Vol. 206, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 5, 1995 Pages 294-301

IDENTIFICATION OF MULTIPLE TRANSCRIBED SEQUENCES FROM THE SPINAL MUSCULAR ATROPHY REGION ON HUMAN CHROMOSOME 5

Antonio Pizzuti[1], Alessia Colosimo[2], Antonia Ratti[1], Francesca Capon [2], Massimo Gennarelli[2], Vincenzo Silani[1], Cristina Ghezzi[1], Stefania Lo Cicero[2], Giuseppe Calabrese[4], Giandomenico Palka [4], Guglielmo Scarlato[1], Giuseppe Novelli[3]*and Bruno Dallapiccola[2,5]

[1]Istituto di Neurologia, Università di Milano, Centro Dino Ferrari, Milano, Italy

[2]Dipartimento di Sanità Pubblica e Biologia Cellulare, Università Tor Vergata, Roma, Italy

[3]Cattedra di Genetica Umana, Università Cattolica del Sacro Cuore, Roma, Italy

[4] Istituto di Biologia e Genetica, Università di Chieti, Italy

[5]Ospedale CSS, San Giovanni Rotondo, Italy

Received November 29, 1994

Summary: We report the isolation and characterization of novel expressed sequences from the spinal muscular atrophy (SMA) region on human chromosome 5q13. Based on the sequence homology studies these cDNAs were grouped in four classes, one of which shows extensive homologies with the β-glucuronidase (BG) gene, differing in exon arrangement. The other cDNAs do not show any strong homology with known DNA sequences.

Introduction: Childhood spinal muscular atrophies (SMAs) are autosomal recessive disorders characterized by degeneration of the anterior horn cells in the spinal cord (1). Based on age of onset and life-span, SMAs are classified into types I, II, and III, which are considered allelic forms of a locus in chromosome 5q11.2-q13.3 (2-4). Several strategies have been used to isolate transcripts from the SMA region between D5S823 and D5S557 (5). Different expressed candidate sequences were characterized by DNA sequencing, database comparison, transcript size, expression pattern, and in a few cases, DNA sequence analysis in affected individuals (5-8). At present, none of these sequences looks appropriate as a candidate gene for SMA. We report the isolation of several novel spinal cord expressed

*To whom correspondence should be addressed at Laboratorio Centrale CRI - Via Ramazzini 15 00151 ROMA (ITALY). FAX:+ 39/6/65741183.

0006-291X/95 $5.00

sequences from the SMA region by direct cDNA selection using cosmid DNA derived from YAC clones mapping to 5q13.

Materials and Methods

Cosmid library construction and screening: Total yeast DNA from YAC clones 821A9 and 908D1 of the CEPH mega-YAC library was partially digested with MboI restriction enzyme (9). Fragments were ligated into BamHI digested superCos cosmid vector (Stratagene). The cosmid library was packaged using Gigapack gold packaging extract (Stratagene). XL-1 blue E.coli cells were infected and 5 X 10^3 clones/dish were plated in 150 mm LB agar Petri dishes. Human genomic clones were isolated by hybridization of the library with total human genomic DNA.

cDNA direct selection: *Biotin-DNA fragments conjugation.* Twenty cosmid clones were digested (100 ng DNA each) to completion with MboI. Fragment protruding ends were filled using a deoxynucleotide mix containing biotin-conjugated dUTP and Klenow fragment of DNA polymerase I.

cDNA synthesis. Total RNA was extracted from human fetal spinal cord using the single step acid-guanidinium thiocyanate-phenol-chloroform extraction method (10). Fetal tissues were obtained from pregnancies interrupted according the current Italian regulation. Poly-A^+ RNA was isolated by filtration on oligo-dT cellulose. Poly-A^+ RNA (5 μg) was treated with RNAse-free DNAse (Pharmacia) 1 U in 20 μl final solution at 37°C for 1 hour. After DNAse heat inactivation ds cDNA was synthesized using the TimeSaver cDNA synthesis kit (Pharmacia). One μg of poly-A^+ RNA was retrotranscribed using 0.5μg of oligodT and 0.75μg of random hexamers as polymerization primers. Two hundred μg cDNAs were ligated to Not I linkers. After NotI digestion, 200-800 bp cDNA molecules were size-selected on 1% agarose gel and ligated to a NotI adapter (5' GGCCGCAAGCATGCGAATTCAGGATCC 3'). One to four ng of cDNA were amplified by PCR using an oligonucleotide complementary to the adapter sequence.

Hybridization of cDNA to cosmid DNA. Amplified cDNA (500 ng) and biotin-labelled cosmid fragments (50 ng) were denatured and hybridized for 48 hours at 65°C. The hybridization was carried out in 100 μl of 5 X SSPE, 5X Denhardt's, 0.1% SDS. Ten μg of human genomic DNA, 5μg of yeast DNA and 1μg of pBR 322 were used to compete for repetitive sequences. The hybridization mix was completed by adding 50 μl of streptavidin-conjugated magnetic beads (DYNABEADS - Dynal, Inc). The beads-biotinylated cosmid fragment-cDNA complexes were separated magnetically. After washing in 0.5 X SSC at 65°C for 15 min, the cDNAs were recovered by boiling (5 min at 100°C). The released cDNAs were amplified by PCR using the oligonucleotide complementary to the NotI adapter sequence. The resulting cDNAs were directly cloned in a TA-cloning vector (Invitrogen).

Analysis of cDNAs: Twenty four cloned cDNAs were amplified by PCR using the adapter primer. PCR products were blotted onto nylon membrane (Hybond N, Amersham Int.) and hybridized to human genomic DNA, yeast genomic DNA and pBR 322 in order to rule out clones containing repetitive sequences, vector DNA contaminants and ribosomal cDNAs.

Negative cDNAs were used to hybridize dotted grids of cosmid DNA. cDNAs hybridizing one or more cosmids were sequenced using the automated 370A DNA Sequencer (Applied Biosystem).

Expression studies: Multiple tissue poly(A)$^+$ RNAs were purchased from Clontech Laboratories. Reverse transcription of mRNAs into single-stranded cDNA was essentially carried out according to the method of Kawasaki (11) and by using the cDNA synthesis kit supplied by Clontech. The reaction was performed with avian myeloblastosis virus reverse transcriptase, oligo $(dT)_{15}$ primers, RNasin (Amersham) and 0.5 μg of mRNAs in a final volume of 25 μl. Aliquots of 10 μl were used in the PCR experiments. Human β-actin (Clontech) was used as an internal standard.

Fluorescent *in Situ* hybridization: Metaphase chromosomes were prepared from human peripheral blood lymphocytes as described (12). Purified cosmid DNA was labelled with biotinylated-dCTP and dUTP by nick translation and hybridized to chromosome spreads according to routine procedures. Chromosomes were counterstained with propidium iodide and visualized as described.

Results

Isolation and characterization of cDNAs: Two groups of ten randomly picked human cosmids were used for cDNA direct selection. Some of them contained D5F149S1-S2 and D5F150S1-S2, which are the most tightly linked markers to the SMA locus (9). Twenty-four plasmid clones were analyzed for the presence of chromosome 5 specific transcripts by hybridization to the cosmids and successive sequencing of the positive clones. Among the 24 screened clones two cDNAs containing Alu sequences and two clones that did not hybridize any cosmid were discarded (Table 1). All the remaining cDNA clones hybridized cosmids derived from chromosome 5, as demonstrated by FISH analysis of the cosmid themselves (data not shown). Based on sequence homologies, positive clones were assigned to four cDNA groups (Table 1). Four partially overlapping sequences represent the OL-1 cDNA group, spanning a 389 bp contig. Two different clones constitute the FS-1 group, spanning 290 bp. The third cDNA group, CB-1 (241 bp), consists of four additional overlapping clones. The last cDNA sequence group (DR) is formed by 10 partially

Table 1: Selected cDNAs

cDNA	No. clones	length (bp)	sequence homology
n.h.	2	n.t.	n.t
AL	2	140	alu
OL-1	4	389	n.f.
FS-1	2	290	n.f.
CB-1	4	241	n.f.
DR:			
DR-1	4	275	BG
DR-2	3	162	BG
DR-3	3	329	BG

n.h., not hybridizing clones; n.t., not tested; n.f., not found. BG, β-glucuronidase (GUSB) cDNA. cDNA sequences were submitted to GenBank.

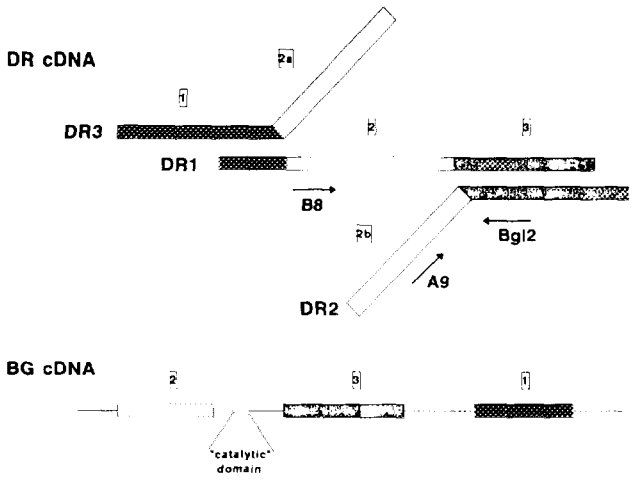

Fig. 1.
Schematic representation of the DR cDNAs and the homologous regions of the BG cDNA (see text). Oligonucleotide B8 (5' GGT GAA TTA CCA GAT CTC CGTC 3'), A9 (5' AGG ATG GGA GGT TGT CTG AGCC 3') and Bgl2 (5' CTT TGG TGT GAG CAA TCA CCA TC 3') are shown with arrows.

homologous sequences. Sequences in OL-1, FS-1, and CB-1 groups did not demonstrate significant homologies to nucleotide sequences in GenBank, using FSTSCAN program (PC/GENE, IntelliGenetics). On the contrary, DR cDNAs showed extensive homologies to translated sequences belonging to the human and rodent β-glucuronidase gene (BG) (GUSB, E.C.3.2.1.31). In humans, the BG gene maps to chromosome 7q11.23-q21 and its deficiency is responsible for mucopolysaccharidosis type VII (MIM*253220) (13,14). The DR group is not homogeneous and three partially overlapping sequences were recognized (DR-1, DR-2 and DR-3) (Table 1). They probably represent alternatively spliced forms of the same gene. As shown in Fig. 1, cDNA DR-1 and cDNA DR-3 overlap at their 5' region, then the sequences diverge. DR-2 overlaps to DR-1 only at the 3' sequence. The organization of homologous regions in the DR cDNAs and in the BG gene are different (Fig.1). DR exon(s) in region 1 are homologous to nucleotides 1545 to 1681 of BG cDNA. Region 2 of DR corresponds to nucleotide 752 to 939 of BG cDNA and region 3 shows homologies from nucleotide 1420 to 1465 (Fig.2). Thus, in DR the arrangement of region 1, 2 and 3 from 5' to 3', corresponds to a different 2, 3, 1 arrangement in the BG cDNA (Fig.1). Regions 2b of DR-2 and 2a of DR-3 are not homologous to any sequence in the BG cDNA and to any known gene. DR cDNAs do not contain homologies to the region probably coding for the BG catalytic domain (15). In BG cDNA this region separates sequences that are contiguous in DR (Fig. 1). The first portion of DR-1 region 2 is homologous to STS UT518 containing a polymorphic tetranucleotide repeat mapped to chromosome 5 and deposited in GeneBank (accession number L18312). The presence of UT518 in the cosmids containing DR-1 sequences, was demonstrated by PCR experiments (data not shown). Cosmids hybridization pattern of cDNA groups indicates that CB-1, FS-1 and OL-1 sequences map very close each

```
REGION  1

BG 1545 TGGTATCACGACTACGGGCACCTGGAGTTGATTCAGCTGCAGCTGGCCACCCAGTTTGA
        ||||||| | ||||||||||||||||||||||||||||||||||||||| ||||||||||
        TGGTATCGCAACTACGGGCACCTGGAGTTGATTCAGCTGCAGCTGGCCGCCCAGTTTGA


        GAACTGGTATAAGAAGTATCAGAAGCCCATTATTCAGAGCGAGTATGGAGCAGAAACGA
        ||| |||| |||||   |||| || ||||||||||||||| ||||||||  ||||||
        GAATTGGTGTAAGAC--ATCACAATCCCATTATTCAGAGCGCGTATGGAGTGGAAACGC


        TTGCAGGGTTTCACCAGG   BG 1680
        ||| |||| |||||||||
        TTGTAGGGCTTCACCAGG

REGION  2

BG 749  GGCTGGTGAATTACCAGATCTCTGTCAAGGGCAGTAACCTGTTCAAGTTGGAAGTGCG
        ||||||||||||||||||||| |||||||||||||| |||||||||||||||| |
        GGCTGGTGAATTACCAGATCTCCGTCAAGTGCAGTAACCAGTTCAAGTTGGAAGTGTG


        TCTTTTGGATGCAGAAAACAAAGTCGTGGCGAATGGGACTGGGACCCAGGGCCAACTT
        ||||||| ||||||||||||||||||||  ||   | |||||||||||||||| ||
        TCTTTTGAATGCAGAAAACAAAGTCGTGGACAACCAGGCTGGGACCCAGGGCCAGCTG


        AAGGTGCCAGGTGTCAGCCTCTGGTGGCCGTACCTGATGCACGAACGCCCTGCCTATC
        |||||||  |||| || ||||||||||||||||||||||||||||| ||| ||||| |
        AAGGTGCTGGGTGCCAACCTCTGGTGGCCGTACCTGATGCACGAACACCCCGCCTACC


        TGTACTCATTGGAGG   BG 939
        ||||||| | |||||
        TGTACTCGTGGGAGG

REGION  3

BG 1419 GATGGTGATCGCTCACACCAAATCCTTGGACCCCTCCCGGCCTGTGACCTTTGTGAG
        ||||||||| |||||||||||| |||||||||||||| ||||||||||||||||
        GATGGTGATTGCTCACACCAAAGCCTTGGACCCCTCCCAGCCTGTGACCTTTGTGAC

        CAACTCTAACTATGCAGCAGACAA   BG 1499
        ||||||   ||| ||||||||||
        CAACTCGTCCTACGCAGCAGACAA
```

Fig. 2.
Sequence homologies between the DR-1 cDNA and the corresponding regions of BG cDNA. The positions of the homologous regions in BG cDNA (15) are shown by numbers at the beginning and the end of the sequences.

other and to the 5' of DR. OL-1 and CB-1 are open sequences in more the one strand, while FS-1 reading frames are interrupted by many stop codons.

DR expression studies: In order to define the expression pattern of DR and of the alternative spliced forms, a set of three oligonucleotide (A9, B8, and Bgl2) were synthesized

and used in RT-PCR experiments (Fig.1). As espected from the DR-1 sequence, B8 and Bgl2 primers amplify a ~220 bp fragment from brain, skeletal muscle, leukocytes, placenta and thymus. Genomic DNA PCR does not show any product, indicating DR regions 2 and 3 are separated by long intronic sequences. As anticipated by the DR-2 cDNA sequence, A9 and Bgl2 primers amplify a ~150 bp in both genomic and RT-PCR.

Discussion

We report the isolation and characterization of four novel human expressed sequences, from fetal human spinal cord RNA, by direct cDNA selection, using cosmid DNA derived from YAC clones mapping to 5q13 (9). cDNA direct selection takes advantage of screening multiple tissue at the same time and more cDNA molecules, compared to protocols based on hybridization of cDNA libraries. Previously, we reported the isolation of cDNA clones from the SMA region using total YAC as driver DNA (16). The limit of this approach is the time-consuming step for the verification of many selected cDNAs using YAC and genomic DNA Southern blotting (16). The subcloning of the YAC in cosmids adds a preliminary step to the direct selection. However, it allows a much faster and easier screening of the selected cDNA, very well compensating for the initial time loss. In the present study we obtained a very high yield of positive cDNA clones (83%) recognizing four sequence groups. Three of these cDNA groups (OL-1, FS-1, CB-1) did not show any significant similarity with other gene sequences in the GeneBank data base. On the contrary, the fourth group (DR) demonstrated over 90% homology to the BG gene cDNA (15). Within the DR cDNA group we recognized three different cDNA forms. DR-1 shows the major identity to the BG cDNA, with three successive regions (1,2 and 3), homologous to different portions of the BG cDNA (Fig. 1 and 2). Differently from DR-1 these regions are not contigous in the BG cDNA, and their arrangement in BG is different, with region 1 following region 2 and 3 (Fig. 1). Homologous regions 2 and 3 in BG are separated by sequences probably coding for the catalytic domain, which are not found in DR. In fact, two different mRNA types have been described for the BG gene, due to alternative splicing (15). The shorter form, lacking bases 939-1091, encodes a protein without glucuronidase activity. As shown in Fig. 2 also the DR-1 cDNA lacks sequences homologous to the "active" region (from base 939 to 1419 of the BG cDNA). The sequence data suggest that DR is not simply another BG gene and that the DR gene product(s) lacks glucuronidase activity. DR-2 and DR-3 cDNAs appear as alternative spliced forms of the DR gene at the region 2, with long portions not homologous to the BG gene (2a and 2b). RT-PCR experiments using primers able to distinguish the amplification of the DR cDNA from BG, conferm its expression in different tissues and in different spliced forms. Moreover preliminary data also suggest that DR alternative forms may be differentially expressed by tissues, for instance during central nervous system development.

OL-1, CB-1 and FS-1 cDNAs map very close to each other (we isolated cosmid clones hybridizing to all of them) and to the 5' region of DR. From present evidences we cannot exclude that they belong either to a common gene or to the same gene as DR. However RT-PCR assays using fetal brain and spinal cord RNA with primers designed on DR sequences

(A9, B8, Bg2) and primers designed on the other cDNAs have failed to show any specific PCR product. This issue will be better addressed when the full length cDNA clones for each of them are isolated. The chromosomal localization and tissue origin (fetal spinal cord) of these four cDNA, suggests a possible role in SMA. We are currently screening SMA patient genomic DNA for gross alterations of these genes by restriction mapping. Full-length cDNA clones are being isolated and evaluated as candidate genes by sequencing and SSCP analysis of normal and SMA patient mRNA. Moreover, we are investigating the tetranucleotide polymorphic repeat UTR518 contained in DR gene to determine linkage disequilibrium with the disease locus.

Acknowledgments

We thank Donata Penso for technical assistance in automated sequencing and D. Toniolo (YAC Screening Center, Milan) for providing YAC clones. This research was supported by the Italian Telethon (Grant No.296), MURST (fondi 60%) and ASM, Milan.

References

1. Dubowitz, V. (1978) Muscle disorders of childhood. Saunders, Philadelphia.
2. Brzustowicz, L.M., Lenher, 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 344, 540-541.
3. Melki, J., Sheth, P., Abdelhak, S., Burlet, P., Bachelot, M.F., Lathrop, M.G., Frezal, J., Munnich, A. (1990) Lancet 336, 271-273.
4. Gillian, T.C., Brzustowicz, L.M., Castilla, L.H, Lenher, T., Penchaszadeh, G.K., Daniels R.J., Byth, B.C., Knowles, J., Hislop, J.E., Shapira, Y., Dubowitz, V., Munsat, T.L., Ott, J., Davies, K.E. (1990) Nature 345, 823-825.
5. Davies, K.E., Theodosiou, A.M., Morrison, K.E., Nesbit, A.M., Daniels, R.J., Campbell, L., Francis, M.J., Christodoulou, Z., Ignatius, J. (1994) Am.J.Hum.Genet. 55 (Suppl.) Abst. 57.
6. Kleyn, P.W., Chang, C.H., Vitale, E., Carter, T.A., Ross, B.M., Jie, P., Palmer, D.A., Grunn, A., Bonaldo, M.F., Shah, A., Penchazadeh, G., Das, K., Brzustowicz, L.M., Lien, L.L., Selig, S., Perls, K.B., Parano, E., Pavone, L., Munsat, T., Warburton, D., Soares, M.B., Kunkel, L.M. and Gilliam, T.C. (1994) Am.J.Hum.Genet. 55 (Suppl.) Abst. 1312.
7. Di Donato, C.J., Carpten, J., Morgan, K., Fuerst, P., Mcpherson, J., Ingraham, S., Prescott, G., Dadze, A., Wirth, B., Wasmuth, J.J., Simard, L., and Burghes, A.H.L. (1994) Am.J.Hum.Genet. 55 (Suppl.) Abst. 153.
8. Carter, T.A., Wang, C.H., Vitale, E., Ross, B.M., Xu, J., Grunn, A., Lien, L.L., Selig, S., Kunkel, L.M. and Gilliam, T.C. (1994) Am.J:Hum:Genet. 55 (Suppl.) Abst. 154.
9. Melki, J., Lefebvre, S., Burglen, L., Burlet, I., Clermont, O., Milasseau, P., Reboullet, S., Benichou, B., Zeviani, M., Le Pasler, D., Cohen, D., Weissenbach, J. and Munnich, A. (1994) Science 264, 1474-1477.

10. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 167, 157-159.
11. Kawasaki, E.S. (1991) In: PCR protocols, a guide to methods and applications, pp. 21-27, Innis et al., eds., Academic Press, Inc., San Diego, CA.
12. Baldini, A. and Word, D.C. (1991) Genomics 9, 770-774.
13. Benn, P., Chern, C.J., Bruns, G.A.P., Graig, I.W. and Croce, C:M. (1977) Cytogenet. Cell Genet. 19, 273-280.
14. Bell, C.E.Jr., Sly, W.S. and Brot, F.E. (1977) J.Clin.Invest. 59, 97-105.
15. Oshima, A., Kyle, J.W., Miller, R.B., Hoffmann, J.W., Powell, P.P., Grubb, J.H., Sly, W.S., Tropak, M., Guise, K.S., Gravel, R.A. (1987) Proc.Natl.Acad.Sci. U.S.A. 84, 685-689.
16. Pizzuti, A., Gennarelli, M., Novelli, G., Colosimo, A., Lo Cicero, S., Caskey, C.T., and Dallapiccola, B. (1993) Biochem.Biophis.Res.Comm. 197, 154-162.