Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 3702–3706, April 1995
Genetics

Expressed cadherin pseudogenes are localized to the critical region of the spinal muscular atrophy gene

(chromosome 5/neuronal expression)

S. SELIG*, S. BRUNO*, J. M. SCHARF†, C. H. WANG‡, E. VITALE‡, T. C. GILLIAM‡, AND L. M. KUNKEL*§

*Howard Hughes Medical Institute and Genetics Division, Children's Hospital, Boston, MA 02115; †Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology, Harvard Medical School, Boston, MA 02115; and ‡Departments of Psychiatry, Genetics, and Development, College of Physicians and Surgeons, Columbia University, and New York State Psychiatric Institute, New York, NY 10032

Contributed by L. M. Kunkel, January 25, 1995

ABSTRACT **Low-copy repeats have been associated with genomic rearrangements and have been implicated in the generation of mutations in several diseases. Here we characterize a subset of low-copy repeats in the spinal muscular atrophy (SMA) region in human chromosome 5q13. We show that this repeated sequence, named c41-cad, is a highly expressed pseudogene derived from an intact neuronal cadherin gene, Br-cadherin, situated on 5p13–14. Br-cadherin is expressed specifically in the brain, whereas the c41-cad transcripts are 10–15 times more abundant and are present in all tissues examined. We speculate that the c41-cad repeats, separately or in concert with other repeats in the SMA region, are involved in the pathogenesis of SMA by promoting rearrangements and deletions.**

Genomic regions containing repeated DNA segments are generally unstable and may trigger unequal crossing-over events that can result in gene conversions, duplications, or deletions. These events can cause mutations leading to genetic disease (1–3). Human chromosome 5 contains such an area at 5q13–14, where the abundance of low-copy repeats has been well documented (4–9). Interestingly, the gene for spinal muscular atrophy (SMA) has been mapped in tight linkage to this region (10, 11).

SMA is a childhood neurological disease characterized by degeneration of spinal cord anterior horn cells and their axons, leading to secondary muscle wasting. It is the most common fatal autosomal recessive neurological illness and occurs in three distinct clinical forms differentiated by their severity (12), all linked to the same locus on 5q13 (10, 11, 13). The repetitive nature and unusual instability of this region hindered the task of narrowing down the critical region and identifying candidate genes (7, 8, 14–17). However, deletions of several regional polymorphic markers in SMA patients (7, 9) led to the discovery of three candidate genes, all present in multiple copies and exhibiting homozygous deletions in SMA patients at very high rates (18–20).

Here, we describe the cloning and characterization of a low-copy repeat which maps in the immediate vicinity of the SMA gene(s) in 5q13. This repeat sequence, named c41-cad, has high homology to the cadherin family of calcium-dependent cell adhesion glycoproteins (reviewed in refs. 21 and 22) and is a pseudogene of the cadherin family. c41-cad is derived from a neuronal cadherin gene residing on 5p13–14, which we have cloned and named Br-cadherin. Only one exon is present in c41-cad, encompassing the first extracellular (EC) repeat of Br-cadherin, and the remaining sequence contains the flanking introns. The c41-cad loci are highly transcribed but, unlike classic transcribed pseudogenes (23), are only partially processed, so that intron sequences are included in the polyadenylylated transcripts. We believe that these transcripts represent a class of unusually transcribed pseudogenes from the SMA region which are similar in their organization and expression to transcripts derived from two of the SMA candidate genes (19, 20). In addition, the genomic organization of c41-cad and Br-cadherin sheds light on the origin of repetitive sequences in the SMA locus.

MATERIALS AND METHODS

cDNA Cloning. Several cDNA libraries were screened by standard procedures (24). An amplified human fetal brain methylated cDNA library, prepared as described (25), and a primary unamplified adult brain methylated cDNA library, prepared as described (26), were screened with the short selected cDNA 335bp-cad (Fig. 1 *Upper*). The probe was prepared by PCR amplifying the 335-bp fragment with primers from phage λgt10 arms [LL136 (5′-CTTCCAGGGTAAAAAGCAAAAAG-3′) and LL135 (5′-AGCCTGGTTAAGTCCAAGCTG-3′)] and digesting with *Eco*RI to release phage arm DNA. After separation in a gel, the *Eco*RI insert was α-^{32}P-labeled by random priming (Boehringer Mannheim) and hybridized. Positive plaques were purified, and phage DNA was prepared (24). The isolated cDNA clones from both libraries were subcloned into pBluescript II SK(+) (Stratagene), with the exception of the 2.6-kb 3′ *Eco*RI fragment of Br-cadherin, which was amplified by PCR using 3′-cad (5′-TGCAGGATATTAATGATAATGAGCC-3′) and LL135 (vector primer mentioned above) and cloned by TA cloning (Invitrogen).

DNA Sequencing and Analysis. Sequencing was performed with an Applied Biosystems sequencer using *Taq* DNA polymerase cycle sequencing, and acquired data were analyzed with SEQUENCHER software (Genecodes, Ann Arbor, MI) or Genetics Computer Group software (program manual for the GCG package, version 7, April 1991). Homology to known sequences was determined by comparison with GenBank on the National Center for Biotechnology Information BLAST Network Service (29).

Southern and Northern Blot Analysis. Southern and Northern blotting and hybridization were done as described (24). Washes were at high stringency. Northern hybridization was performed with a multitissue blot (Clontech). The Br-cadherin 5′ probe, containing mostly the 5′ untranslated region (UTR; bp 287–1238), was prepared by PCR of the 5′ *Eco*RI fragment of Br-cadherin using SZ31 (5′-ATCTGAAGGGTAAATGTGGAGG-3′, forward primer) and SZ30 (5′-TTGGAAATGTGACCGTTGTCC-3′, reverse primer).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: FISH, fluorescence *in situ* hybridization; ORF, open reading frame; RT, reverse transcription; SMA, spinal muscular atrophy; UTR, untranslated region; YAC, yeast artificial chromosome.
§To whom reprint requests should be addressed.

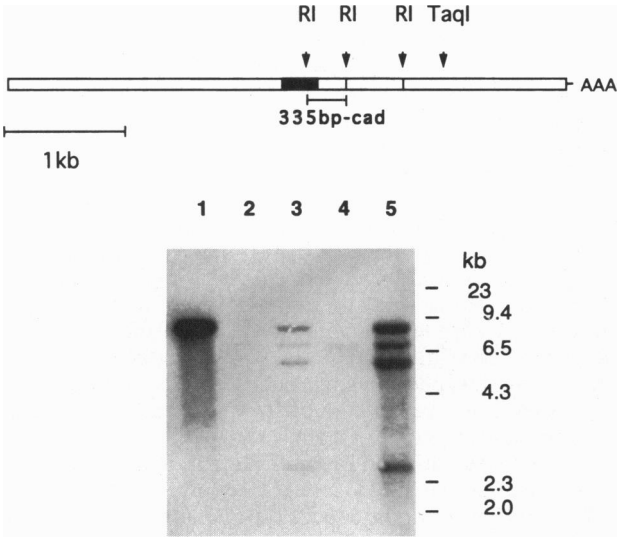

FIG. 1. c41-cad transcript. (*Upper*) An *Eco*RI (RI) and *Taq* I restriction map of c41-cad. The black box represents the open reading frame (ORF). 335bp-cad is the originally selected 335-bp *Eco*RI fragment used to detect the 4.6-kb transcript. (*Lower*) Southern blot analysis maps c41-cad to three chromosome 5 loci. A mapping panel was constituted from *Taq* I-digested DNA from yeast artificial chromosome YAC 98 (27) (lane 1); YAC 88 (27) (lane 2); HHW105, a human–Chinese hamster ovary (CHO) hybrid cell line containing one normal human chromosome 5 (28) (lane 3); HHW1064, a human–CHO hybrid cell line with a sole human chromosome 5 missing the region, 5q11.2–13.3, that contains the SMA critical region (28) (lane 4); and total human placental (46XX).

Genomic Phage Cloning. The genomic phage libraries screened were a singly amplified human EMBL 3 phage library (30) and two EMBL 3 phage libraries prepared from YACs 88 and 98 (8). Intron/exon borders were determined by subcloning the restriction fragments carrying exon sequence from genomic phages and sequencing across the borders.

Fluorescence *in Situ* Hybridization (FISH) Analysis. Labeling, hybridization, and detection were done by a standard protocol (31). Biotinylated probes were hybridized to metaphase spreads prepared from phytohemagglutinin-stimulated whole blood.

Reverse Transcription (RT)–PCR. Ten micrograms of total RNA was reverse transcribed with 40 units of Superscript in 1× first-strand buffer (GIBCO/BRL) in a volume of 20 μl. Then 40 μl of water was added and 1 μl was incubated with 50 ng of each primer at 94°C for 4 min, and 25 cycles of 60°C for 1 min, 72°C for 45 sec, and 94°C for 30 sec. RT–PCR for both c41-cad and Br-cadherin used a common forward primer, 3′-cad (described above), end-labeled with [γ-^{32}P]ATP. The reverse primers were SZ55 (5′-AACCGGTGGCTCATCTACG-3′) for Br-cadherin and SZ101 (5′-ACATTATTCATTTGCTGCTTGGC-3′) for c41-cad. Labeled 3′-cad primer (10^6 cpm) was added to each PCR mixtures. Ten microliters of each reaction mixture was electrophoresed in a 2% agarose gel and transferred to Hybond N^+ membrane (Amersham). The membrane was exposed to a storage phosphor plate and scanned with a PhosphorImager using IMAGEQUANT software (Molecular Dynamics).

RESULTS

Isolation of a Pseudo-Cadherin Transcript, c41-cad, from the SMA Locus. To isolate candidate genes for SMA we performed direct cDNA selection (unpublished work) on two YACs, 88 and 98, which span the critical region on 5q13–14 (27). Of the nine distinct short cDNAs obtained, one consisted of a 335-bp *Eco*RI fragment which contained a partial ORF exhibiting high amino acid sequence homology to a segment of a cadherin extracellular repeat. Because different cadherin molecules have been associated with neuronal development (22), a gene from this family was considered a favorable candidate for SMA. To obtain a full-length transcript, the original 335-bp cadherin (335bp-cad) sequence was used to screen a human fetal brain cDNA library. Three hybridizing clones were detected, the longest being a 4.6-kb polyadenylylated transcript, c41-cad (Fig. 1 *Upper*). Full sequencing of c41-cad revealed an ORF of 360 bp in the center of the transcript, encompassing an entire extracellular cadherin repeat plus 51 bp of an adjacent repeat. Surprisingly, the remaining cDNA contained no additional significant ORFs. This organization suggests a potentially unprocessed transcript or a pseudogene transcript.

The complex genomic organization of c41-cad was apparent by Southern hybridization of a mapping panel of *Taq* I-digested DNA (Fig. 1 *Lower*) with the full-length 4.6-kb c41-cad as a probe. Four *Taq* I fragments of 8, 6.5, 5.5, and 2.5 kb were detected in both the rodent–human cell line HHW105 (lane 3), which includes one normal human chromosome 5, and total human DNA (lane 5). When hybridizing with the original 335bp-cad-selected cDNA, the three upper *Taq* I fragments were detected as well (result not shown). Since 335bp-cad contains no *Taq* I site and is linear with genomic sequence (see below), this probe must be detecting three separate chromosome 5 loci. The additional 2.5-kb *Taq* I fragment was visible only when the complete 4.6-kb c41-cad was the probe, as a consequence of an internal *Taq* I site positioned in the most 3′ *Eco*RI fragment of c41-cad (Fig. 1 *Upper*). The intensities of the three upper *Taq* I fragments are not identical. This finding indicates that some individual bands on the Southern blot originated from multiple loci, suggesting the existence of more than three copies of c41-cad or related sequences.

One of the loci on chromosome 5, represented by the 6.5-kb fragment, is clearly situated outside of the SMA region, since it is present in the hybrid HHW1064 DNA, which lacks 5q11.2–13.3 (lane 4). However, the genomic localization of c41-cad to the 5q13-selecting SMA YACs was confirmed by the 8-kb signal detected in YAC 98 DNA (lane 1). Absence of a signal in YAC 88 DNA (lane 2) indicates that c41-cad is either unique to YAC 98 or present in the 200-kb deletion in the overlapping YAC 88 region (27). The region represented by the 5.5-kb fragment is clearly included in the 5q11.2–13.3 area; however, it may be deleted from the YACs.

Cloning and Sequence Analysis of Br-cadherin cDNA. To determine whether a real cadherin gene was present in the SMA locus in addition to the c41-cad copies, additional cDNAs were obtained from a primary adult brain cDNA library by using 335bp-cad as a probe. Characterization of the 18 positive plaques revealed that 16 were similar to c41-cad and one was a rearranged c41-cad transcript. One of the 18 positively hybridizing phages was found to be a 4184-bp transcript containing an ORF of 794 aa which is highly homologous to other cadherin genes. During characterization of this clone, a nearly identical sequence including the entire ORF was published as a newly isolated brain-expressed cadherin gene named cadherin 12 (32). Based on the presence of this transcript in the brain, and the common naming system of cadherins according to their tissue of expression, we named this gene Br-cadherin (Genebank accession no. L33477; detailed analysis of this gene will be published elsewhere). Comparison of the Br-cadherin transcript with c41-cad reveals that it contains an almost identical sequence across the region of the original c41-cad ORF, diverging only at the 5′ end of c41-cad's ORF (Fig. 2).

Comparison of Br-cadherin and c41-cad Expression. Hybridization of c41-cad to a human adult multi-tissue Northern blot (Clontech) detected no discrete band, but rather a smeared signal in many tissues, strongest in lung and kidney,

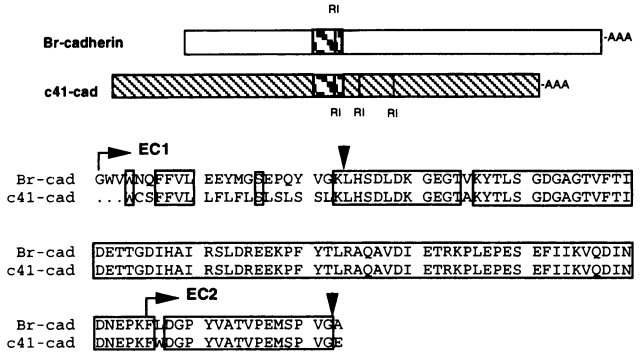

FIG. 2. Comparison of the predicted amino acid sequences for the ORF of c41-cad and the corresponding region in Br-cadherin (Br-cad). Identical amino acids are boxed, and start points of cadherin extracellular region repeats 1 and 2 are marked by EC1 and EC2. Arrows point to positions of conserved splice signals. The positions of these ORFs on the cDNAs are designated on the upper transcript diagrams by the shaded rectangles. The hatched region in c41-cad represents an intron-like sequence; the open boxes in Br-cadherin represent cDNA sequence. *Eco*RI (RI) restriction sites are indicated.

ranging from ≈2.5 to 7.5 kb in lung and from ≈7.5 kb to very high molecular weights in kidney (Fig. 3*a*). In contrast, a 5′ UTR-specific Br-cadherin probe detected a 4.4-kb transcript in brain only, very close in size to the cDNA clone we have obtained (Fig. 3*b*).

During our analysis of the two types of transcripts, it was apparent that the c41-cad pseudogene transcript was unusually highly represented in the cDNA libraries. Since the multiple c41-cad copies reside in the critical SMA region, we found it important to further characterize their pattern of expression and determine the level of expression relative to that of Br-cadherin. We were able to differentiate specifically between c41-cad and Br-cadherin transcripts by RT–PCR, using a common oligonucleotide from a homologous sequence present in both (3′-cad) combined separately with a unique sequence-specific oligonucleotide for each transcript (Fig. 3*c*). Dramatic differences between the two transcripts were revealed in all tissues examined; whereas c41-cad was expressed ubiquitously, Br-cadherin's expression was restricted to a limited number of tissues as demonstrated previously by Northern hybridization. However, due to high sensitivity of the PCR, some tissues showed RT–PCR products for Br-cadherin although no signal was detected by Northern analysis. Since the PCR products for both transcripts are of approximately the same size and were labeled radioactively with the same common primer, the intensity of the PCR products could be compared by phosphorimaging. Such analysis indicated that c41-cad's level of expression in brain was at least 10–15 times that of Br-cadherin (data not shown).

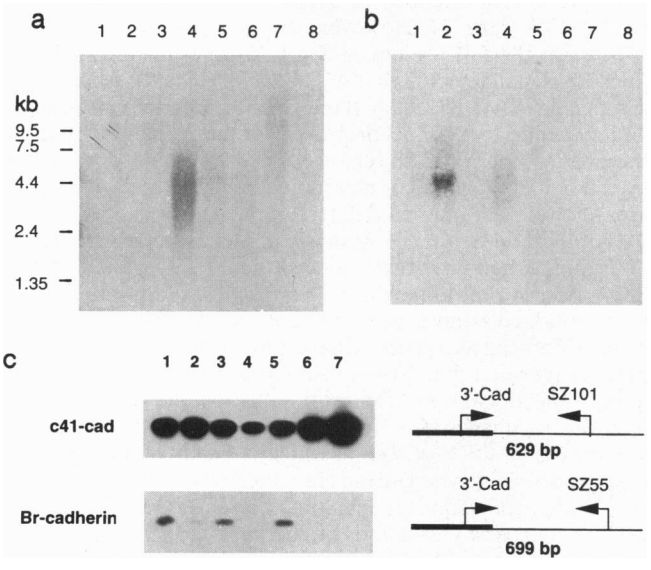

FIG. 3. c41-cad and Br-cadherin expression. (*a* and *b*) Expression in human adult tissues. Lanes: 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas. c41-cad expression was detected with the 335bp-cad probe (*a*), and Br-cadherin expression with the 5′ UTR probe (*b*). (*c*) Comparison of expression between Br-cadherin and c41-cad by radioactive PCR. RT–PCR was performed on cDNA from various tissues, utilizing a common forward primer for Br-cadherin and c41-cad (3′-cad) and a specific reverse primer for each of the transcripts (SZ101 for c41-cad and SZ55 for Br-cadherin). The 3′-cad primer is positioned on the common ORF (thick line) of Br-cadherin and c41-cad, as shown at right. Lanes: 1, cerebral cortex; 2, spinal cord; 3, skeletal muscle; 4, smooth muscle; 5, vascular endothelium; 6, heart; 7, liver.

Genomic Localization of Br-cadherin and c41-cad. Since by Southern analysis the pseudogene transcript, c41-cad, detected an additional chromosome 5 locus external to the SMA region (Fig. 1), we examined the genomic origin of the various transcripts, including Br-cadherin. For this purpose genomic phage libraries of YACs 88 and 98, spanning the SMA region (8) and a total human genomic phage library were screened with the full-length Br-cadherin and c41-cad cDNAs. Positively hybridizing phage then were partially sequenced to determine whether they carried Br-cadherin or c41-cad sequence. This sequence analysis revealed that only two short sequences which share near, but not complete, identity with Br-cadherin were present in the YAC 98 genomic library: (*i*) the c41-cad sequence (in phages 98-1 and 98-91) and (*ii*) a sequence which shares near identity with 5′ UTR bp 757–908 (in phages 98-28 and 98-87). In contrast, we detected in the total genomic library numerous positively hybridizing phage with Br-cadherin sequences, which form a contig across most of the genomic region of Br-cadherin (data not shown). Partial sequencing of these phage enabled the characterization of intron/exon borders of the Br-cadherin gene (partial data in Fig. 4). Comparison of the YAC 98 sequence of c41-cad with the Br-cadherin genomic phage 3′-7, from the homologous region, revealed that most of the ORF of c41-cad is exon 6 from Br-cadherin with a few differences, and the surrounding sequences of c41-cad are intron sequences adjacent to exon 6 (Figs. 4 and 2). The 5q13 sequence homologous to the 5′-UTR is similar to Br-cadherin exon 4 and surrounding intron sequences.

The chromosomal location of the various phages was determined by FISH (Fig. 5). Phage 98-91, containing the c41-cad sequence, hybridized to both 5p13–14 and 5q13 (Fig. 5*a*). Similar hybridizing loci were detected with phage 98-87, containing the 5′ UTR (result not shown). In contrast, phage CA-15A1, cloned from the total genomic library and containing the 3′ end of Br-cadherin, hybridized only to the 5p arm at 5p13–14 (Fig. 5*b*). From the latter result we concluded that Br-cadherin is localized to 5p13–14 and is likely to be the source of the 6.5-kb *Taq* I fragment detected in HHW1064 DNA (Fig. 1). Genetic linkage of the *D5S411* CA marker (34) which appears in the 3′ UTR of Br-cadherin, further confirmed the localization of Br-cadherin to the p arm.

The c41-cad pseudogene copies and the 5′-UTR pseudogene were indeed localized to the 5q13 region by FISH, and the intense hybridization of phage 98-91 to the 5q locus illustrates the repetitive nature of c41-cad in the SMA region in comparison to the single copy on 5p (Fig. 5*a*).

DISCUSSION

In our analysis of the SMA critical region, we have come upon a cadherin gene, Br-cadherin, that has an unusual genomic organization with related exons on both the p and q arms of

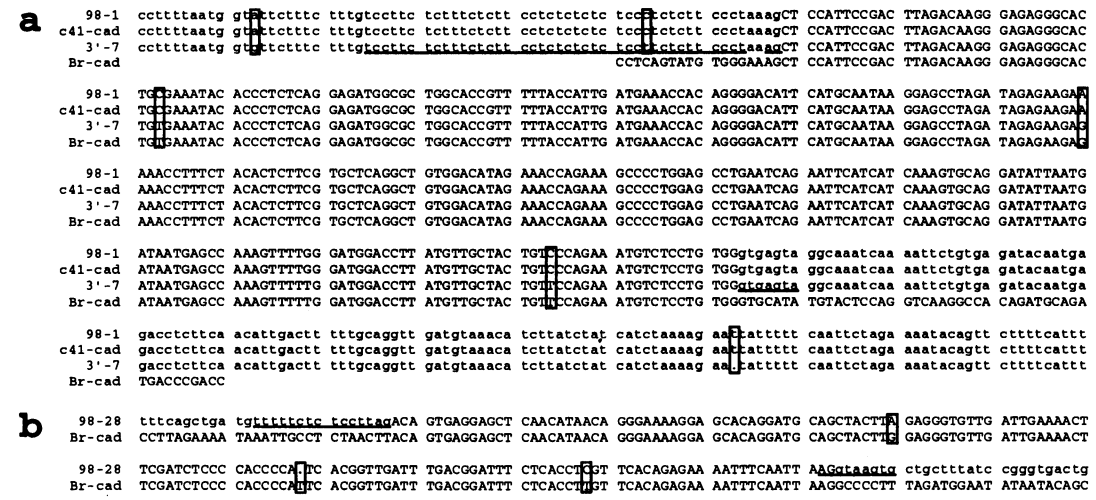

FIG. 4. Nucleic acid comparison between Br-cadherin and homologous regions in 5q13. The sequences shown are derived from several sources: Br-cadherin cDNA, c41-cad cDNA, and phages from the YAC 98 genomic library and total human genomic library. Uppercase letters present cDNA sequence of Br-cadherin, and lowercase letters present sequences that do not appear in Br-cadherin's cDNA. Comparisons were done with the PILEUP program (Genetics Computer Group), and differences are indicated by boxes. Underlined sequences are consensus sequences for donor and acceptor splice sites (33). (*a*) Comparison between the homologous regions of Br-cadherin and c41-cad. Base pairs 1302–1661 from Br-cadherin cDNA (Br-cad) are compared with the corresponding genomic phage 3′-7, with the c41-cad transcript (c41-cad), and with c41-cad's corresponding genomic phage from YAC 98 (98-1). Homology between both cDNAs occurs across bp 1302–1661 of Br-cadherin cDNA. The borders of the homology are putative intron/exon borders in the Br-cadherin locus. Intron sequence of Br-cadherin matches with almost 100% identity to sequences surrounding the ORF of c41-cad. (*b*) Comparison of a portion of the 5′ UTR of Br-cadherin with the homologous region in 5q13. Base pairs 730–929 from the Br cadherin 5′ UTR region (Br-cad) and the homologous region in YAC 98 (phage 98-28) demonstrate almost 100% identity across the region covering bp 757–902 of Br-cadherin. As shown in *a*, the borders of the homology are putative intron/exon borders in the Br-cadherin locus.

chromosome 5. The locus of Br-cadherin, 5p13–14, represents a genomic site for intact human cadherin genes in addition to chromosomes 16 and 18 (35–37). The 5q13 region contains cadherin pseudogenes containing partial duplications of a 5′-UTR exon and an additional coding exon (c41-cad) of Br-cadherin.

Southern hybridization and FISH analysis indicate that at least two copies of the c41-cad pseudogenes exist in the 5q13 region. However, based on the relative intensities of the *Taq* I fragments appearing on the Southern blots, we believe that at least five copies of c41-cad exist in the critical SMA region and that the number of copies probably varies among individuals (results not shown). Many other low-copy repeated sequences have been found on both arms of chromosome 5 (4–8, 14, 15, 20), but none of the other 5q13 repeats has a functional homologous gene associated with it on the p arm. The genomic organization of c41-cad and Br-cadherin suggests that at least several of the 5q13 pseudogenes arose by transposition of genomic p-arm material to the q arm of the same chromosome. After the initial event of transposition, additional copies were probably generated by amplification in the 5q13 locus, since none of the characterized sequences in the critical SMA region appears to be unique.

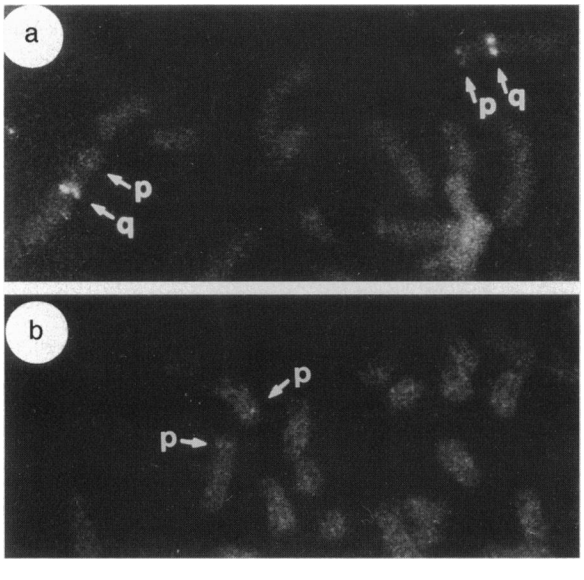

FIG. 5. (*a*) FISH with phage 98-91 from the c41-cad locus, which is homologous to a portion of the Br-cadherin locus. (*b*) FISH with phage CA-15A1, containing the 3′ end of Br-cadherin. Arrows point to the hybridization signals; p and q depict the chromosome arms. Hybridization to 5p13–14 is shown in *b*; hybridization to both 5p13–14 and 5q13 in *a* demonstrates large regions of genomic homology on both arms and multiple copies on the q arm.

The low-copy repeats at 5q13 may be the source of this region's high instability, since repeats can trigger unequal crossing-over leading to gene conversions, duplications, or deletions in the area. This is the case with several human diseases such as Gaucher, 21-hydroxylase deficiency, Charcot–Marie–Tooth type 1A, and defects with visual pigment genes (1–3, 38). Such an assumption coincides well with finding that deletions in the SMA region occur at extremely high rates (7, 14, 17, 19, 20), reaching almost 100% for *SMN*, one of the SMA candidate genes (18).

Although transcribed pseudogenes have been previously described, the characteristics of c41-cad suggest a new class of partially processed pseudogenes which incorporate introns as part of the polyadenylylated transcript. Although the expression level of these pseudogenes is extremely high, 10–15 times the level of the Br-cadherin transcript, no distinct hybridization band appears on Northern blots. Rather, a smeared signal is detected, indicating a size heterogeneity of the pseudogene transcripts. This phenomena is not exclusive for c41-cad but can be detected for two of the SMA candidate genes as well (19, 20). Consistent with the Northern blot result, we have detected variably sized c41-cad transcripts in both fetal and adult cDNA libraries: Analysis of 19 cDNA clones, 3 from the fetal brain library and 16 from the adult brain library, revealed

various distinct start and end points (results not shown). Varying-length transcripts and different exon combinations, sometimes from different genes, have also been reported by other groups (16, 19, 20, 39). These findings emphasize the unusual patterns of transcription occurring in the SMA region which may be explained by the genomic organization of the genes and pseudogenes in the region; several of the reported genes are clearly intertwined in each other and are being transcribed in opposite orientations [the XS2G3 gene (20) contains exon 7 of the NAIP gene in the opposite orientation (19)]. Due to mechanisms as yet unexplained, irregular transcription of the pseudogenes in the SMA region may be initiating and ending at various points, thus creating complex transcripts composed of exons from unrelated transcription units. In addition, these bizarre transcripts may be spliced in variable combinations—hence the smeared signal.

Our results indicate that the c41-cad pseudogenes are located in the SMA region and may be involved in the instability of the 5q13 region. Further studies of the nature of these different pseudogenes may reveal whether their high transcriptional level is associated in some way with the pathogenesis of SMA.

We thank Elizabeth McNally, Hart Lidov, Emanuela Gussoni, Carsten Bonnemann, and Tejvir Khurana for advice and comments on the manuscript. We appreciate Elizabeth McNally's assistance with RT–PCR experiments and Richard Bennet's assistance with sequencing. T.C.G. is supported by the Families of SMA (Chicago), Andrew's Buddies Inc., the Muscular Dystrophy Association of America, and National Institutes of Health Grant NS28877. L.M.K. is supported by National Institutes of Health Grant NS23740 and is an Investigator of the Howard Hughes Medical Institute.

1. He, G. S. & Grabowski, G. A. (1992) *Am. J. Hum. Genet.* **51,** 810–820.
2. Collier, S., Tassabehji, M. & Strachan, T. (1993) *Nat. Genet.* **3,** 260–265.
3. Chance, P. F. & Fischbeck, K. H. (1994) *Hum. Mol. Genet.* **3,** 1503–1507.
4. Francis, M. J., Morrison, K. E., Campbell, L., Grewal, P. K., Christodoulou, Z., Daniels, R. J., Monaco, A. P., Frischauf, A.-M., McPherson, J., Wasmuth, J. & Davies, K. E. (1993) *Hum. Mol. Genet.* **2,** 1161–1167.
5. Thompson, T. G., Morrison, K. E., Kleyn, P., Bengtsson, U., Gilliam, T. C., Davies, K. E., Wasmuth, J. J. & McPherson, J. D. (1993) *Hum. Mol. Genet.* **2,** 1169–1176.
6. Burghes, A. H. M., Ingraham, S. E., McLean, M., Thompson, T. G., McPherson, J. D., *et al.* (1994) *Genomics* **21,** 394–402.
7. Melki, J., Lefebvre, S., Burglen, L., Burlet, P., Clermont, O., Millasseau, P., Reboullet, S., Benichou, B., Zeviani, M., Paslier, D. L., Cohen, D., Weissenbach, J. & Munnich, A. (1994) *Science* **264,** 1474–1477.
8. Wang, C. H., Kleyn, P. W., Vitale, E., Ross, B. M., Lien, L., Xu, J., Carter, T., Brzustowicz, L. M., Selig, S., Pavone, L., Parano, E., Penchaszadeh, G. K., Munsat, T., Kunkel, L. M. & Gilliam, T. C. (1995) *Am. J. Hum. Genet.* **56,** 202–209.
9. DiDonato, C. J., Morgan, K., Carpten, J. D., Fuerst, P., Ingraham, S. E., Prescott, G., McPherson, J. D., Wirth, B., Zerres, K., Hurko, O., Wasmuth, J. J., Mendell, J. R., Burghes, A. H. M. & Simard, L. R. (1994) *Am. J. Hum. Genet.* **55,** 1218–1229.
10. Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmsen, K. C., Daniels, R., Davies, K. E., Leppert, M., Ziter, F., Wood, D., Dubowitz, V., Zerres, K., Hausmanowa-Petrusewicz, I., Ott, J., Munsat, T. L. & Gilliam, T. C. (1990) *Nature (London)* **344,** 540–541.
11. Melki, J., Abdelhak, S., Sheth, P., Bachelot, M. F., Burlet, P., *et al.* (1990) *Nature (London)* **344,** 767–768.
12. Pearn, J. (1980) *Lancet* **i,** 919–921.
13. Gilliam, T. C., Brzustowicz, L. M., Castilla, L. H., Lehner, T., Penchaszadeh, G. K., Daniels, R. J., Byth, B. C., Knowles, J., Hislop, J. E., Shapira, Y., Dubowitz, V., Munsat, T., Ott, J. & Davis, K. E. (1990) *Nature (London)* **345,** 823–825.
14. Carpten, J. D., DiDonato, C. J., Ingraham, S. E., Wagner-McPherson, C., Nieuwenhuijsen, B. W., Wasmuth, J. J. & Burghes, A. H. M. (1994) *Genomics* **24,** 351–356.
15. Wirth, B., Voosen, B., Rohrig, D., Knapp, M., Piechaczek, B., Rudnik-Schoneborn, S. & Zerres, K. (1994) *Genomics* **20,** 84–93.
16. Theodosiou, A. M., Morrison, K. E., Nesbit, A. M., Daniels, R. J., Campbell, L., Francis, M. J., Christodoulou, Z. & Davis, K. E. (1994) *Am. J. Hum. Genet.* **55,** 1209–1217.
17. McLean, M. D., Roy, N., MacKenzie, A. E., Salih, M., Burghes, A. H. M., Simard, L., Korneluk, R. G., Ikeda, J. & Surh, L. (1994) *Hum. Mol. Genet.* **3,** 1951–1956.
18. Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Paslier, D. L., Frezal, J., Cohen, D., Weissenbach, J., Munnich, A. & Melki, J. (1995) *Cell* **80,** 155–165.
19. Roy, N., Mahadevan, M. S., McLean, M., Shutler, G., Yaraghi, Z., *et al.* (1995) *Cell* **80,** 167–178.
20. Thompson, T. G., DiDonato, C. J., Simard, L. R., Ingraham, S. E., Burghes, A. H. M., Crawford, T. O., Rochette, C., Mendell, J. R. & Wasmuth, J. J. (1995) *Nat. Genet.* **9,** 56–62.
21. Kemler, R. (1993) *Trends Genet.* **9,** 317–321.
22. Geiger, B. & Ayalon, O. (1992) *Annu. Rev. Cell Biol.* **8,** 307–332.
23. Weiner, A. M., Deininger, P. L. & Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55,** 631–661.
24. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
25. Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C. & Kunkel, L. M. (1987) *Cell* **50,** 509–517.
26. McNally, E. M., Yoshida, M., Mizuno, Y., Ozawa, E. & Kunkel, L. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9690–9694.
27. Kleyn, P. Q., Wang, C. H., Lein, L. L., Vitale, E., Pan, J., Ross, B. M., Grunn, A., Palmer, D. A., Waeburton, D., Brzustowicz, L. M., Kunkel, L. M. & Gilliam, T. C. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6801–6805.
28. Gilliam, T. C., Freimer, N. B., Kaufmann, C. A., Powchik, P. P., Bassett, A. S., Bengtsson, U. & Wasmuth, J. J. (1989) *Genomics* **5,** 940–944.
29. Altschul, S. F., Gish, W., Miller, W., Meyer, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
30. Monaco, A. O., Bertelson, C. J., Middlesworth, W., Colletti, C. A., Aldridge, J., Fischbeck, K. H., Bartlett, R., Pericak-Vance, M. A., Roses, A. D. & Kunkel, L. M. (1985) *Nature (London)* **316,** 842–845.
31. Lichter, P., Tang, C. C., Call, K., Hermanson, G., Evans, G. A., Housman, D. & Ward, D. C. (1990) *Science* **247,** 64–69.
32. Tanihara, H., Sano, K., Heimark, R. L., John, T. S. & Suzuki, S. (1994) *Cell Adhes. Commun.* **2,** 15–26.
33. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472.
34. Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M. & Weissenbach, J. (1994) *Nat. Genet.* **7,** 246–339.
35. Mansouri, A., Spurr, N., Goodfellow, P. N. & Kemler, R. (1988) *Differentiation (Berlin)* **38,** 67–71.
36. Walsh, F. S., Barton, C. H., Putt, W., Moore, S. E., Kelsell, D. & Spurr, N. (1990) *J. Neurochem.* **55,** 805–812.
37. Kaupmann, K., Becker-Follmann, J., Scherer, G., Jockusch, H. & Starzinski-Powitz, A. (1992) *Genomics* **14,** 488–490.
38. Deeb, S. S., Lindsey, D. T., Hibiya, Y., Sanocki, E., Winderickx, J., Teller, D. Y. & Motulsky, A. G. (1992) *Am. J. Hum. Genet.* **51,** 687–700.
39. Sargent, C. A., Chalmers, I. J., Leversha, M. & Affara, N. A. (1994) *Mamm. Genome* **5,** 791–796.