

Genomic Structure of a Novel LIM Domain Gene (ZNF185) in Xq28 and Comparisons with the Orthologous Murine Transcript

Nina S. Heiss,* Gernot Gloeckner,† Dietmar Bächner,‡ Petra Kioschis,* Sabine M. Klauck,* Bernd Hinzmann,† André Rosenthal,† Gail E. Herman,§[¶] and Annemarie Poustka*¹

*Department of Molecular Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany; †Department of Molecular Genome Analysis, Institut für Molekulare Biotechnologie, Beutenbergstraße 11, 07745 Jena, Germany; ‡Department of Gene Regulation and Differentiation, Gesellschaft für Biotechnologische Forschung, Mascheroder Weg 1, 38124 Braunschweig, Germany; §Department of Molecular and Human Genetics and Pediatrics, Baylor College of Medicine, Houston, Texas 77030; and [¶]Children's Hospital Research Foundation, Department of Pediatrics, The Ohio State University, Columbus, Ohio 43210

Received December 23, 1996; accepted May 7, 1997

Construction of a transcript map in the DXS52 region in Xq28 had previously led to the isolation of a cDNA with a LIM zinc finger domain in the carboxyl terminus. In parallel, the orthologous murine transcript was isolated from the syntenic region. The human and mouse cDNAs have been designated ZNF185 and *Zfp185*, respectively. By integrating the cDNA sequence with the cosmid-derived genomic sequence the exon-intron structure of the 3' end of the ZNF185 gene was resolved. Comparative sequence analyses of the human genomic sequence with the full-length murine cDNA facilitated prediction of the 5' end of the gene. The selective expression of three transcripts corresponding to the ZNF185 gene and a related gene was shown by Northern and Southern blots. *In situ* hybridizations revealed a nonubiquitous and stage-specific expression of *Zfp185*, especially in differentiating connective tissue. Since LIM proteins regulate cellular proliferation and/or differentiation by diverse mechanisms, and some have directly been associated with disease, conceivably ZNF185 may represent a candidate for a disease-causing gene linked to Xq28. Knowledge of the genomic structure will permit detailed mutation analyses. © 1997 Academic Press

INTRODUCTION

As part of a long-range positional cloning approach in Xq28, transcript maps were established in the 300-kb region around the G6PD gene (Sedlacek *et al.*, 1993), the 900-kb candidate region for myotubular myopathy (MTM-1; Kioschis *et al.*, 1996), the 1200-kb candidate region for the hereditary form of incontinentia pigmenti (IP2; Rogner *et al.*, 1996), and the 700 kb around

the DXS52 loci (Heiss *et al.*, 1996). Although the identification of novel genes linked to Xq28 is ongoing, an association with a particular disease is lacking for most. Unclassified disease-causing genes include the genes for Waisman syndrome (Gregg *et al.*, 1991), otopalato-digital syndrome (Biancalana *et al.*, 1991), Geminine syndrome (Zuffardi *et al.*, 1982), X-linked mental retardation (MRX3; Gedeon *et al.*, 1991; Nordström *et al.*, 1992), the X-linked dominant form of chondrodysplasia punctata (CDPX2 or Happle syndrome; Traupe *et al.*, 1992), dyskeratosis congenita (DKC; Connor *et al.*, 1986; Arngrimsson *et al.*, 1993), neuronal chronic idiopathic X-linked intestinal pseudoobstruction (CIIP; Auricchio *et al.*, 1996), and periventricular heterotopia (Eksioglu *et al.*, 1996). The candidate regions for most of these diseases cover several megabases and span the polymorphic DXS52 loci. The transcript map of the DXS52 region (Heiss *et al.*, 1996) provided scope for the further characterization of single genes and allowed first associations to be made among tissue expression, homologies to known protein domains, possible function, and disease. Of the novel cDNAs located in the DXS52 region, XAP105 was interesting because of its homology to a murine cDNA that maps to the syntenic region on the mouse X chromosome. Both the human and the mouse cDNAs contain a LIM zinc finger domain in the carboxyl terminus and have been designated ZNF185 and *Zfp185*, respectively.

LIM domain proteins represent a diverse family of regulatory proteins occurring in mammals, birds, amphibians, fish, yeast, and *Drosophila* (Dawid *et al.*, 1995). The term "LIM" stems from the first letters of the *Caenorhabditis elegans lin11* gene (Freyd *et al.*, 1990), the rat *isl1* gene (Karlsson *et al.*, 1990), and the *C. elegans mec3* gene (Way and Chalfie, 1988), in which LIM domains were originally identified. The conserved cysteine and histidine residues that make up a LIM domain form a C₂HC and a C₄ motif, each of which

¹ To whom correspondence should be addressed. Telephone: (49) 6221 424741. Fax: (49) 6221 423454. E-mail: N.Heiss@DKFZ-Heidelberg.de.

coordinately binds one zinc ion (Sanchez-Garcia and Rabbitts, 1994). Strengthened by findings that LIM proteins are distributed in the cytoplasm or nucleus of cells, the assumption that LIM proteins carry out regulatory functions by protein-protein interactions rather than by interacting directly with DNA is generally favored (Perez-Alvarado *et al.*, 1994; Schmeichel and Beckerle, 1994). Although at present no consistent LIM protein dimerization partners are known, accumulating evidence suggests that these proteins regulate cellular proliferation and differentiation through diverse mechanisms (Dawid *et al.*, 1995). This is reflected by their involvement in a wide range of cellular functions including transcriptional activation (*isl-1*; Karlsson *et al.*, 1990), somatic patterning (*lin-11*, *mec3*; Freyd *et al.*, 1990), focal cell adhesion (*zyxin*; Sadler *et al.*, 1992), and immediate-early response to serum stimulation (CRP; Wang *et al.*, 1992; Feuerstein *et al.*, 1994). Conceivably, when mutated, LIM proteins may play a role in the etiology of hereditary and somatic diseases.

Currently no direct association between the novel ZNF185 LIM domain gene and a specific disease linked to Xq28 is apparent. Although it remains open to speculate about its potential role in the causation of disease, more information was gained by comparing the human and the orthologous mouse cDNAs at the sequence and expression levels. Furthermore, the genomic structure of the ZNF185 gene was determined.

MATERIALS AND METHODS

Sequencing and database comparisons. For the genomic sequencing of cosmids Qc11F9, Qc10G10, LA1733, and Qc17B8 (Heiss *et al.*, 1996), the DNA was isolated by using the standard alkaline lysis method and purified on CsCl gradients (Radloff *et al.*, 1967). A total of 10 μ g cosmid DNA was sonified, and protruding ends were subsequently removed by treating the DNA with 30 U mung bean nuclease for 10 min at 30°C in 1 \times nuclease buffer (New England Biolabs). After phenol extraction and ethanol precipitation the DNA was size fractionated on a 0.8% agarose gel. DNA fragments ranging in size from 1.2 to 1.6 kb were excised from the gel, purified with Qiaex (Qiagen), and subcloned into the *Sma*I site of the M13mp18 vector. Prior to carrying out the sequencing reaction, the DNA templates were prepared by using the Triton preparation method. For the sequencing of cloned cDNAs, the plasmid DNA was extracted using the Bio Robot 9600 and Qiawell Ultra System (Qiagen). All sequence reactions were carried out with the *Taq* DyeDeoxy Terminator Cycle Sequencing kit (Perkin-Elmer, ABI) and run on the Applied Biosystems 377 automated sequencer. The raw data were collected and assembled using the XGAP program (STADEN package). Exon predictions were made using the XGRAIL 1.3c, XPOUND, and FEX-HB programs. In addition, the DNA and protein analysis programs of the Heidelberg Unix Sequence Analysis Resources (HUSAR) were used.

Synthesis of cDNA templates for verification of exons. Prior to cDNA synthesis, 5 μ g total RNA and 33 ng oligo(dT18) primer or 6 ng random hexamer primer per microgram of RNA was denatured at 70°C for 5 min. Reverse transcription was carried out in 1 \times first-strand buffer containing 0.01 mM DTT, 10 mM each dNTP, and 200 U Superscript II (Gibco, BRL) for 45 min at 42°C. The mixture was then incubated at 55°C for 10 min, and remaining RNA was removed with 2 U RNaseH (Gibco, BRL) for 10 min at 55°C.

PCR analyses. Exons were PCR amplified in 50- μ l volumes. The cDNA templates were initially denatured for 2 min at 94°C. This was followed by 35 cycles at 94°C, 30 s, 60°C, 30 s, 72°C, 30 s and terminated with a final extension of 10 min at 72°C. To exclude the possibility that failure to amplify 5' exons was owing to incompatible primers, these were tested at the genomic level. To obtain fragments of adequate length, most exons were amplified by combining forward and reverse primers of adjacent exons. Primer sequences were:

GRAIL exonAF	CCTGGGCCATCAGGAGAACTTCC
GRAIL exonAR	GGAGATTCTCCTGATGGCCAGG
GRAIL exonBF	GCAAGCGCTTAGTCTGTGTAAGGC
GRAIL exonBR	GCTCATGTCCGGCTGCCTCC
GRAIL exonCF	GGCACAGGACGTGTGCCCTGC
GRAIL exonCR	GCGTGCTCGCCGAGGTGCACAGG
GRAIL exonEF	GAAGTTCATGAGGACCTCGGACAGG
GRAIL exonER	GCTCTGTGGACAATGTAACATCC
GRAIL exonFF	GGAAAGCTTTTCTATCGAGGAGG
GRAIL exonFR	GCACAGGGCTCATTGTTTCAGG
Exon 5F	GGTGGCAGCCTATGACAGGAAAGC
Exon 5R	GGAGAATGGCACCACCTCCTCC
Exon 6F	GGTGTCTGAGGAGGACAGCTCCC
Exon 6R	CCGCTGACAGGACGTAGGAGTGC
Exon 7F	TACCCAGGAGACACAGGCACCG
Exon 7R	CGATAAACGGTGCCTGTGTCTCC
Exon 8F	CCTTCTGAGAAGAGCAGGACCC
Exon 8R	AGGAGTGGATCTTGCCAGGACAGG
Exon 9F	CAAGAGGTGAGGAAATTGTCCGC
Exon 9R	CCCTGGGTGTCAGGATCTGC
Exon 10F	CCCTGCTGATAGGAAGAGCAACAGC
Exon 10R	CCTTTGGGTCTGCCCTTTGCATCC
Exon 11F	CCTTGGCTGATTGTGAGGGGAAGG
Exon 11R	CAACTCTCTCGTCTGACAAAGTTGC
Exon 12F	GCGTGTTGACTGATTTTGAGGG
Exon 12R	GCTGGTGCAGCTCTCTCGTCTGACG
Exon 13F	GGTGAGGAGCCCCTCGAGCTGC
Exon 13R	GCTGGAGTCCAATTCATCTGC
Exon 14F	GGGATTCTCTTCGTGAAGGAGTACG
Exon 14R	GCTATAGCGTGCAGATACTGGC
Exon 15F	CGTCAGCAGCATTGAGGACTATTCC
Exon 15R	CCTCTCAGAGTATGGAGTGCTGCC
Exon 16F	CAACTGGAGGGATCTGTACTTACTCC
Exon 16R	GCAATATTCATGGCAGCAGATACC
Exon 17F	CGGTAACCCCGATGTCTGTGCC
Exon 17R	CCCACAGTGAATGGTGTACCG

Probe labeling, competition, and filter hybridizations. Fifty to 100 ng DNA was labeled to a specific activity of 5×10^6 to 10×10^6 cpm by random priming, isopropanol precipitated, and then subjected to at least 2 h of competition with 25 μ g human Cot-1 DNA (Gibco, BRL) at 65°C. Hybridizations were carried out at 65°C for 16 h in Church solution, and filters were washed twice at 65°C for 20 min in 2 \times SSC containing 0.2% SDS. For Northern hybridizations, blots from Clontech were used.

RNA *in situ* hybridizations. Embryos and tissue samples from adult animals were fixed overnight in 4% paraformaldehyde in PBS at 4°C. Riboprobes were generated from the cloned 2-kb *Zfp185* cDNA by RNA *in vitro* transcription. Prior to transcription, the vector was linearized with *Xba*I or *Mlu*I, and antisense and sense probes were then transcribed from the T3 and T7 promoters using T3 and T7 polymerases, respectively. *In situ* hybridization was performed as described previously (Bächner *et al.*, 1993). Briefly, the probes were labeled to a specific activity of $>10^9$ dpm/ μ g with [α - 35 S]dUTP. Probe length was reduced to 150–200 nucleotides by alkaline hydrolysis. The slides were prehybridized at 54°C in a solution containing 50% formamide, 10% dextran sulfate, 0.3 M NaCl, 10 mM Tris, 10 mM sodium phosphate, pH 6.8, 20 mM dithiothreitol, 0.2 \times Denhardt's reagent, 0.1% Triton X-100, 1.25 mg/ml yeast RNA, and "cold" 0.1 mM [S]dUTP. For hybridization, 80,000 dpm/ μ l [α - 35 S]dUTP-labeled riboprobe was added to the hybridization mix, and hybridization was continued at 54°C for 16 h in a humid chamber. The slides were washed in hybridization salt solution to which dithiothreitol was

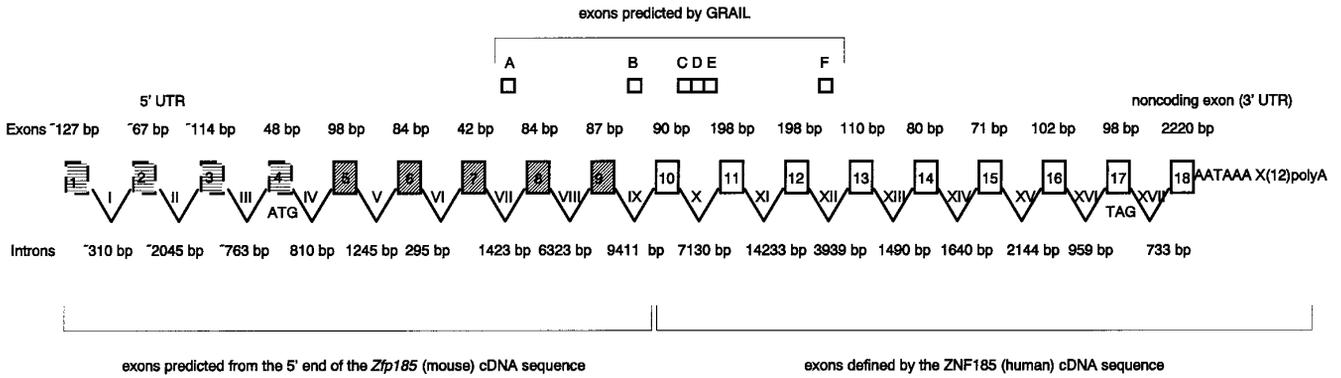


FIG. 1. Schematic diagram showing the genomic structure of the ZNF185 gene. Exons 10–18 were derived from assembled cDNA clones isolated by cDNA selection and sequenced ESTs (open boxes). Exons 13–16 were simultaneously isolated by exon trapping (open boxes). Exons 5–9 (diagonally hatched boxes in solid lines) and exons 1–4 (horizontally hatched boxes in broken lines) were predicted based on homology of the mouse *Zfp185* cDNA to the human genomic sequence. Exons 5–9 were verified at the peptide level, while prediction of exons 1–4 was solely based on nucleotide sequence homology. The positions of exons predicted by GRAIL are shown in small open boxes and are numbered A–F. The exons and introns are not drawn to scale.

added. After RNase A digestion (40 µg/ml) the slides were washed consecutively for 30 min at 37°C with 2× SSC and 0.1× SSC containing 0.1% SDS. The tissues were subsequently dehydrated by passage through increasing concentrations of ethanol. The slides

were coated with Ilford K5 photoemulsion for autoradiography. After 3 weeks of exposure at 4°C, the slides were developed and stained with Giemsa solution. The embryos and sections were analyzed using bright- and dark-field illumination with a Zeiss SV11 stereo micro-

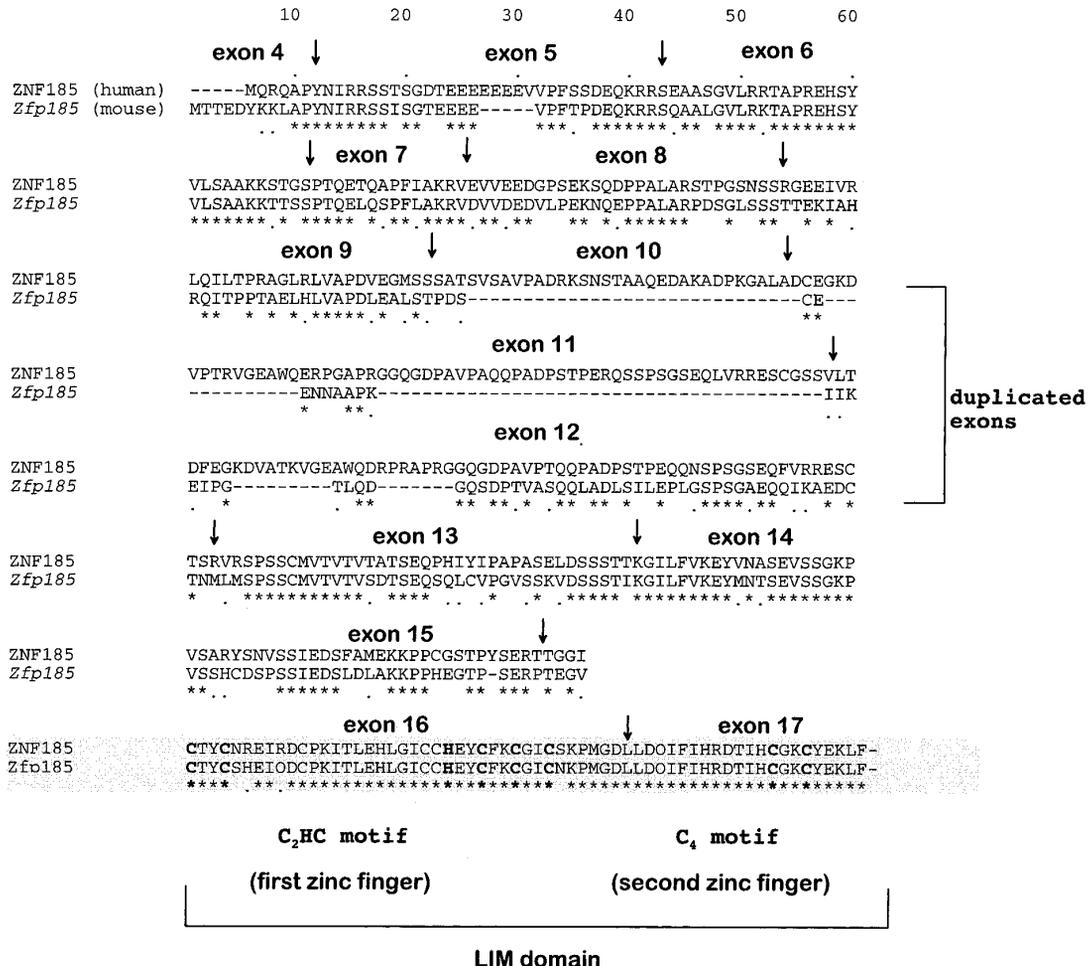


FIG. 2. CLUSTAL alignment of the ZNF185 and *Zfp185* peptide sequences. The positions of the exons demarcated by arrows refer to the structure of the human gene. The LIM domain is shaded and conserved cysteine (C) and the histidine (H) are shown in boldface type. (*) Positions of identical amino acids; (:) positions of similar but conserved amino acids; spaces, different amino acid positions that may alter the protein structure; gaps, regions of dyshomology.

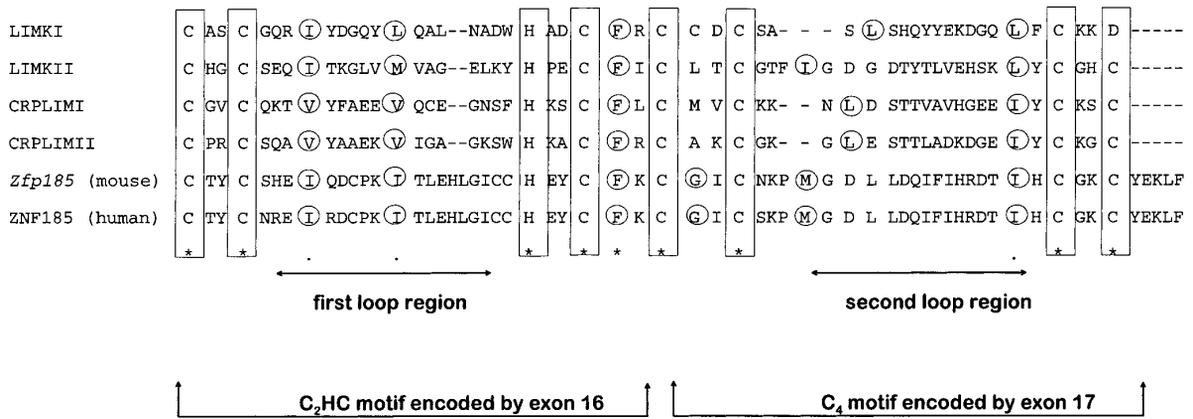


FIG. 3. CLUSTAL alignments of the ZNF185 and *Zfp185* LIM domains with those of other LIM domain proteins. LIMKI and LIMKII represent the first and second LIM domains of the protein kinase double LIM domain protein (Mizuno *et al.*, 1994), CRPLIMI and CRPLIMII represent the two LIM domains of the CRP protein (human cysteine-rich protein) (Liebhaber *et al.*, 1990). The conserved cysteine and histidine residues are surrounded by boxes. Conserved hydrophobic residues are circled.

scope and an Axioplan microscope and photographed using Kodak Ektachrome 320T tungsten film.

RESULTS

Structure of the ZNF185 Gene

The ZNF185 cDNA (EST XAP105; Accession No. X96622) had been assembled from four overlapping cDNA selection clones isolated by using complex exon trap products as probes on enriched cDNA libraries (Heiss *et al.*, 1996). By sequencing 1 of the 19 EST clones identified by BLASTN searches (clone 26810; Accession No. R14025), a final cDNA sequence with a length of 3140 bp was obtained. Although the 5' end of the cDNA was missing, a continuous open reading frame of 307 amino acids terminating in a TAG stop codon was present. The 3' UTR consisted of 2220 bp bearing a AATAAA poly(A) signal and a poly(A) tail.

To determine the genomic structure of the ZNF185 gene, cosmids to which ZNF185 mapped were sequenced. Alignment of the cDNA sequence with the genomic sequence revealed the genomic structure of the 3' end of the gene and consisted of exons 10–18 and introns X–XVII (Fig. 1). Exons 13–16 had previously been defined by four adjacent exon trap products (Heiss *et al.*, 1996), and their authenticity was verified by the availability of the genomic sequence (Fig. 1). Attempts to extend the 5' end of the ZNF185 cDNA by performing 5' RACE and screening conventional cDNA libraries were unsuccessful. Furthermore, GRAIL did not reliably predict the 5' exons. Subsequently, the 5' end of the human cDNA was constructed by comparing the human genomic sequence with the full-length murine *Zfp185* cDNA sequence (clone P1A; Accession No. U46687; Levin *et al.*, 1996). The homology of the mouse cDNA and the human genomic sequence was sufficiently high to permit an accurate prediction of the size and location of the 5' ZNF185 exons. In this way, exons 4–9 and introns IV–IX were

identified (Fig. 1). Except for the splice junctions flanking exon 4, all other predicted exon–intron boundaries represented conserved splice junctions and maintained the open reading frame. Although homology between the human and the mouse first coding exons was low at the nucleotide level, a higher degree of conservation at the peptide level facilitated determination of the position and sequence of human exon 4. Alignment of the mouse cDNA with the human genomic sequence indicated that the 5' UTR probably contains three additional noncoding exons (exons 1–3) with their respective introns (introns I–III; Fig. 1).

While exons 5, 6, 10, 11, 12, and 14 present in the human and mouse cDNA sequences were also recognized by GRAIL, the remaining exons were not predicted (Fig. 1). GRAIL, however, predicted six additional exons, numbered A–F (Fig. 1). The predicted exons were flanked by classical splice junctions, but because they were not present in the cDNA it was necessary to verify their authenticity. Failure to amplify GRAIL exons A–F by RT-PCR implied that they do not represent true exons. It cannot be excluded, however, that the predicted exons represent alternatively spliced exons. By adding up the length of the cDNA with the predicted 5' end and by taking into account an average length of 200 bp for the poly(A) tail, a length of ~4.2 kb was calculated for the full-length transcript. From the combined sequence data, the ZNF185 gene was found to span ~59 kb of genomic DNA and to exhibit a centromere to telomere direction of transcription.

A number of diseases are associated with regions of expanded triplet repeats that may occur in the coding or noncoding sequence of the candidate gene (Mandel, 1997). On searching the cDNA sequence for such repeats, although atypical, a 7× GAG repeat was detected in exon 5 of the ZNF185 cDNA. This repeat was absent in the murine *Zfp185* cDNA. In addition, a 5× and 3× GGA repeat interrupted by TGA in intron XI was identified in the genomic sequence of ZNF185. The

cDNA and the genomic sequences have been given Accession Nos. Y09538 and U82671, respectively.

Comparisons of the ZNF185 and Zfp185 cDNA and Peptide Sequences

FASTA alignments demonstrated that the homology of the human and mouse cDNAs was unevenly distributed, with an average identity of only 71.6 and 54% at the nucleotide and peptide levels, respectively. Regions of dyshomology and regions of very high homology were confined to certain exons (Fig. 2). While at the nucleotide level, the first coding exon, exon 4, was not very conserved (53.8% identity); exons 5–9 showed higher levels of homology (71.4–89% identity). The degree of homology dropped within exons 10–12 and was largely due to the presence of a duplicated exon, exons 11 and 12, found only in the human cDNA (Fig. 2). The duplicated exons exhibited 92% identity at the nucleotide level, while at the peptide level, 81.8% of the amino acids were identical, and 7.6% were different but conserved. Owing to the high homology of human exons 11 and 12, it was impossible to determine which one of the two exons is missing in the murine counterpart. The duplicated exons were also present in the independently derived genomic sequence and confirmed the authenticity of the result. Although the intron sequences surrounding exons 11 and 12 were not wholly conserved, short blocks of repetitive sequences bordering the exons were identified. The sequence TTCATAAGG-GTGCTTATTCTTCTT(or CCT)CAG immediately flanked the 3' intron splice site of both exons. The only difference was the lack of the G (underlined) and a CCT (in parentheses) instead of CTT in the sequence preceding exon 12. The sequence flanking the 5' intron splice site was less well conserved, but contained the sequence TGTTTTTGT 40 and 42 nucleotides into the intron sequence following exons 11 and 12, respectively. Downstream of exon 12, the homology to the murine cDNA rose dramatically, reaching 93.9% in the last coding exons 16 and 17. This coincided with the presence of a conserved LIM zinc finger domain in both cDNAs.

In the segment encoding the LIM zinc finger domain only four diverging residues were identified, of which two were conserved and two were unique (Figs. 2 and 3). The motif consisted of a stretch of 56 amino acids containing the conserved sequence CysX₂CysX₁₉HisX₂. CysX₂CysX₂CysX₁₉CysX₂Cys. The first "finger" demarcated by the C₂HC motif was encoded by exon 16, while the second "finger" defined by the C₄ motif was encoded separately by exon 17 (Figs. 2 and 3). In most LIM domains the amino acids X_{17–19} forming the loop regions, although less rigidly conserved, contain hydrophobic isoleucine (I), valine (V), leucine (L), phenylalanine (F), and/or methionine (M) residues at conserved positions (Dawid *et al.*, 1995). In agreement with this, both the ZNF185 and the Zfp185 peptides contained such conserved hydrophobic residues (Fig. 3). In addition, a conserved glycine (G) residue at the

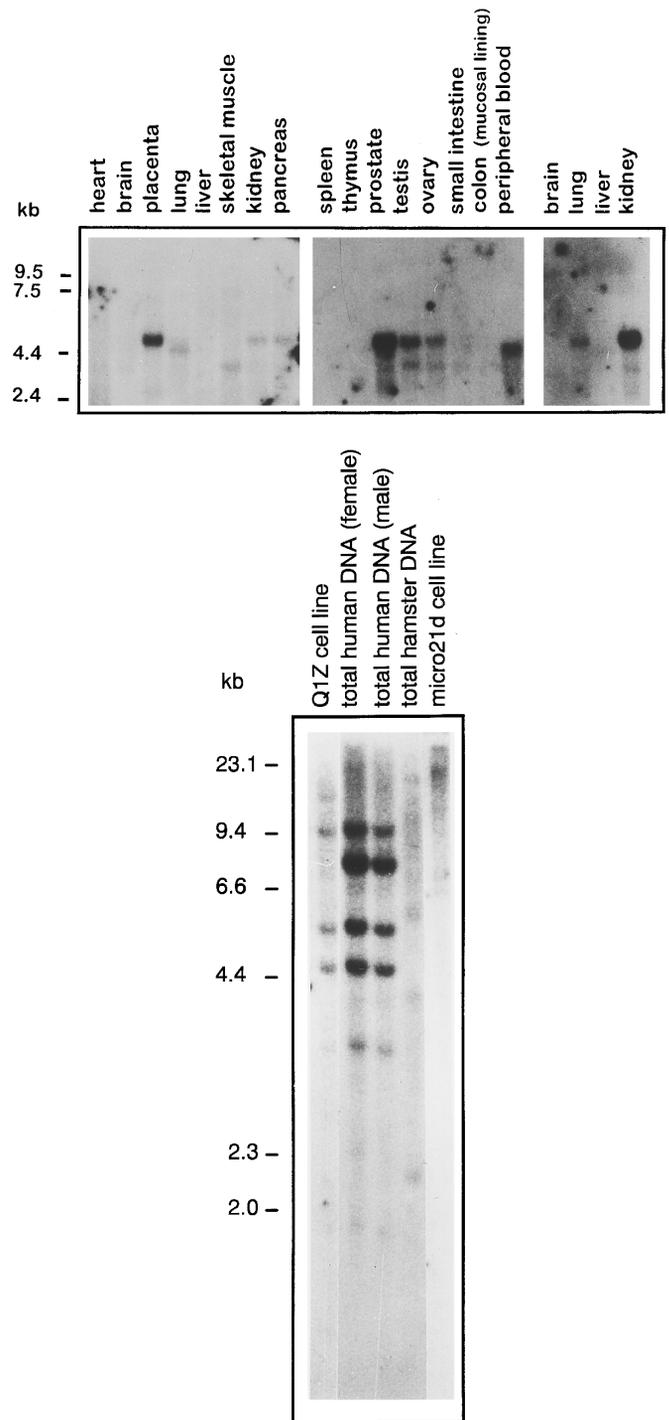
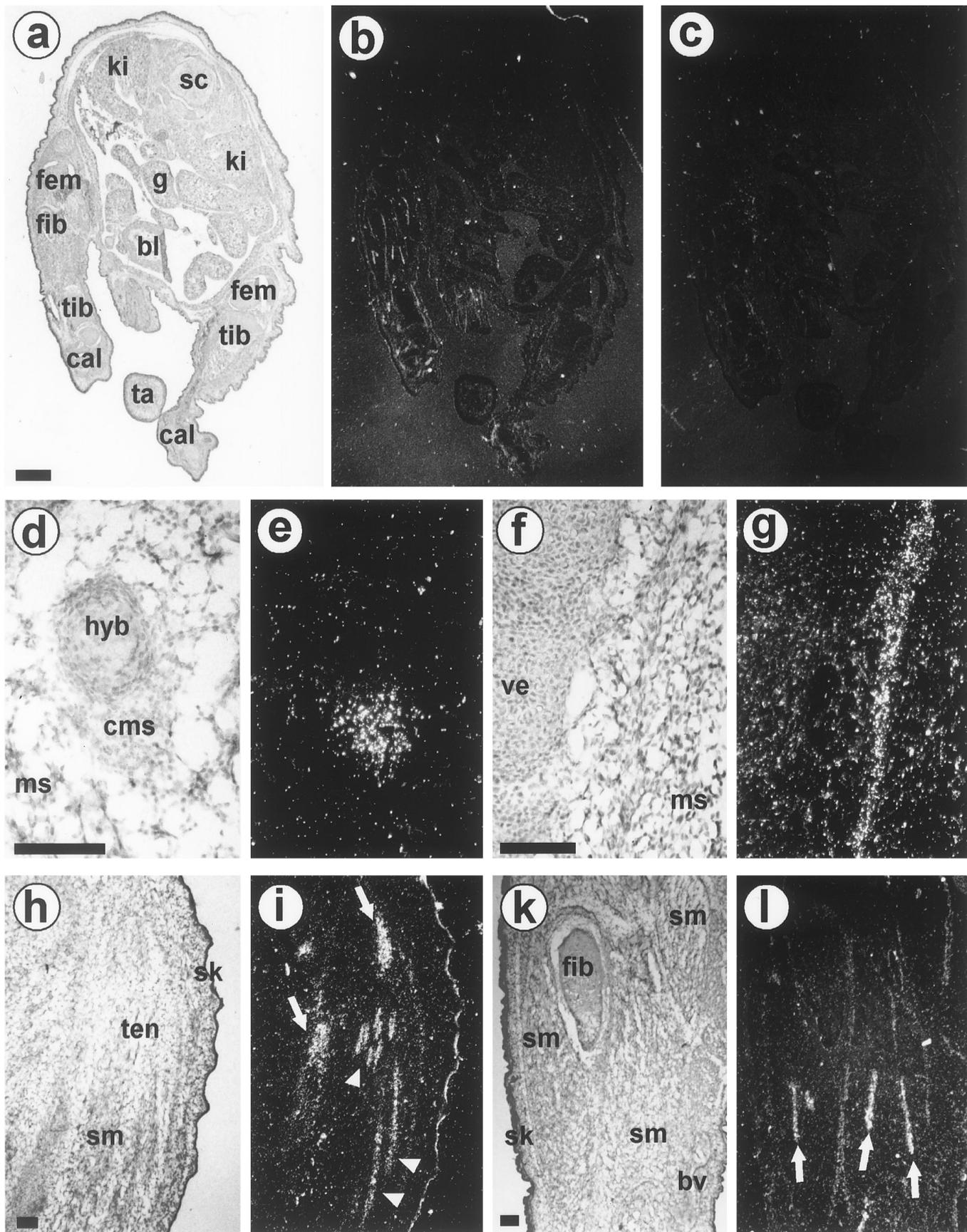


FIG. 4. Northern blots hybridized with cDNA selection clones containing exons 10–17 of the ZNF185 cDNA. (**Top**) From left to right, the first two blots contain adult tissues and the third blot contains fetal tissues. (**Bottom**) Southern blots hybridized with cDNA selection clones containing exons 10–17 of the ZNF185 cDNA. Total genomic DNA was isolated from blood; the Q1Z cell line carries Xq28 on a hamster background; the micro21d cell line carries the X chromosome on a hamster background minus Xq28.

fourth amino acid position following the histidine (H) residue found in many but not all LIM domains was present (Fig. 3). Hydrophobicity profiles showed that the peptide sequences 5' of the LIM domain do not



contain hydrophobic regions of sufficient length as is typical of signal peptides or membrane-spanning proteins. In accordance with the low hydrophobicity profiles and a probable nuclear localization, both the human and the mouse peptides have a low *pI* of 5.29 and 4.86, respectively. Although no further regulatory homeodomains or kinase domains were identified, a proline (P) content of 8.850 and 8.239% for the ZNF185 and *Zfp185* peptides, respectively, was evident. Considering that the proline content of most eukaryotic proteins typically lies at 4.5–5.5%, this is remarkably high.

Expression of the ZNF185 and Zfp185 Genes

Northern blot hybridizations. Northern blots previously showed that ZNF185 is expressed as a single ~5-kb transcript in placenta, pancreas, and kidney (Heiss *et al.*, 1996). On acquiring more cDNA sequence and extending the Northern blot analyses, however, additional ~3.8- and ~4.4-kb transcripts expressed variably in different tissues were identified (Fig. 4). The cDNA discussed in the present report, including the predicted 5' end, probably represents the ~4.4-kb transcript. To investigate whether the presence of more than one transcript was owing to the existence of a related gene, panels with total genomic DNA, genomic DNA from the Q1Z cell line carrying Xq28 on a hamster background, and micro21d bearing the remainder of the X chromosome on a hamster background were hybridized. One extra band was observed in the lanes containing total genomic DNA, suggesting that ZNF185 belongs to a gene family (Fig. 4).

RNA *in situ* hybridizations. RNA *in situ* hybridizations were performed on mouse embryo sections from 10.5 to 18.5 days post coitum (dpc) and on different adult tissue sections. No expression was evident in 10.5- and 12.5-dpc mice and was only visible in mid to late gestation embryos. Expression was first detected in 14.5-dpc embryos and was restricted to distinct condensing mesenchymal cells, for example, adjacent to the hyoid bone (Figs. 5d and 5e). In midgestation development, enhanced expression was also observed as an elongated stripe of connective tissue adjacent to the vertebrae (Figs. 5f and 5g), in condensing mesenchymal cells of the limbs (not shown), and in the proximal tail region (arrows in Fig. 5i). Possibly, these cells represent differentiating tendons. This assumption was supported by the presence of high levels of expression in more differentiated tendons (arrowheads in Fig. 5i).

While expression in tendons of the limbs was maintained in late embryogenesis (arrows in Fig. 5l), a lower level of expression was additionally observed in the connective tissue sheaths (epimysium) surrounding the skeletal muscles (Figs. 5k and 5l). Expression was also observed in mesenchymal cells adjacent to the developing distal skeletal elements of the limbs, for example, in the distal tibia and the calcaneum (Figs. 5a and 5b) and in the epithelia of the epididymis of the testis (not shown). Hybridization with a sense control probe showed only background grain distribution (Fig. 5c).

DISCUSSION

The candidate regions for many diseases linked to Xq28 are in the megabase range and span the DXS52 loci. The chromosomal region surrounding these markers appears to contain a hot spot of recombination owing to the highly polymorphic and repetitive nature of the DXS52 loci (Oberlé *et al.*, 1985; Bell *et al.*, 1989; Feil *et al.*, 1990). Furthermore, the presence of expressed repetitive sequence elements consisting of the MAGE gene cluster in humans (Rogner *et al.*, 1995; Heiss *et al.*, 1996) and the *Xlr3* gene cluster in the syntenic region on the mouse X chromosome (Levin *et al.*, 1996) suggests a general and "conserved" chromosomal instability. As part of a long-range positional cloning approach in Xq28, we previously established a transcript map in a 700-kb region surrounding the DXS52 markers (Heiss *et al.*, 1996). The ZNF185 cDNA (EST XAP105) contains a LIM zinc finger domain, suggesting a function in the regulation of cellular proliferation and/or differentiation and a possible association with disease. For this reason, we characterized this gene further.

The genomic structure of the 3' part of the ZNF185 gene became apparent by comparing the 3140-bp cDNA with the cosmid-derived genomic sequence. The cDNA did not represent a full-length transcript, and attempts to isolate its 5' end were unsuccessful. By using the genomic sequence and the 5' end of the orthologous mouse cDNA as a guide (Levin *et al.*, 1996), an accurate prediction of the remaining transcript was accomplished. Since GRAIL predicted six false and six true exons, but failed to predict the remaining eight coding and four noncoding exons, this indicated that although GRAIL is useful in predicting the genomic structure of some genes (Chen *et al.*, 1996), it is unsuccessful in predicting others (Lopez *et al.*, 1994). This justifies the

FIG. 5. RNA *in situ* analysis of *Zfp185* expression in mouse embryogenesis. Bright-field (a, d, f, h, k) and corresponding dark-field (b, c, e, g, i, l) images of transversal sections through 18.5-dpc (a–c, k, l) and sagittal sections through 14.5-dpc (d, e) and 16.5-dpc (f–i) embryos hybridized with a *Zfp185* antisense (b, e, g, i, l) and a sense (c) riboprobe are shown. At Day 14.5 dpc enhanced expression was visible at condensed mesenchymal cells adjacent to the hyoid bone (d, e). At Day 16.5 enhanced expression was further visible in cells lining the vertebrae (f, g) and tendons of the proximal tail (arrowheads and arrows in i). In late embryogenesis enhanced expression was detected in mesenchymal cells adjacent to distal limb bones, for example tibia and calcaneum (b), tendons (arrows in l), and the connective tissue sheaths surrounding the skeletal muscles (l). Only unspecific grain distribution is obtained with the sense control hybridization (c). bl, bladder; bv, blood vessel; cal, calcaneum, cms, condensed mesenchyme; fem, femur; fib, fibula; g, gut; hyb, hyoid bone; ki, kidney; ms, mesenchyme; sc, spinal cord; sk, skin; sm, skeletal muscle; ta, tail; ten, tendons; tib, tibia; ve, vertebra. Bar a–c, 1 mm; d–l, 100 μ m.

importance of using cDNA clones and not relying solely on the genomic sequence for gene identification. In view of the ultimate goal aimed at establishing gene maps, our data also outline the value of integrating the physical maps with the transcript and sequence maps.

Northern blots revealed three transcripts expressed selectively in different tissues. Because Southern blots detected a second locus mapping outside of Xq28, it currently remains uncertain which of the transcripts are derived from the ZNF185 gene and which correspond to the related gene. Furthermore, alternative splicing or alternative polyadenylation may give rise to the third transcript. In agreement with the Northern blots, *in situ* hybridizations revealed a nonubiquitous and stage-specific expression of *Zfp185*. In contrast to the Northern blots, however (Fig. 4a; Levin *et al.*, 1996), *in situ* hybridization detected no expression in organs at any stage of development, but was rather confined to differentiating connective tissue.

The high degree of conservation between genes on the human and the mouse X chromosomes implies a conservation in function (Davisson, 1987; Lundin, 1993). The homologous regions are preserved in the form of subchromosomal blocks and include Xq28. Although exceptions do exist (Faust *et al.*, 1992), gene order is largely conserved within the blocks (Davisson, 1987; Blair *et al.*, 1995; Rivella *et al.*, 1995). This was again demonstrated by a comparative analysis of the human DXS52 region and the syntenic region on the mouse X chromosome (Levin *et al.*, 1996). Although ZNF185 and *Zfp185* represented orthologous genes, they exhibited regions of dyshomology at the nucleotide and peptide sequence level. The largest segment of dyshomology extended from exons 10 to 12 and was largely owing to the presence of a duplicated exon contained only in the ZNF185 cDNA. This points to a sequence divergence after separation of the ancestral human and mouse lineages and is indicative of a recent evolutionary event. It remains open to speculation whether the human exon duplicated or the mouse exon became deleted or rearranged. One may presume with considerable certainty, however, that the function imposed by the LIM domain remained conserved.

According to the LIM protein classification system, ZNF185 and *Zfp185* belong to the third group of LIM proteins. Group 3 LIM proteins contain between one and five LIM domains in the carboxyl-terminus, usually in the absence of other regulatory domains (Taira *et al.*, 1995), and include proteins such as the cysteine-rich intestinal protein (CRIP; Birkenmeier and Gordon, 1986; Liebhaber *et al.*, 1990), the ESP1/CRP2 protein (Karim *et al.*, 1996), the cell-adhesion molecule zyxin (Sadler *et al.*, 1992), and the rhombotin gene family (RBTN1-3; Foroni *et al.*, 1992). In contrast to the RBTN3 (Foroni *et al.*, 1992), the mouse *Lhx3* (Zhadanov *et al.*, 1995), and the *LPP* (Petit *et al.*, 1996), proteins whose LIM domains are encoded by one exon, the two motifs composing the LIM domain of ZNF185 are encoded by two separate exons. Several LIM proteins

contain additional kinase domains (Mizuno *et al.*, 1994; Bernard *et al.*, 1996), homeodomains (Zhadanov *et al.*, 1995), and serine threonine kinase domains (Cheng and Robertson, 1995). Although ZNF185 and *Zfp185* did not contain such extra regulatory domains, both peptide sequences preceding the LIM domain were proline-rich. Proline-rich regions are typically associated with transcriptional activation domains and interact specifically with other factors to initiate transcription (Mermod *et al.*, 1989; Freyd *et al.*, 1990; Cheng and Robertson, 1995). Furthermore, proline-rich regions are frequently interdigitated with negatively charged amino acids (Struhl, 1987) and have been proposed to mediate contacts with positively charged regions of other transcription factors (Ptashne, 1988). The proline-rich domains following and preceding the amino- and carboxyl-terminal LIM domains of the rat *isl-1* (Karlsson *et al.*, 1990) and *LPP* LIM (Petit *et al.*, 1996) proteins, respectively, appear to exhibit such functional activity. Conceivably, the proline-richness and the acidic nature of the ZNF185 and *Zfp185* peptides may be of similar significance.

The composition of the ZNF185 and *Zfp185* peptide sequences permits prediction of a function in regulating cellular proliferation and/or differentiation, thus implicating a role in the etiology of disease. Levin *et al.* (1996) previously suggested that *Zfp185* may be a candidate for the *Bpa* mutant, which is the murine counterpart of the human dominant form of chondrodysplasia punctata (CDPX2) linked to Xq28. In the present report, however, expression analyses did not directly support this supposition. It is nonetheless of interest to note that candidate triplet repeat regions were found in exon 5 and intron XI of the ZNF185 gene. Although no anticipation effects have been observed in the mouse mutant (G. Herman, unpublished data) and the question of whether CDPX2 is a triplet repeat disorder remains open, increasing severity of the disease in later generations has been observed in some families. ZNF185 thus remains a good candidate for CDPX2. Knowledge of the genomic structure of ZNF185 will permit extensive mutation analyses. Because numerous diseases are linked to the DXS52 region, but a direct association with a specific disease is lacking, ZNF185 also remains a candidate for all of these diseases.

ACKNOWLEDGMENTS

We thank Petra Wilgenbus and Alexandra Ochs for carrying out the cDNA sequencing work. We thank Zdenek Sedlacek for critically reading the manuscript. This research was supported by grants from the BMBF (Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie) and the Human Genome Analysis Program of the European Community and in part by the National Institutes of Health R01 NS34953 to G.E.H.

REFERENCES

- Arngrimsson, R., Dokal I., Luzzatto L., and Connor, J.-M. (1993). Dyskeratosis congenita: Three additional families show linkage to a locus in Xq28. *J. Med. Genet.* **30**: 618–619.

- Auricchio, A., Brancolini, V., Casari, G., Milla, P. J., Smith, V. V., Devoto, M., and Ballabio, A. (1996). The locus for a novel syndromic form of neuronal intestinal pseudoobstruction maps to Xq28. *Am. J. Hum. Genet.* **58**: 743–748.
- Bächner, D., Manca, A., Steinbach, P., Wöhrle, D., Just, W., Vogel, W., Hameister, H., and Poustka, A. (1993). Enhanced expression of the murine *Fmr-1* gene during germ cell proliferation suggests a special function in both the male and the female gonad. *Hum. Mol. Genet.* **2**: 2043–2050.
- Bell, M. V., Patterson, M. N., Dorkins, H. R., and Davies, K. E. (1989). Physical mapping of DXS134 close to the DXS52 locus. *Hum. Genet.* **82**: 27–30.
- Bernard, O., Burkitt, V., Webb, G. C., Bottema, C. D. K., Nicholl, J., Sutherland, G. R., and Matthew, P. (1996). Structure and chromosomal localization of the genomic locus encoding the *Kiz1* LIM-kinase gene. *Genomics* **35**: 593–596.
- Biancali, V., Le Marec, B., Odent, S., van den Hurk, J. A. M. J., and Hanauer, A. (1991). Oto-palato-digital syndrome type I: Further evidence for assignment of the locus to Xq28. *Hum. Genet.* **88**: 228–230.
- Birkenmeier, E. H., and Gordon, J. I. (1986). Developmental regulation of a gene that encodes a cysteine-rich intestinal protein and maps near the murine immunoglobulin heavy chain locus. *Proc. Natl. Acad. Sci. USA* **83**: 2516–2520.
- Blair, H. J., Ho, M., Monaco, A. P., Fisher, S., Craig, I. W., and Boyd, Y. (1995). High-resolution comparative mapping of the proximal region of the mouse X chromosome. *Genomics* **28**: 305–310.
- Chen, E. Y., Zollo, M., Mazzarella, R., Ciccociola, A., Chen, C., Zuo, L., Heiner, C., Burrough, C., Repetto, M., Schlessinger, D., and D'Urso, M. (1996). Long-range sequence analysis in Xq28: Thirteen known and six candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and G6PD loci. *Hum. Mol. Genet.* **5**: 659–668.
- Cheng, A. K., and Robertson, E. J. (1995). The murine LIM-kinase gene (*link*) encodes a novel serine threonine kinase expressed predominantly in trophoblast giant cells and the developing nervous system. *Mech. Dev.* **52**: 187–197.
- Connor, J. M., Gatherer, D., Gray, F. C., Pirrit, L. A., and Affara, N. A. (1986). Assignment of the gene for dyskeratosis congenita to Xq28. *Hum. Genet.* **72**: 348–351.
- Davisson, M. T. (1987). X-linked genetic homologies between mouse and man. *Genomics* **1**: 213–227.
- Dawid, I. B., Toyama, R., and Taira, M. (1995). LIM domain proteins. *C.R. Acad. Sci. Ser. B* **318**: 295–306.
- Eksioglu, Y. Z., Scheffer, I. E., Caedenas, P., Knoll, J., DiMario, F., Ramsby, G., Berg, M., Kamuro, K., Berkovic, S. M., Duyk, G. M., Parisi, J., Huttenlocher, P. R., and Walsh, C. A. (1996). Periventricular heterotopia: An X-linked dominant epilepsy locus causing aberrant cerebral cortical development. *Neuron* **16**: 77–87.
- Faust, C. J., Levinson, B., Gitschier, J., and Herman, G. E. (1992). Extension of the physical map in the region of the mouse X chromosome homologous to human Xq28 and identification of an exception to conserved linkage. *Genomics* **13**: 1289–1295.
- Feil, R., Palmieri, G., d'Urso, M., Heilig, R., Oberlé, I., and Mandel, J.-L. (1990). Physical and genetic mapping of polymorphic loci in Xq28 (DXS15, DXS52, DXS134): Analysis of a cosmid clone and a yeast artificial chromosome. *Am. J. Hum. Genet.* **46**: 720–728.
- Feuerstein, R., Wang, X., Song, D., Cooke, N. E., and Liehaber, S. A. (1994). The LIM/double zinc-finger motif functions as a protein dimerization domain. *Proc. Natl. Acad. Sci. USA* **91**: 10655–10659.
- Feroni, L., Boehm, T., White, L., Forster, A., Sherrington, P., Liao, X. B., Brannan, C. I., Jenkins, N. A., Copeland, N. G., and Rabbitts, T. H. (1992). The rhombotin gene family encode related LIM-domain proteins whose differing expression suggests multiple roles in mouse development. *J. Mol. Biol.* **226**: 747–761.
- Freyd, G., Kim, S. K., and Horvitz, H. R. (1990). Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*. *Nature* **344**: 876–879.
- Gedeon, A., Kerr, B., Mulley, J., and Turner, G. (1991). Localisation of the MRX3 gene for non-specific X linked mental retardation. *J. Med. Genet.* **28**: 372–377.
- Gregg, R. G., Metzberg, A. B., Hogan, K., Sekhon, G., and Laxova, R. (1991). Waisman syndrome, a human X-linked recessive basal ganglia disorder with mental retardation: Localization to Xq27.3-qtter. *Genomics* **9**: 701–706.
- Heiss, N. S., Rogner, U. C., Kioschis, P., Korn, B., and Poustka, A. (1996). Transcriptional mapping in a 700 kb region around the DXS52 locus in Xq28: Isolation of six novel transcripts and a novel ATPase isoform (hPMCA5). *Genome Res.* **6**: 478–491.
- Karim, M. A., Ohta, K., Masayuki, E., Jinno, Y., Niikawa, N., Matsuda, I., and Indo, Y. (1996). Human *ESPI/CRP2*, a member of the LIM domain protein family: Characterization of the cDNA and assignment of the gene locus to chromosome 14q32.3. *Genomics* **31**: 167–176.
- Karlsson, O., Thor, S., Norberg, T., Ohlsson, H., and Edlund, T. (1990). Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature* **344**: 879–882.
- Kioschis, P., Rogner, U. C., Pick, E., Klauck, S. M., Heiss, N., Siebenhaar, R., Korn, B., Coy, J. F., Laporte, J., Liechti-Gallati, S., and Poustka, A. (1996). A 900 kb cosmid contig and eleven new transcripts within the candidate region for myotubular myopathy (MTM1). *Genomics* **33**: 365–373.
- Levin, M. L., Chatterjee, A., Pragliola, A., Worley, K. C., Wehnert, M., Zhuchenko, O., Smith, R. F., Lee, C. C., and Herman, G. (1996). A comparative transcription map of the murine bare patches (*Bpa*) and striated (*Str*) critical regions and human Xq28. *Genome Res.* **6**: 465–477.
- Liehaber, S. A., Emery, G. J., Urbanek, M., Wang, X., and Cooke, N. E. (1990). Characterization of a human cDNA encoding a widely expressed and highly conserved cysteine-rich protein with an unusual zinc-finger motif. *Nucleic Acids Res.* **13**: 3871–3879.
- Lopez, R., Larsen, F., and Prydz, H. (1994). Evaluation of the exon predictions of the GRAIL software. *Genomics* **24**: 133–136.
- Lundin, L. G. (1993). Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* **16**: 1–19.
- Mandel, J.-L. (1997). Breaking the rule of three. *Nature Genet.* **386**: 767–769.
- Mermoud, N., O'Neill, E. A., Kelly, T. J., and Tijan, R. (1989). The proline-rich transcriptional activator of CTF-NF-1 is distinct from the replication and DNA binding domain. *Cell* **58**: 741–753.
- Mizuno, K., Okano, I., Ohashi, K., Nunoue, K., Kuma, K., Miyata, T., and Nakamura, T. (1994). Identification of a human cDNA encoding a novel protein kinase with repeats of the LIM/double zinc finger motif. *Oncogene* **9**: 1605–1612.
- Nordström, A. M., Penttinen, M., and von Koskull, H. (1992). Linkage to Xq28 in a family with nonspecific X-linked mental retardation. *Hum. Genet.* **90**: 263–266.
- Oberlé, I., Drayna, D., Camerino, G., White, R., and Mandel, J.-L. (1985). The telomeric region of the human X chromosome long arm: Presence of a highly polymorphic DNA marker and analysis of recombination frequency. *Proc. Natl. Acad. Sci. USA* **82**: 2824–2828.
- Perez-Alvarado, G. C., Miles, C., Michelsen, J. W., Louis, H. A., Winge, D. R., Beckerle, M. C., and Summers, M. F. (1994). Structure of the carboxyl-terminal LIM domain from the cysteine rich protein CRP. *Struct. Biol.* **1**: 388–398.
- Petit, M. M. R., Mols, R., Schoenmakers, E. F. P. M., Mandahl, N., Van de Ven, W. J. M. (1996). *LPP*, the preferred fusion partner of *HMGIC* in lipomas, is a novel member of the LIM protein gene family. *Genomics* **36**: 118–129.
- Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature* **335**: 683–689.
- Radloff (1967). A dye-buoyant-density method for the detection and

- isolation of closed circular duplex DNA: The closed circular DNA in HeLa cells. *Proc. Natl. Acad. Sci. USA* **57**: 1514–1521.
- Rivella, S., Tamanini, F., Bione, S., Mancini, M., Herman, G., Chatterjee, A., Maestrini, E., and Toniolo, D. (1995). A comparative transcriptional map of a region of 250 kb on the human and mouse X chromosome between the G6PD and the FLN1 genes. *Genomics* **28**: 377–382.
- Rogner, U. C., Wilke, K., Steck, E., Korn, B., and Poustka, A. (1995). The melanoma antigen gene (MAGE) family is clustered in the chromosomal band Xq28. *Genomics* **29**: 725–731.
- Rogner, U. C., Heiss, N. S., Kioschis, P., Wiemann, S., Korn, B., and Poustka, A. (1996). Transcriptional analysis of the candidate region for incontinentia pigmenti (IP2) in Xq28. *Genome Res.* **6**: 922–934.
- Sadler, I., Crawford, A. W., Michelsen, J. W., and Beckerle, M. C. (1992). Zyxin and cCRP: Two interactive LIM domain proteins associated with the cytoskeleton. *J. Cell Biol.* **119**: 1573–1587.
- Sanchez-Garcia, I., and Rabbitts, T. H. (1994). The LIM domain: A new structural motif found in zinc-finger-like proteins. *Trends Genet.* **10**: 315–320.
- Schmeichel, K. L., and Beckerle, M. C. (1994). The LIM domain is a modular protein-binding interface. *Cell* **79**: 211–219.
- Sedlacek, Z., Korn, B., Konecki, D. S., Siebenhaar, R., Coy, J. F., Kioschis, P., and Poustka, A. (1993). Construction of a transcription map of a 300 kb region around the human G6PD locus by direct cDNA selection. *Hum. Mol. Genet.* **11**: 1865–1869.
- Struhl, K. (1987). Promoters, activator proteins and the mechanism of transcriptional initiation in yeast. *Cell* **49**: 295–297.
- Taira, M., Evrard, J.-L., Steinmetz, A., and Dawid, I. B. (1995). Classification of LIM proteins. *Trends Genet.* **11**: 431–432.
- Traupe, H., Müller, D., Atherton, D., Kalter, D. C., Cremers, F. P. M., van Oost, B. A., and Ropers, H. H. (1992). Exclusion mapping of the X-linked dominant chondrodysplasia punctata/ichthyosis/cataract/short stature (Happle) syndrome: Possible involvement of an unstable pre-mutation. *Hum. Genet.* **89**: 659–665.
- Wang, X., Lee, G., and Liebhaber, S. A. (1992). Human cysteine-rich protein: A member of the LIM/double-finger family displaying coordinate serum induction. *J. Biol. Chem.* **267**: 9176–9184.
- Way, J. C., and Chalfie, M. (1988). *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* **54**: 5–16.
- Zhadanov, A. B., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., and Westphal, H. (1995). Genomic structure and chromosomal localization of the mouse LIM/homeobox gene *Lhx3*. *Genomics* **27**: 27–32.
- Zuffardi, O., and Fraccaro, M. (1982). Gene mapping and serendipity. The locus for torticollis, keloids, cryptorchidism and renal dysplasia (31430, McKusick) is at Xq28, distal to the G6PD locus. *Hum. Genet.* **62**: 280–281.