: *D* der (3) t(3;17) (q26.3;q23.1), *E* chromosome 17 and der(17) t(3;17) (q26.3;q23.1). the original colour FISH images can be viewed at <http://www.ncl.ac.uk/ihg/cdls>

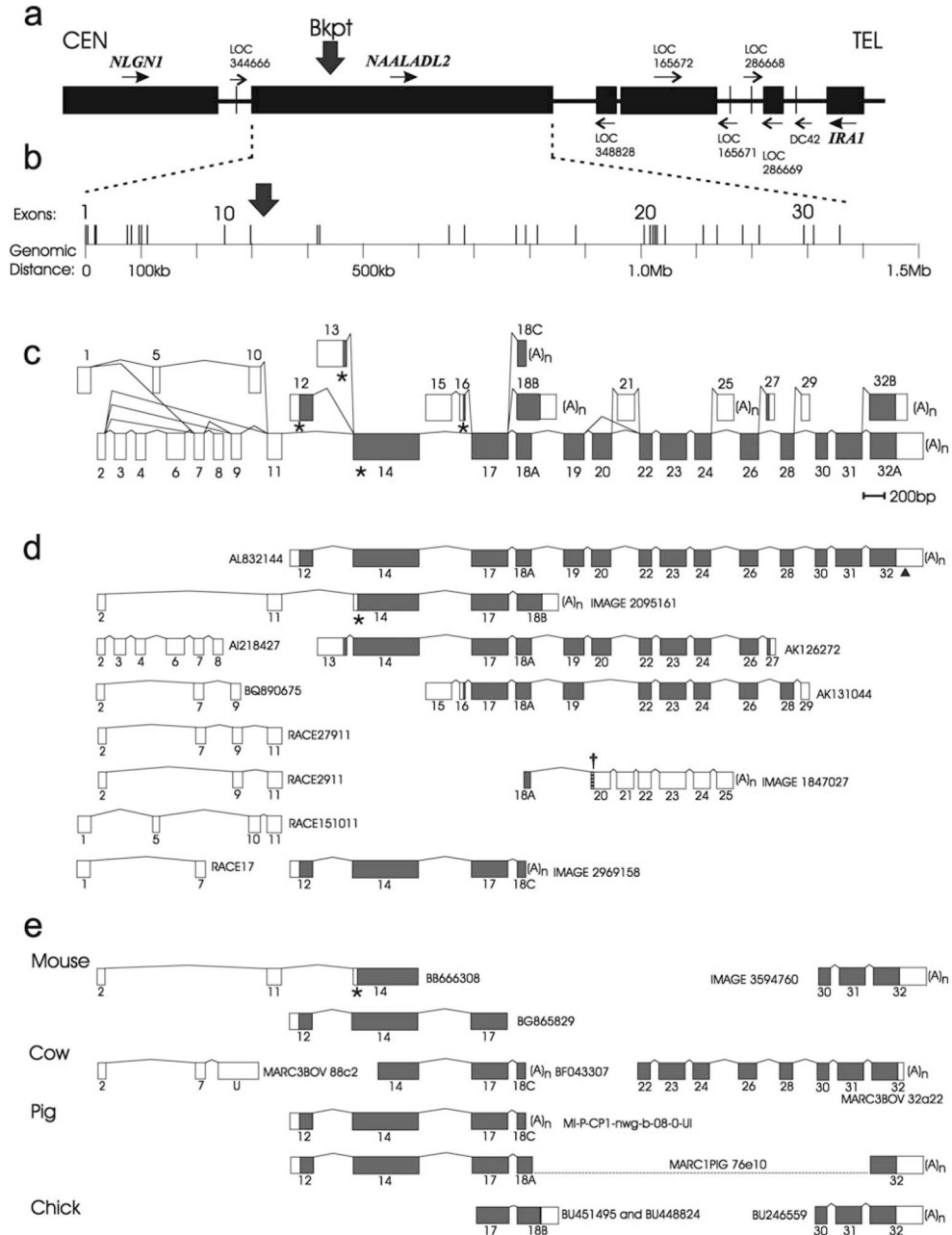


Fig. 2.
a–e Genomic structure, exon-intron organisation and transcript analysis of the *NAALADL2* gene. **a** Genomic organisation of human *NAALADL2* relative to the flanking genes neurologigin 1 (*NLGN1*) and the orthologue of mouse *Ira1*. These three genes and the position of additional loci predicted by computational analysis (see <http://www.ncbi.nih.gov/mapview>) are illustrated by *filled boxes* according to genomic size. *Horizontal arrows* Transcriptional direction, *solid downward-pointing arrow* position of the 3q26.3 breakpoint. **b** *NAALADL2* exon organisation. Intron sizes are to scale. **c** Exon structure and alternative splicing events in *NAALADL2*. *Boxes* Exons, *black fill-in* coding sequence, *absence of fill-*

in untranslated sequences. Exons are drawn to scale with the exception of the 3'UTR following exon 32 (drawn to 1/10th scale). The 2.4-kb 3'UTR (exon 32A) and a shorter 1-kb form (exon 32B) are shown with the alternative polyadenylation site producing the short form (found in AL832931) indicated by a *filled triangle* at the corresponding position in the exon 32 sequence of AL832144 in **d**. *Asterisks* Predicted alternative translational starts. The *asterisk below* exon 14 (see also **d**, **f**) denotes the predicted alternative start used when this exon is spliced directly to exon 11 (when exon 14 is spliced instead to either exon 12 or exon 13, the entire exon contributes to the open reading frame). Exon 18 has been observed in three forms: exon 18A (spliced form), 18B (extended form with a stop codon and 3'UTR) and exon 18C (abbreviated form, with a terminating polyA tail [(A)_n], but without a stop codon or 3'UTR). **d** cDNA clones and 5'RACE products used to determine the exon structure of *NAALADL2*. Accession numbers are given for cDNA sequences obtained by in silico screens. Clone names are given for cDNAs where the full-length sequence was established by in-house sequencing. The *dagger* above exon 20 indicates the presence of a stop codon, resulting from a change in reading frame when this exon is spliced directly downstream of exon 18. **e** cDNA clones from mouse, cow, pig and chick contain sequences aligning with human exons as indicated. The exon organisation is largely conserved at the genomic level in mouse but was unclear in cow, pig and chick at the time of writing

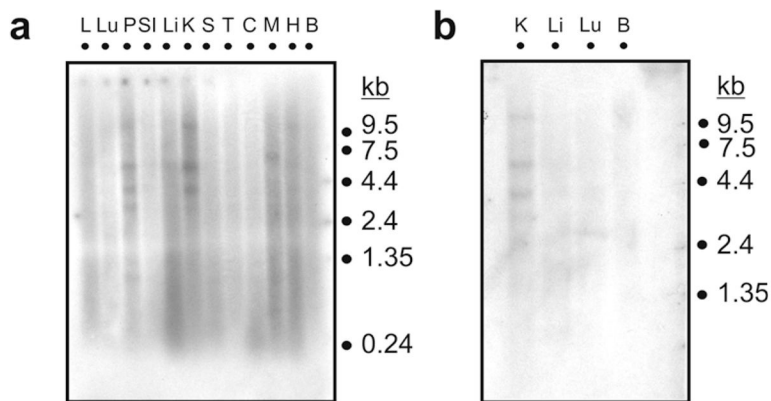


Fig. 3.

a, b Northern blot hybridisation identifies multiple *NAALADL2* transcripts in adult and fetal tissues. Human adult (**a**) and fetal (**b**) Northern blots containing poly (A)⁺ RNA and normalised to β -actin (Clon-tech) were hybridised with a 581-bp probe *NAALADL2* cDNA probe. *Right* Size markers. Sources of tissue RNA are abbreviated as follows: *L* peripheral blood leucocytes, *Lu* lung, *P* placenta, *SI* small intestine, *Li* liver, *K* kidney, *S* spleen, *T* thymus, *C* colon, *M* skeletal muscle, *H* heart, and *B* brain

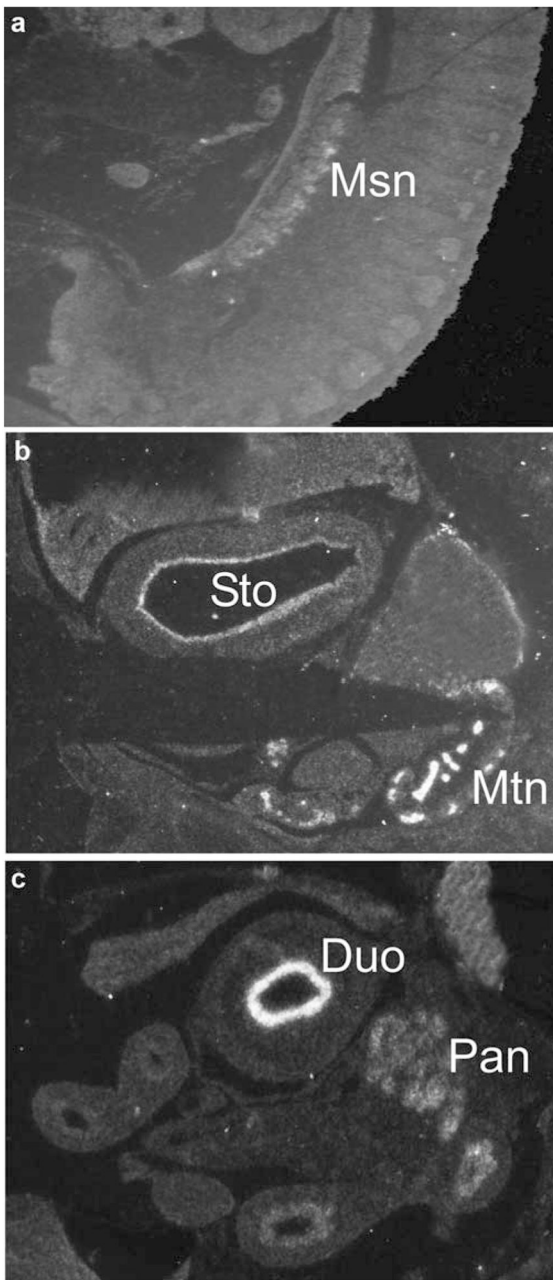


Fig. 4. **a–c** Embryonic expression of *NAALADL2*. Dark-field microscopy images following hybridisation of a ^{35}S -labelled antisense *NAALADL2* probe to sagittal embryonic sections representing Carnegie stages 16 (**a**) and 21 (**b**, **c**). *Msn* mesonephros, *Sto* stomach, *Mtn* metanephros, *Duo* duodenum, *Pan* pancreas