Cell, Vol. 80, 155–165, January 13, 1995, Copyright © 1995 by Cell Press

Identification and Characterization of a Spinal Muscular Atrophy-Determining Gene

Suzie Lefebvre,*† Lydie Bürglen,*†
Sophie Reboullet,* Olivier Clermont*,
Philippe Burlet,* Louis Viollet,*
Bernard Benichou,* Corinne Cruaud,‡
Philippe Millasseau,§ Massimo Zeviani,*‖
Denis Le Paslier,§ Jean Frézal,*
Daniel Cohen,§ Jean Weissenbach,‡
Arnold Munnich,* and Judith Melki*
*Unité de Recherches sur les Handicaps Génétiques de l'Enfant
Institut National de la Santé et de la Recherche Médicale
Unité 393
Institut Necker
Hôpital des Enfants Malades
75743 Paris Cedex 15
France
‡Genethon
91002 Evry
France
§Centre d'Etudes du Polymorphisme Humain
75010 Paris
France

Summary

Spinal muscular atrophy (SMA) is a common fatal autosomal recessive disorder characterized by degeneration of lower motor neurons, leading to progressive paralysis with muscular atrophy. The gene for SMA has been mapped to chromosome 5q13, where large-scale deletions have been reported. We describe here the inverted duplication of a 500 kb element in normal chromosomes and narrow the critical region to 140 kb within the telomeric region. This interval contains a 20 kb gene encoding a novel protein of 294 amino acids. An highly homologous gene is present in the centromeric element of 95% of controls. The telomeric gene is either lacking or interrupted in 226 of 229 patients, and patients retaining this gene (3 of 229) carry either a point mutation (Y272C) or short deletions in the consensus splice sites of introns 6 and 7. These data suggest that this gene, termed the survival motor neuron (*SMN*) gene, is an SMA-determining gene.

Introduction

Spinal muscular atrophies (SMAs) are characterized by degeneration of the anterior horn cells of the spinal cord, leading to progressive symmetrical limb and trunk paralysis associated with muscular atrophy. SMA represents the second most common fatal autosomal recessive disorder after cystic fibrosis (1 in 6000 newborns) (Roberts et al., 1970; Pearn, 1973, 1978; Czeizel and Hamula, 1989). Childhood SMA is classically subdivided into three clinical groups on the basis of age of onset and clinical course (Munsat, 1991). The acute form of Werdnig–Hoffmann disease (type I; Werdnig, 1894; Hoffmann, 1900) is characterized by severe, generalized muscle weakness and hypotonia at birth or within the next 3 months. Death from respiratory failure usually occurs within the first 2 years. This disease may be distinguished from the intermediate (type II) and juvenile (type III or Kugelberg–Welander disease; Kugelberg and Welander, 1956) forms. Type II children are able to sit, although they cannot stand or walk unaided, and survive beyond 4 years. Type III patients have proximal muscle weakness, starting after the age of 2. The underlying biochemical defect(s) remains unknown.

By means of linkage analysis, we and others have shown that all three forms of SMA map to chromosome 5q11.2-q13.3 (Brzustowicz et al., 1990; Melki et al., 1990a, 1990b, 1993; Gilliam et al., 1990; Sheth et al., 1991; Lien et al., 1991; Morrisson et al., 1992; Soares et al., 1993; Clermont et al., 1994). Various yeast artificial chromosome (YAC) contigs of the 5q13 region spanning the disease locus have been constructed and the presence of low copy repeats in this region demonstrated (Kleyn et al., 1993; Francis et al., 1993; Melki et al., 1994). Allele segregation was analyzed at the closest genetic loci, detected by markers derived from the YAC contig (*C212*, *C272*, and *C161*), in 201 SMA families. These markers revealed either two (*C212* and *C272*) or three (*C161*) loci in the 5q13 region. Inherited or de novo deletions were observed in nine unrelated SMA patients. Moreover, deletions were strongly suggested in at least 18% of SMA type I patients by the observation of marked heterozygosity deficiency for the loci studied. These results indicated that deletion events are statistically associated with the severe form of SMA (type I). Recently, similar results were obtained with the *Ag1-CA* marker in type I SMA individuals (DiDonato et al., 1994). Studying all polymorphic DNA markers derived from the YAC contig, we observed that the smallest rearrangement occured within a region bordered by loci detected by *C161* and *C212–C272* and entirely contained in a 1.2 Mb YAC clone, 903D1 (Melki et al., 1994).

In this report, we present the characterization of the small (140 kb) nested critical SMA region by a combination of genetic and physical mapping in SMA patients. This region suggested a precise location for the SMA gene and, therefore, a limited region within which to search for candidate genes. We identified a duplicated gene in the 5q13 region, one of which, the survival motor neuron (*SMN*) gene, was located within the critical region. This gene was lacking in 213 of 229 (93%) or interrupted in 13 of 229 (5.6%) SMA patients. In patients in which the *SMN* gene was neither lacking nor interrupted, the presence of deleterious mutations provided strong evidence that this gene is an SMA-determining gene.

†The first two authors contributed equally to this work.

‖Present address: Istituto Nazionale Neurologico Carlo Besta, 20113 Milano, Italy.

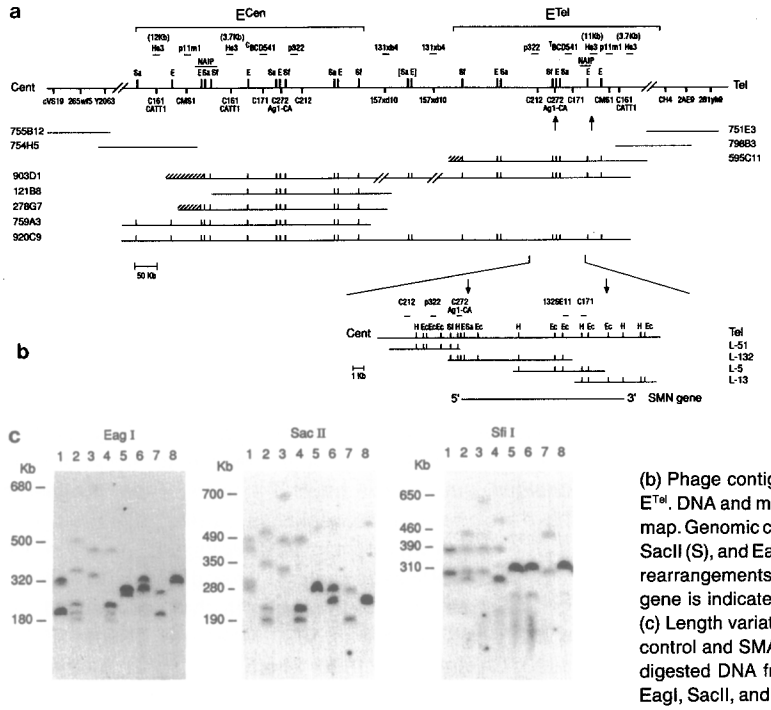

Figure 1. Physical Map of the SMA Locus

(a) Organization of the YAC contig spanning the SMA region. DNA and microsatellite markers are indicated above and below the genomic map, respectively. Above the *He3* marker are indicated in parenthesis the sizes of the HindIII restriction fragments detected by this marker. The genomic cut sites for the enzymes SfiI (Sf), EagI (E), and SacII (S) are shown. The positions of the restriction sites and markers provided evidence for an inverted duplication of an element that was divided into E^{Tel} and E^{Cen} according to location. Arrows indicate the limits of the genomic rearrangements observed in SMA patients. The positions of the *NAIP* genes with respect to ^{C}*BCD541* and ^{T}*BCD541* (*SMN*) genes are indicated. Abbreviations: Cent, centromere; Tel, telomere. Hatched bars indicate chimeric YAC clones.

(b) Phage contig spanning the *SMN* gene in the critical region of the E^{Tel}. DNA and microsatellite markers are indicated above the genomic map. Genomic cut sites for the enzymes HindIII (H), EcoRI (Ec), SfiI(Sf), SacII (S), and EagI (E) are shown. Arrows indicate the limits of genomic rearrangements observed in SMA patients. The orientation of the *SMN* gene is indicated.

(c) Length variation of SfiI, EagI, and SacII restriction fragments from control and SMA individuals identified with probe *132SE11*. PFGE of digested DNA from controls (lanes 1–4) and SMA (lanes 5–8) using EagI, SacII, and SfiI was probed with the clone 132SE11. PFGE analysis revealed a high degree of restriction fragment variability using both enzymes insensitive (SfiI) and sensitive (EagI and SacII) to sequence-specific methylation.

Results

Pulsed Field Gel Electrophoresis Analysis Reveals a Large Inverted Duplication and a Complex Genomic Organization of the 5q13 Region in Normal Chromosomes

A YAC contig of the 5q13 region spanning the disease locus has been constructed. The haplotypes of YACs at the polymorphic loci detected by microsatellite markers *C212*, *C272*, *C171*, *C161* (Melki et al., 1994), *CMS1* (Kleyn et al., 1993), and *2AE9* (Francis et al., 1993) were compared with that of the human donor whose genomic DNA was used to construct the Centre d'Etudes du Polymorphisme Humain (CEPH) YAC libraries. *Ag1-CA* and *CATT1* primers are within the sequence of markers *C272* and *C161*, respectively (DiDonato et al., 1994; Burghes et al., 1994). All the subloci detected by these markers on each chromosome of the human donor were present in this YAC contig and confirmed the extension of the YAC contig over the entire SMA region (Melki et al., 1994; Figure 1a). Consequently, the resulting contig was larger than those reported by the other groups (Kleyn et al., 1993; Francis et al., 1993; McLean et al., 1994; DiDonato et al., 1994).

To elucidate the physical organization of the SMA candidate region, we performed long-range restriction mapping of the YAC contig using pulsed field gel electrophoresis (PFGE), and the restriction enzymes SacII, SfiI, and EagI were used to digest the YACs containing the telomeric loci detected by markers *C212*, *C272*, *C171*, and *C161* (YAC clone 595C11); the centromeric loci detected by these markers (YAC clones 121B8, 759A3, and 278G7); or both loci (YAC clones 903D1 and 920C9). λ phage libraries of YACs 595C11, 121B8, and 903D1 were constructed, and subclones p322, 132SE11, He3, 131xb4, and p11M1 were isolated from phages containing microsatellite markers *C212*, *C272*, *C161*, *AFM157xd10* (Melki et al., 1994), and *CMS1* (Kleyn et al., 1993), respectively. These were used as probes for PFGE analysis. The restriction map of the region (Figure 1a) shows that probes *132SE11*, p11m1, and p322 revealed two loci and probe *He3* revealed four loci on the YAC contig, whereas probe *131xb4* revealed several loci on 5q13 and 5p (data not shown). The positions of the restriction sites and markers provided evidence for a large inverted duplication of an element of approximately 500 kb, termed E^{Tel} and E^{Cen} for the telomeric and centromeric elements, respectively (Figure 1a). Interestingly, PFGE analysis of the region revealed a high degree of variability in both SMA and control individuals. This variability could be due to a variable number of repeated sequences in the general population and hampered our ability to distinguish E^{Tel} from E^{Cen} and to recognize abnormal restriction fragments in SMA patients (Figure 1c).

Southern Blot Analysis Enables the E^{Tel} and E^{Cen} of the Duplication to Be Distinguished

DNA from the YAC clones of the contig and from the SMA patients and controls was digested with restriction enzymes SacI, KpnI, MspI, PstI, PvuII, EcoRI, HindIII, BglII, and XbaI for Southern blotting and hybridized with clones *132SE11*, p11m1, *He3*, *131xb4*, and p322 as probes. Only one probe, *He3*, derived from phage containing microsatellite marker *C161*, detected a difference between the two

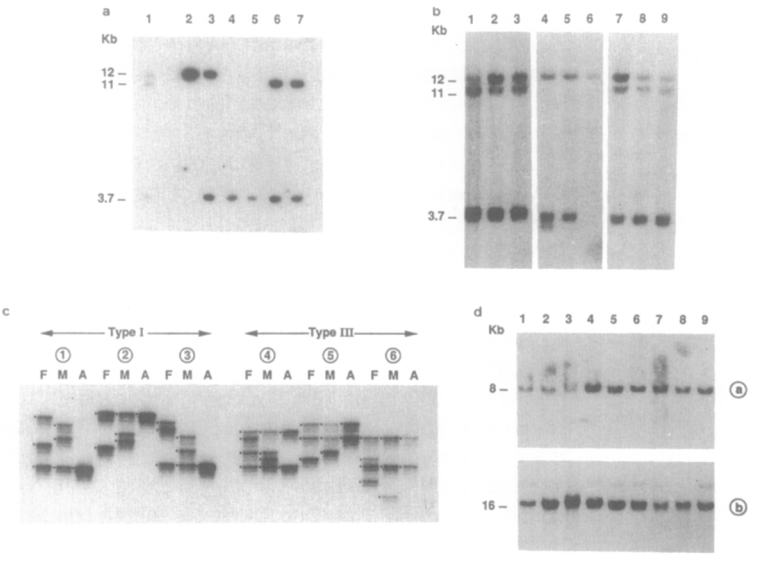

Figure 2. Genetic and Physical Mapping of Genomic Rearrangements in SMA Individuals

(a) Southern blot analysis of HindIII-digested DNA from YAC clones and the human donor probed with *He3*. The genomic probe *He3*, derived from the phage containing marker *C161*, detected three restriction fragments of 12 kb, 11 kb, and 3.7 kb in the human donor whose genomic DNA was used to construct the CEPH YAC libraries. Southern blot analysis from HindIII-digested DNA of YAC clones and the human donor (lane 1), probed with *He3*, detected the 12 kb fragment in YAC clones 754H5 (lane 2) and 759A3 (lane 3), indicating that this fragment was specific for the most proximal locus detected by this probe. The presence of the 11 kb fragment in YAC clones 903D1 (lane 7) and 595C11 (lane 6) indicated that this fragment was specific for the E^{Tel}. The detection of the 3.7 kb fragment in the nonoverlapping YAC clones containing E^{Cen} (278G7, lane 4; 121B8, lane 5) or B^{Tel} (595C11, lane 6) indicated that this fragment corresponded to two different loci in each element.

(b) Southern blot analysis of HindIII-digested DNA from controls (lanes 1–3) and SMA patients (lanes 4–9) probed with *He3*. The 11 kb telomeric-specific HindIII fragment, detected by probe *He3*, was absent in consanguineous type I patients (lane 4–6) and, in one patient, also involved the 3.7 kb HindIII fragment (lane 6). By contrast, the 11 kb telomeric-specific fragment was present in consanguineous type III patients (lane 7–9).

(c) Allele segregation analysis at loci detected by marker *C272* in consanguineous SMA patients. The figure shows family studies based on microsatellite DNA marker *C272*. The SMA patients belonged to type I (families 1, 2, and 3) or type III (families 4, 5, and 6). Consanguineous type I SMA patients had one single PCR amplification product, as compared with 0 of 59 controls (Melki et al., 1994). By contrast, consanguineous type III SMA patients had two *C272* amplification products, inherited from both parents, indicating homozygosity by descent, at each locus detected by marker *C272*. Abbreviations: F, father; M, mother; A, affected. Dots indicate allelic fragments.

(d) Gene dosage analysis of the 5q13 region with the *132SE11* probe in consanguineous SMA patients and controls. Total human DNA from SMA patients or controls was digested with HindIII restriction enzyme for Southern blotting. Filters were consecutively hybridized with *132SE11* (a) and an internal anonymous control probe *122P1* (b). Gene dosage was determined by densitometric scanning of the hybridization signals. A marked decrease in *132SE11* band intensity was observed in consanguineous type I SMA patients (40%, 14%, and 20% in lanes 1–3, respectively) as compared with the controls (115%, 85%, and 100% in lanes 4–6, respectively). By contrast, no gene dosage effect was observed in consanguineous type III patients, indicating the absence of deletions involving the locus detected by *132SE11* (120%, 92%, and 88% in lanes 7–9, respectively).

duplicated elements in control individuals. Upon analysis of total human DNA, probe *He3* detected three HindIII restriction fragments of 12 kb, 11 kb, and 3.7 kb, respectively (Figure 2a). The 12 kb HindIII restriction fragment was present in YAC clones 754H5 and 759A3, indicating that this fragment corresponded to the most proximal locus of the E^{Cen} (Figure 2a). The 11 kb HindIII fragment was present in YAC clones 595C11, 903D1, and 920C9, indicating that this fragment corresponded to a single locus of the E^{Tel} (Figure 2a). Finally, the 3.7 kb HindIII fragment was present in the nonoverlapping YAC clones 595C11 and 121B8, containing E^{Tel} and E^{Cen}, respectively, indicating that this fragment corresponded to two different loci (Figure 2a). Similar results were obtained using restriction enzymes SacI and KpnI (data not shown). The three restriction fragments detected by *He3* were observed in the monochromosomal hybrid HHW105 (Carlock et al., 1985) and in 30 unrelated healthy individuals, confirming that these fragments were not due to allelic polymorphisms (data not shown). These results allowed E^{Tel} to be distinguished from E^{Cen} in both controls and SMA patients.

The Combined Use of Genetic and Physical Mapping to Identify Genomic Rearrangements in the E^{Tel} of SMA Patients

Among 201 SMA patients previously analyzed with markers *C212*, *C272*, and *C161*, nine carried large-scale deletions of a single mutant chromosome and were selected in the present study (Melki et al., 1994). In addition, 29 inbred SMA families of various ethnic origins were studied at the most closely flanking polymorphic loci (D5S629, D5S637, D5S351, D5S435, D5S1364, D5S1365, and D5S1370; data not shown but available on request). Homozygosity at these loci in affected children born to consanguineous parents provided evidence for identical mutations on both mutant chromosomes. We investigated further 12 of 13 type I, 1 of 6 type II, and 9 of 10 type III SMA patients who were homozygous by descent.

The genomic DNA of the nine patients harboring large-scale deletions of a single mutant chromosome and the 21 type I and III consanguineous patients showing homozygosity by descent was digested with HindIII for Southern blotting and hybridized with probe *He3* (Figure 2b). The 11 kb telomeric-specific fragment was absent in 11 of 12 consanguineous type I patients. In 2 of 11 patients, the deletion also involved the 3.7 kb *He3* fragment. By contrast, the 11 kb telomeric-specific fragment was absent in only 1 of 9 consanguineous type III patients. In 4 of 9 patients harboring large-scale deletions of a single mutant chromosome, the 11 kb HindIII fragment was absent. Of particular interest was the absence of the 11 kb telomeric-specific fragment in the patient harboring a deletion of one

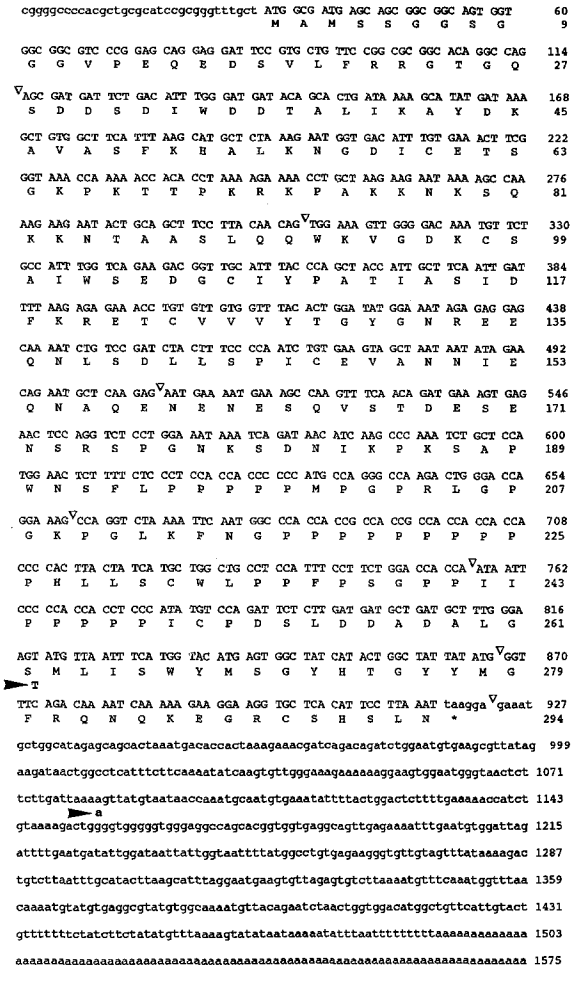

Figure 3. Nucleotide Sequence of *SMN* cDNA

The predicted amino acid sequence encoded by the *SMN* cDNA is shown below the DNA sequence, assuming that translation begins at the first in-frame methionine of the ORF. Triangles indicate the intron–exon junctions. Arrowheads indicate the position of the single base differences observed in exons 7 and 8, between the *SMN* and the centromeric genes.

of the two loci detected by markers *C212* and *C272* (Melki et al., 1994), indicating the presence of a deletion involving the telomeric loci detected by markers *C212*, *C272*, and *He3*. Taken together, these observations provided evidence for genomic rearrangements of the E^{Tel} in SMA patients. In addition, as all but one of the consanguineous type III patients were not deleted for this locus, these data supported the location of the SMA gene centromeric to the telomeric-specific *He3* locus (Figure 1a).

To define the centromeric boundaries of the genomic rearrangements in SMA, we analyzed the allele segregation at loci detected by microsatellite marker *C272* in the 21 type I and III consanguineous SMA patients showing homozygosity by descent (Figure 2c). All consanguineous type I SMA patients had only one single polymerase chain reaction (PCR) amplification product, a feature that was not observed in 59 controls (Melki et al., 1994). This reduction in the number of *C272* loci was due to the deletion of one of the two loci detected by *C272*, as indicated by the marked decrease in gene dosage with probe *132SE11*, which maps close to this marker (Figure 2d). By contrast, 8 of 9 consanguineous type III SMA patients had two *C272* amplification products, inherited from both parents, indicating homozygosity at each locus detected by marker *C272* (Figure 2c). Thus, no reduction in the number of *C272* loci was observed in all but one of the consanguineous type III patients. Similarly, no gene dosage effect was detectable with probe *132SE11*, indicating the absence of deletions involving these loci in consanguineous type III patients (Figure 2d). Our results suggested that the disease-causing gene is distal to the telomeric locus detected by *C272* (Figure 1a).

These studies placed the SMA gene within the E^{Tel}, between the telomeric locus detected by marker *C272* and the 11 kb telomeric-specific locus detected by *He3* (Figure 1a). Based on long-range restriction mapping of the YAC contig using PFGE, this critical region is entirely contained in a terminal 140 kb SacII–pYAC fragment of YAC clone 903D1 (or a 150 kb SacII–pYAC fragment of YAC clone 920C9; Figure 1a).

Construction of a Phage Contig Spanning the Critical Region of the E^{Tel}: Identification and Characterization of a Candidate Gene

Phage clones containing markers *C212*, *C272*, *C171*, and *C161* were isolated from the λ phage library constructed from YAC clone 595C11 and used as starting points for bidirectional walking. A phage contig of the E^{Tel} encompassing markers *C212*, *C272*, and *C171* was constructed based on the restriction map of the phage clones (Figure 1b). The probe *132SE11*, derived from the phage containing marker *C272*, gave positive hybridization signals with hamster DNA, indicating the presence of interspecies-conserved sequences. The screening of a λgt10 human fetal brain cDNA library with probe *132SE11* resulted in the selection of seven overlapping λ clones spanning 1.6 kb. Sequence analysis of the clones revealed an 882 bp open reading frame (ORF) and a 580 bp noncoding region (Figure 3). A 1.5 kb cDNA clone (*BCD541*) contained the entire coding sequence and most of the 3′ noncoding region. The 3′ end of the cDNA along with its poly(A)$^+$ tail was obtained by PCR amplification of a lymphoblastoid cell line cDNA library (see Experimental Procedures). Northern blot analysis of poly(A)$^+$ RNA from various tissues, including heart, brain, liver, muscle, lung, kidney, and pancreas, revealed the presence of a widely expressed 1.7 kb transcript (Figure 4a). This gene is also expressed in spinal cord (data not shown). The ORF encodes a putative protein of 294 amino acids with a predicted molecular mass of approximately 32 kDa (Figure 3). An homology search using the FASTA and BLAST networks failed to detect any homology at either the nucleotide or the amino acid levels. Anticipating the results presented below, this candidate gene was called the survival motor neuron (*SMN*) gene.

Evidence for a Duplication of the *SMN* Gene in the 5q13 Region

The *BCD541* cDNA clone was used as a probe for South-

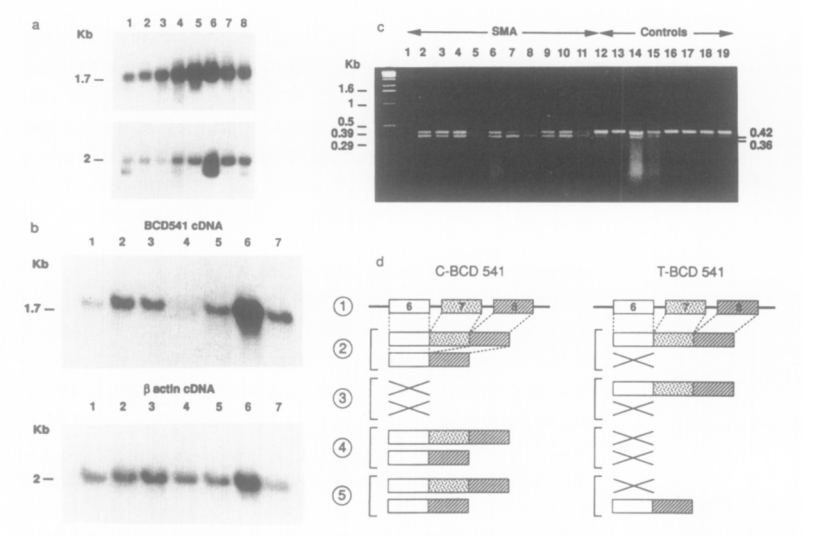

Figure 4. RNA Analysis of *BCD541* Gene Transcripts in Controls and SMA Individuals

(a) The *BCD541* cDNA, hybridized to a Northern blot containing 2 μg of poly(A)$^+$ RNA per lane from heart (lane 1), brain (lane 2), placenta (lane 3), lung (lane 4), liver (lane 5), skeletal muscle (lane 6), kidney (lane 7), and pancreas (lane 8) (Clonetech) revealed the presence of a widely expressed 1.7 kb transcript in all tissues tested. Filters were consecutively hybridized with *BCD541* cDNA and an internal control probe β-actin.

(b) Northern blot analysis of *SMN* in SMA patients and controls. Northern blot analysis of poly(A)$^+$ RNA isolated from lymphoblastoid cell lines of controls (lanes 6 and 7) and SMA patients (lanes 1–5) hybridized with the *BCD541* cDNA. SMA patients belonged to type I (lanes 1–4) or type III (lane 5). Filters were consecutively hybridized with *BCD541* cDNA and an internal control probe β-actin. Gene dosage was determined by densitometric scanning of the hybridization signals. A significant decrease in *BCD541* cDNA band intensity was observed in 2 of 4 type I SMA patients (lanes 1 and 4) compared with the controls (lanes 6 and 7).

(c) RT–PCR analysis of *BCD541* transcripts in controls and SMA patients. RT–PCR amplification using oligonucleotides located in exons 6 and 8 (*541C618* and *541C1040*) from lymphoblastoid cell lines in SMA patients (lanes 1–11) and controls was performed (lanes 12–19). All patients presented here, except one (patient SA, carrying both the *SMN* and centromeric genes; lane 7), were lacking telomeric exons 7 and 8. All controls but two (carrying only the *SMN* gene; lanes 13 and 16) carried both the *SMN* and the centromeric genes. Electrophoresis on 2% Seaplaque agarose of RT–PCR amplification products revealed two bands corresponding to differential splicing of exon 7 in the *BCD541* transcripts. Controls carrying only the *SMN* gene expressed only the 0.42 kb RT–PCR products containing exon 7 (lanes 13 and 16). Controls carrying both the centromeric and the *SMN* genes expressed both the 0.42 kb RT–PCR products and the 0.36 kb RT–PCR products lacking exon 7 (lanes 12, 14–15, and 17–19). RT–PCR analysis showed that in controls the 0.36 kb RT–PCR products were far less abundant than the 0.42 kb RT–PCR products. In SMA patients lacking the *SMN* gene on both mutant chromosomes, both species were observed, and the amount of the 0.36 kb RT–PCR products was comparable or even more abundant than that of the 0.42 kb RT–PCR products (lanes 1–11). In patient SA (lane 7), carrying both the centromeric gene(s) and the *SMN* gene with a 7 bp deletion in the 3′ splice acceptor site of intron 6, the amount of the 0.36 kb RT–PCR products was more abundant than that of the 0.42 kb RT–PCR products, demonstrating the deleterious effect of the deletion on the processing of the mutant *SMN* gene.

(d) Schematic representation of the differential splicing of exon 7 of the *BCD541* transcripts in controls and SMA patients. Line 1, genomic organization of the centromeric (^{C}*BCD541*) and the *SMN* genes (^{T}*BCD541*); line 2, splicing patterns of *BCD541* transcripts in controls carrying both genes (^{C}*BCD541* and ^{T}*BCD541*); line 3, splicing patterns of transcripts of the ^{T}*BCD541* (*SMN*) gene in controls carrying the *SMN* gene only; line 4, splicing patterns of the *BCD541* transcripts in SMA patients carrying only the ^{C}*BCD541* gene; line 5, abnormal splicing of the mutant *SMN* gene transcripts (^{T}*BCD541*) in patient SA carrying a deletion in the 3′ splice acceptor site of intron 6.

ern blot and PFGE analyses of YACs spanning the disease locus. Specific hybridization with nonoverlapping YACs containing only E^{Cen} (YAC clones 759A3, 121B8, and 278G7) or only E^{Tel} (YAC clone 595C11) provided evidence for the duplication of the *SMN* gene (Figure 1a). Each gene encompassed approximately 20 kb and displayed an identical restriction pattern (data not shown). Hybridization experiments with probes *BCD541* and p322 that closely flanked the SacII, SfiI, and EagI sites of each element gave evidence for head-to-head orientation of the two genes (Figure 1a).

To search for differences between the two genes, the organization of the *SMN* gene was characterized and compared with that of the centromeric counterpart (see Experimental Procedures). Genomic sequence analysis revealed that the *SMN* gene is composed of eight exons (Figure 3). Starting from either the centromeric or telomeric gene locus (in YAC clones 121B8 and 595C11, respectively), PCR amplification and sequence analysis of each exon and its flanking regions revealed five discrepancies between the centromeric and the *SMN* genes. These were a synonymous mutation in exon 7 (codon 280; TTC, telomeric; TTT, centromeric; Figure 3), a 3′ noncoding region change in exon 8 (nucleotide 1155; TGG, telomeric; TGA, centromeric; Figure 3), and three other single base substitutions in the sixth and seventh introns (data not shown). The presence of both versions of each exon (exon 7 and 8) on a YAC clone containing both gene loci (YAC clone 920C9) and in the monochromosomal hybrid HHW105 (Carlock et al., 1985) demonstrated that these substitutions are neither allelic nor due to polymorphisms (Figure 5). Band shifts on single strand conformation polymorphism (SSCP) analysis of exons 7 and 8 PCR-amplified products allowed the *SMN* (also termed ^{T}*BCD541*) and centromeric (^{C}*BCD541*) genes in both controls and SMA patients to be readily distinguished. All (100%) unrelated controls tested (n = 246) carried the *SMN* gene, as shown by SSCP analysis of exons 7 and 8. The majority (235 of 246; 95.5%) also carried the centromeric ^{C}*BCD541*, but 11 of 246 (4.4%) lacked ^{C}*BCD541* (Figure 5).

Absence of the *SMN* (^{T}*BCD541*) Gene in SMA Patients

SSCP analysis of amplified exons 7 and 8 was carried out in 229 SMA patients. Of these, 103 belonged to type I, 91 to type II, and 35 to type III. A total of 213 of 229 SMA patients (93%) lacked the *SMN* exons 7 and 8 on both mutant chromosomes as compared with 0 of 246 controls

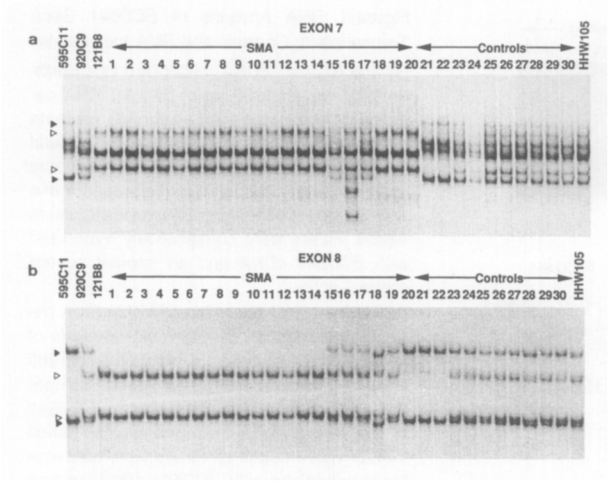

Figure 5. SSCP Analysis of Amplified Exons 7 and 8 from YAC Clones, the Monochromosomal Hybrid HHW105, Controls, and SMA Patients

Band shifts on SSCP analysis of PCR-amplified products of exons 7 (a) and 8 (b) allowed the *SMN* and centromeric genes to be distinguished. Bands indicated by closed arrowheads are present in YAC clone 595C11 containing only the E^{Tel}, indicating that these bands are specific to the telomeric exons 7 and 8. The presence of different bands in YAC clone 121B8 containing only the E^{Cen} (indicated by open arrowheads) showed that these fragments are specific to centromeric exons 7 and 8. Both conformations are present in YAC 920C9 and the monochromosomal hybrid HHW105, indicating that both genes are present and that they are not allelic. Control individuals are shown in lanes 21–30. The majority of these carried both telomeric and centromeric exons. Controls in lanes 21 and 22 carried only the telomeric exons 7 and 8. Lanes 1–20 contained amplified exons 7 and 8 from SMA patients. The majority of these carried only centromeric exons 7 and 8 (lanes 1–14). Patients in lanes 15 (HU), 16 (SA), and 17 (BI) carried both telomeric and centromeric exons 7 and 8. The abnormal band size seen in patient SA (lane 16) corresponds to the 7 bp deletion in the sixth intron of the 3′ splice acceptor site. Patients in lanes 18–20 lacked the telomeric exon 7 on both mutant chromosomes, but carried the telomeric exon 8, indicating an interruption of the *SMN* gene between exons 7 and 8 on both mutant chromosomes. Lane 20 shows the SSCP analysis in the consanguineous type III patient in whom the exon 7–8 PCR product contained a centromeric exon 7 but a telomeric exon 8.

(0%; Figure 5). Interestingly, 13 of 229 SMA patients (5.6%) lacked the *SMN* exon 7 on both mutant chromosomes but retained the *SMN* exon 8, as compared with 0 of 246 controls (0%; Figure 5). Further, in one of these, a consanguineous type III patient, amplification of exon 7 to exon 8 resulted in a product in which sequence analysis showed that exon 7 was centromeric but that exon 8 was derived from the telomeric gene (Figure 5). As both exonic and intronic sequences were completely normal and no abnormal restriction fragment was present on Southern blot analysis (data not shown), this suggested that a gene conversion event may have occurred. Finally, only 3 of 229 (1.3%) SMA patients carried the *SMN* exons 7 and 8 as compared with all the controls (100%; Figure 5). SSCP analysis of amplified exon 7 was carried out in 127 parents of SMA patients. All parents tested carried at least one copy of the *SMN* gene (100%; data not shown but available on request).

These results showed that the *SMN* gene is either absent or truncated in 98.6% of SMA patients. In addition, the observation of patients lacking the *SMN* exon 7 but not the *SMN* exon 8 (13 of 229) supported the view that the gene defect is centromeric to *SMN* exon 8 (Figure 1b). Thus, the critical region is included in a 20 kb region between the *C272* telomeric sublocus and *SMN* exon 8, partly containing the *SMN* gene, which can be regarded, therefore, as a determining gene for SMA (Figure 1b).

Intragenic Mutations of the *SMN* Gene

To demonstrate that the *SMN* gene (*^{T}BCD541*) is indeed a determining gene for SMA, we searched for mutations in the three SMA patients whose *SMN* gene showed no rearrangements. PCR amplification of exon 7 and its flanking regions revealed, in addition to the normal-sized PCR product, a smaller fragment in patient SA (Figure 5a). Direct sequence analysis of the smaller fragment showed a 7 bp deletion in the 3′ splice acceptor site of *SMN* intron 6 (Figure 6). Sequence analysis of the normal-sized PCR product detected the sequence specific to the centromeric exon 7 but not that of the *SMN* exon 7 (data not shown). These results demonstrated that patient SA carried an intronic deletion of the *SMN* gene on one mutant allele and lacked the *SMN* exon 7 on the other mutant allele. This intronic deletion was inherited from the mother, and, in contrast with that of her affected child, the allelic fragment was specific to the *SMN* gene, confirming that the unaffected mother is, indeed, heterozygous for the mutation (data not shown). In patient BI, a 4 bp deletion was found in the 5′ consensus splice donor site of *SMN* intron 7 (Figure 6). Sequence analysis of the nondeleted amplification product recognized the centromeric exon 7, but not the *SMN* exon 7, suggesting that the other mutant chromosome lacked the *SMN* exon 7 (data not shown). These two mutations were absent in 246 control individuals. Such mutations have been found in other genes and typically abolish or disrupt splicing (Bottema et al., 1990; Murru et al., 1991). Finally, patient HU was heterozygous for a point mutation at codon 272 (TAT→TGT), changing a tyrosine into a cysteine in the protein (Figure 6). This mutation was absent in 100 normal chromosomes, ruling out rare polymorphisms.

RNA Analysis Revealed Differential Splicing of Exon 7 in the Centromeric (*^{C}BCD541*) and *SMN* Genes

Eight unrelated controls and 10 of 11 SMA patients lacking the *SMN* gene on both mutant chromosomes but carrying the *^{C}BCD541* gene (type I SMA, 7 patients; type II SMA, 1 patient; type III SMA, 3 patients) were selected for Northern blot analysis of their lymphoblastoid cell line poly(A)$^{+}$ RNAs using a full-length *BCD541* cDNA probe and a β-actin cDNA probe as the internal control. Using the *BCD541* cDNA, we detected a single 1.7 kb mRNA in the controls and the patients. Interestingly, however, the relative amount of the specific mRNA was markedly reduced in 4 of 7 SMA type I patients as compared with controls (Figure 4b). As all 10 patients lacked the *SMN* gene on both mutant chromosomes but carried at least one copy of the *^{C}BCD541* gene, this result suggested that the *^{C}BCD541* gene is transcribed.

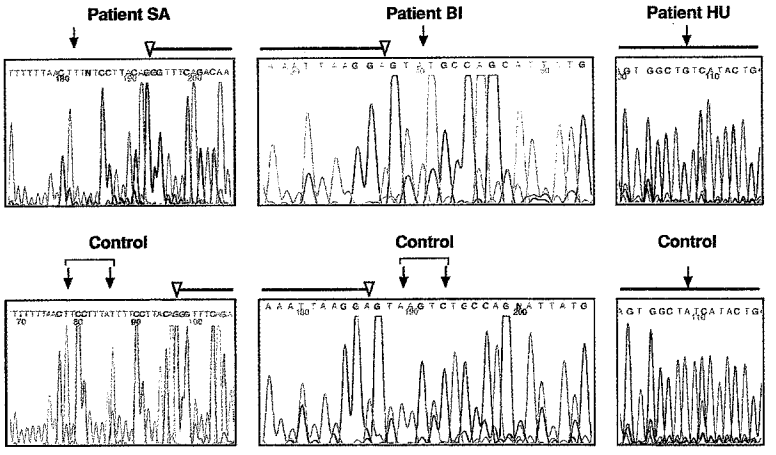

Figure 6. Intragenic Mutations in Patients SA, BI, and HU Carrying the Telomeric Gene

Mutant sequences of patients SA, BI, and HU (above) and control sequences (below) are shown. Arrows indicate the position of deletions (patient SA and patient BI) or the point mutation (patient HU). Triangles indicate the intron–exon junctions and solid lines the sequences of exons 7 (patient SA and BI) and 6 (patient HU).

Amplification and sequence analysis of exons 6–8 from reverse transcribed RNAs confirmed that the *SMN* transcripts were absent and that the ^{C}BCD541 transcripts were solely present in patients lacking the *SMN* gene on both mutant chromosomes, while control individuals expressed both RNA transcripts (^{C}BCD541 and *SMN*; Figure 4d). These studies also revealed that the centromeric transcripts, but not the *SMN* gene transcripts, may normally undergo alternative splicing of exon 7 to produce transcripts lacking this exon. This was shown by the observation that controls lacking the ^{C}BCD541 gene on both chromosomes expressed only the full-length RNA while controls carrying both the centromeric and the *SMN* genes expressed full-length and truncated RNA (Figures 4c and 4d). These results suggested that the truncated RNA transcripts are specific for the centromeric gene (Figure 4d). In patient SA, who carried a 7 bp deletion in the 3′ splice acceptor site of *SMN* intron 6, *SMN* exon 7 was shown to be absent in *SMN* transcript (Figures 4c and 4d). These results showed the deleterious effect of this intronic deletion on *SMN* transcript splicing.

Finally, reverse transcription–PCR (RT–PCR) analysis of control RNAs showed that the truncated ^{C}*BCD541* transcripts were far less abundant than full-length RNA transcripts (Figure 4c). By contrast, the amount of truncated ^{C}*BCD541* transcripts was comparable or even more abundant than that of the full-length RNA in SMA patients (Figure 4c). These results suggested that the amount of full-length and truncated RNA differed between controls and SMA patients.

Discussion

We and other groups (Brzustowicz et al., 1990; Melki et al., 1990a, 1990b, 1993; Gilliam et al., 1990; Sheth et al., 1991; Lien et al., 1991; Morrisson et al., 1992; Soares et al., 1993; Clermont et al., 1994) have previously localized the SMA locus to chromosome 5q13, a region characterized by the presence of low copy repeat elements (Kleyn et al., 1993; Francis et al., 1993; Melki et al., 1994). The construction of a YAC contig encompassing the SMA locus enabled the detailed physical mapping of this region and led to the identification of a large inverted duplication of a 500 kb element within this region (E^{Cen} and E^{Tel}). Physical mapping studies by PFGE on genomic DNA from normal individuals revealed a high degree of variability of this region that could be explained by the variable number of repeated sequences in the general population. It was, however, possible using Southern blot analysis to distinguish E^{Tel} from E^{Cen} using probe *He3*, which detected distinct restriction fragments corresponding to specific loci on the E^{Cen} and E^{Tel}.

In this report, we present the identification and characterization of a gene, which we have called the *SMN* gene, that is duplicated within the 5q13 region and that we believe is a determining gene for SMA. The genomic organization of both the telomeric (*SMN* or ^{T}*BCD541*) and the centromeric (^{C}*BCD541*) genes was identical, suggesting a recent duplication event. Five discrepancies enabled the *SMN* and the centromeric (^{C}*BCD541*) genes to be distinguished. Two of these, in exons 7 and 8, were readily distinguished by SSCP analysis and were used to show the presence or absence of these genes. Analysis of a control population showed that the *SMN* gene was present in all individuals while the ^{C}*BCD541* gene was present in 95.5%. This indicated that the telomeric one was the ancestral gene (*SMN*).

Northern blot analysis of RNA from a range of tissues, including spinal cord, showed that this gene was widely expressed and resulted in a 1.7 kb mRNA transcript. The ORF encoded a putative protein of 294 amino acids with a predicted molecular mass of 32 kDa. No homology was found at the nucleotide or amino acid levels with any previously identified protein, suggesting that this gene encodes a novel protein of unknown function. The presence of interspecies-conserved sequences within the *SMN* gene will enable the isolation of homologous sequences and will enable the identification of functionally conserved domains within the *SMN* protein. Sequence analysis of reverse-transcribed RNAs from controls showed that both genes are expressed, although it is unclear whether both are translated into functional proteins. In addition, the centromeric gene (^{C}*BCD541*), specifically, was shown to undergo alternative splicing of exon 7, resulting in a trun-

cated transcript lacking this exon and a putative protein with a different C-terminal end. This truncated transcript was less abundant than the full-length transcript in normal individuals carrying both centromeric and telomeric genes. These results enabled the *SMN* and the *^{c}BCD541* genes to be distinguished at the RNA level.

We have recently reported the presence of large-scale deletions encompassing the disease locus on one mutant chromosome in nine unrelated SMA patients (Melki et al., 1994). These deletions allowed us to assign the SMA locus to a 1.2 Mb critical region flanked by the microsatellite marker *C161*. Moreover, deletions were strongly suggested in at least 18% of SMA type I patients by the observation of a reduced number of *C212* and *C272* subloci. These results indicated that deletion events are statistically associated with the severe form of SMA (type I). Recently, DiDonato et al. (1994) also demonstrated a highly significant association between *Ag1-CA* alleles and type I SMA, confirming our previous results, as *Ag1-CA* primers are a subset of the *C272* sequence. As a result of the high degree of variability of this region in the general population, we were unable to recognize abnormal restriction fragments or to narrow the SMA critical region further by PFGE analysis in SMA patients. Nevertheless, the combined genetic and physical analysis of consanguineous patients carrying identical mutations on both mutant chromosomes and of patients with large-scale deletions on one mutant chromosome placed the SMA gene between the telomeric subloci detected by markers *C272* and *He3*, in a 140 kb fragment of E^{Tel}.

Among the candidate genes identified within the 5q13 region, one, the *SMN* gene, was located between markers *C272* and *He3*. Specific sequence variations between the centromeric (*^{c}BCD541*) and *SMN* genes showed that in 93% of patients (213 of 229), the *SMN* gene was absent on both mutant chromosomes. Moreover, the interruption of the *SMN* gene between exons 7 and 8 in 13 of 229 SMA patients (5.6%) reduced the critical region to a 20 kb fragment, between the telomeric loci detected by marker *C272* and *SMN* exon 8. Mutation analysis in the three patients retaining the *SMN* gene identified deleterious mutations on one mutant chromosome in all three cases and showed that the *SMN* gene was absent on the other mutant chromosome in two of these. The fact that either intragenic mutations or lack of the *SMN* gene resulted in an SMA phenotype, while the absence of the *^{c}BCD541* gene had no apparent phenotypic effect, strongly suggested a critical function for *SMN* as compared with *^{c}BCD541*.

Northern blot analysis in patients lacking *SMN* genes on both mutant chromosomes but carrying at least one copy of the *^{c}BCD541* gene showed a marked decrease of RNA in only 4 of 7 type I patients. On the other hand, RT–PCR analysis of lymphoblastoid cell lines from patients lacking both *SMN* genes showed the presence of *^{c}BCD541* gene transcripts only. In addition, the proportion of the truncated transcripts specific to the *^{c}BCD541* gene, relative to the full-length species, was increased in SMA patients, whereas in controls the full-length RNA represented the major species. As the truncated transcripts may result in a putative protein with a different C-terminal end, the relevance of these observations to the disease phenotype requires further investigation.

Taken together, our data suggest that the *SMN* gene is an SMA-determining gene. Unfortunately, lack of polymorphism at the *SMN* locus hampered our ability to estimate the actual frequency of de novo rearrangements, previously reported in SMA (Melki et al., 1994). On the other hand, no phenotype–genotype correlation between the gene defect and the type of SMA was observed, as the *SMN* gene was absent or truncated in 98.6% of the patients independent of the type of SMA. We have previously shown that deletion events are, statistically, associated with the severe form (type I) of SMA (Melki et al., 1994). Consanguineous patient analysis, using markers flanking the *SMN* gene (*C272* and *He3*), showed that deletions are more frequent and larger in type I than in type III patients. Indeed, 11 of 12 type I consanguineous patients had deletion involving the telomeric loci detected by markers *C272* and *He3*, while 0 of 9 type III consanguineous patients had deletion involving both loci. One could hypothesize that regulatory elements or gene(s) mapping close to the *SMN* gene, as the one reported by Roy et al. (1995 [this issue of *Cell*]), may be involved in the large genomic rearrangements present in type I SMA, thus altering the clinical phenotype of the patients. In keeping with this, we were able by PFGE analysis and PCR amplification to localize the neuronal apoptosis inhibitory protein (*NAIP*) gene and its truncated forms with respect to the *SMN* gene using primers 1926–1933 and 1258–1343 for the 5′ and the 3′ ends of the *NAIP* gene, respectively (Roy et al., 1995). Both PCR amplification and PFGE analysis on our YAC and phage contigs provided evidence that the *NAIP* gene was at least duplicated (Figure 1a). Moreover, the telomeric version of the *NAIP* gene was distal to the 3′ end of the *SMN* gene with the following order: *C272*—5′-*SMN*-3′—3′-*NAIP*-5′—Tel (data not shown but available on request). Although the telomeric *NAIP* gene was contained in the SMA critical region defined by the telomeric loci detected by markers *C272* and *He3*, the *NAIP* gene was distal to the *SMN* exon 8 (Figure 1a). Consequently, the *NAIP* gene mapped outside the nested critical SMA region defined by *C272* and *SMN* exon 8. As this gene is very close to the *SMN* gene, it might be involved in the large genomic deletions present in type I but not in type III SMA patients. Further investigations are required to confirm the involvement of both a determining gene and a modifying gene in SMA.

On the other hand, the absence of the *SMN* gene without gene dosage effect in type III may be accounted for by gene conversion events changing the *SMN* into the *^{c}BCD541* gene, thus resulting in an increased number of *^{c}BCD541* copies. This hypothesis is supported by the observation of the type III patient in whom a single exon 7–8 amplification product contained a centromeric exon 7 linked to a telomerically derived exon 8. A similar mechanism had been previously suggested in steroid 21-hydroxylase deficiency (Higashi et al., 1988). Under this hypothesis, the number of copies of the *SMN* gene would be reduced in all types of SMA. Thus, consanguineous SMA patients harboring identical mutant haplotypes would be

expected to carry two *cBCD541* genes in type I (one on each chromosome) and four *cBCD541* genes in type III (two on each chromosome). Interestingly, the absence of both *SMN* and *cBCD541* genes was not observed in our series. Such a genotype would be expected to result in an extremely severe form of SMA or in a nonviable fetus. This situation may be similar to that previously reported in α-thalassemia (Orkin, 1987). Assuming that the *cBCD541* gene is translated into, at least, a partially functional protein, the different phenotypes of SMAs would depend on the number of *cBCD541* genes on each chromosome.

Our ability to detect 98.6% of SMA patients by analysis of exon 7 of the *SMN* gene will revolutionize clinical and preclinical testing for this disorder and will greatly facilitate genetic counseling by eliminating the need for complex linkage analysis. Further, it will extend the option of prenatal diagnosis to families in which affected individuals are not available to establish genetic phase.

Although the exact nature of the genetic mechanisms resulting in SMAs remains to be clarified, this study provides important clues to the unraveling of this complex puzzle and forms a base from which to explore the molecular biology, biochemistry, and cell biology of this devastating disease. Furthermore, this work will contribute to the fundamental understanding of the survival of motor neurons.

Experimental Procedures

YAC Library Screening

YAC libraries from CEPH were screened by PCR with microsatellite markers *C212*, *C272*, *C171*, *CMS1*, and *C161* according to the three-dimensional procedure (Green and Olson, 1990). YAC genotypes were established by electrophoresis of amplification products on denaturing polyacrylamide gels (Hazan et al., 1992). YAC size was estimated by PFGE.

Construction of Phage Libraries from YAC Clones 121B8, 595C11, and 903D1

Total yeast DNA from YAC clone 595C11, containing the telomeric loci detected by *C212*, *C272*, and *C161*; from YAC clone 121B8, containing the centromeric loci detected by the same markers; or from YAC clone 903D1, containing both loci, was purified and partially digested with restriction enzyme Sau3A. DNA in the 12–23 kb size range was excised after 0.5% Seaplaque GTG agarose gel electrophoresis and precipitated with ethanol after β-agarase digestion. After partial fill-in of the Sau3A site, DNA was subcloned at the partially filled XhoI site of bacteriophage FIXII (Stratagene). Phage clones containing the microsatellite DNA markers *C212* (L-51), *C272* (L-51 and L-132), *C171* (L-5 and L-13), *C161* (595B1), *CMS1* (L-11), and *AFM157xd10* (L-131) were digested with the restriction enzyme EcoRI or HindIII or with both and subcloned into the Puc18 plasmid vector. Subclones from phages containing markers *C212* (p322), *C272* (132SE11), *C161* (He3), *AFM157xd10* (131xb4), and *CMS1* (p11M1) were used as probes.

PFGE Analysis

High molecular weight DNA was isolated in agarose plugs from Epstein–Barr virus–transformed lymphoblastoid cell lines established from controls and patients or from YAC clones as described previously (Herrmann et al., 1987). Plugs were rinsed twice for 30 min each in 10–20 vol of TE buffer. The plugs were equilibrated for 30 min at 4°C with 0.3 ml of the appropriate restriction enzyme buffer containing 0.1 mg/ml BSA (Pharmacia). Excess buffer was removed, and the plugs were incubated at the appropriate temperature for 16 hr with 40 U of restriction enzyme per reaction. DNA was digested with the restriction enzymes EagI, SfiI, and SacII. Double digestions of DNA from YAC clones with SfiI–EagI, SfiI–SacII, and EagI–SacII were also performed. Separation of DNA fragments was achieved using a CHEF-III-DR PFGE apparatus (Bio-Rad). Fragments of 50–1200 kb were separated by electrophoresis through 1% Seakem agarose at 200 V for 24 hr at 14°C in 0.5× TBE running buffer using a 30–70 s ramping pulse time. The separation of fragments of 5–100 kb was performed by electrophoresis at 200 V for 19 hr at 14°C in 0.5× TBE buffer using a 5–20 s ramping pulse time. After treatment with 0.25 N HCl for 20 min, pulsed field gels were blotted onto Hybond N^+ nylon membranes (Amersham) in 0.4 N NaOH, 0.4 M NaCl for 20 hr. Probes were successively hybridized to the same filters to ensure accurate data.

Southern Blot and Dinucleotide Repeat Polymorphism Analysis

DNA was extracted from YAC DNA clones or peripheral blood leukocytes or lymphoblastoid cell lines. DNA was digested with the restriction enzyme SacI, KpnI, EcoRI, HindIII, BglII, XbaI, PvuII, XmnI, RsaI, PstI, or BamHI, separated by electrophoresis on a 0.8% agarose gel for Southern blotting, and hybridized with radioactively labeled probes (Sambrook et al., 1989). For genotyping using marker *C272* (D5F150S1 and D5F150S2), amplification conditions were as follows: denaturation at 94°C, annealing at 55°C, and extension at 72°C for 1 min each for 30 cycles (Melki et al., 1994).

cDNA Screening and Characterization

Recombinant clones (2×10^6) of a λgt10 human fetal brain cDNA library were plated according to the instructions of the manufacturer (Clonetech). Prehybridization and hybridization with labeled probes were carried out in a solution containing 5× SSEP, 1% SDS, and 5× Denhardt's solution with 100 μg/ml denaturated sheared human placental DNA (Sigma) for 16 hr at 65°C. The filters were washed in 0.1× SSEP, 0.1% SDS at 65°C and autoradiographed for 24 hr. The DNA of positive clones was purified, digested with EcoRI, and subcloned in M13mp18 bacteriophage. Single-stranded DNAs were sequenced using the DyeDeoxy terminator cycle sequencing kit protocol supplied by Applied Biosystems and were analyzed on an ABI model 373A DNA automated sequencer. To obtain the 3′ end of the cDNA along with its poly(A)$^+$ tail, we performed PCR amplification of a lymphoblastoid cell line cDNA library using a primer complementary to the 3′ end of the *BCD541* cDNA and a primer specific to the vector arms of the cDNA library as previously described (Fournier et al., 1994). Specific PCR products were sequenced directly with both primers. Homology searches using the FASTA and BLAST systems through the CITI2 French network (Dessen et al., 1990) were performed at either the nucleotide or the amino acid level.

Genomic Sequence Analyses of the *cBCD541* and *SMN* Genes

DNA from phage clones containing the *SMN* gene (L-132, L-5, and L-13) was digested with either HindIII or EcoRI or with both and subcloned into the Puc18 plasmid vector. DNA sequencing reactions were performed on double-stranded plasmid DNA with oligonucleotides derived from the cDNA sequence. Genomic DNA sequences were compared with the cDNA sequence, and oligonucleotides derived from exon flanking regions were designed to amplify and directly to sequence each exon of the *cBCD541* and *SMN* genes.

Control and SMA patient DNA was sequenced directly using the amplification products of each exon. PCR products were purified by polyacrylamide gel electrophoresis and subjected to a second round of PCR amplification. When the size of the amplification products differed, each band was separated by polyacrylamide gel electrophoresis before reamplification and sequencing using both primers. In cases in which sequence analysis revealed heterozygosity for mutations, PCR amplification products were cloned into the pBluescript SK(+) vector (Stratagene) to confirm nucleotide substitutions.

SSCP Analysis

For SSCP analysis, DNA from peripheral leukocytes (200 ng) was amplified by PCR using unlabeled primers (20 μM) in 25 μl of amplification mixture containing 200 μM dNTPs, 1 U of Taq polymerase (GIBCO BRL), and 0.1 μl of [α-^{33}P]dCTP (10 mCi/ml; New England Nuclear). Amplified DNA was mixed with an equal volume of formamide loading dye (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). The samples (5 μl) were denatured for 10 min at 95°C and loaded onto a polyacrylamide gel (Hydrolink MED, Bioprobe) and electrophoresed at 4°C for 18–24 hr at 4 W. Gels were transferred

onto 3MM Whatman paper, dried, and autoradiographed using Kodak X-OMAT films for 24–48 hr. To amplify genomic DNA containing divergent exon 7 sequences, we used oligonucleotides *R111* (5′-AGACTATCAACTTAATTTCTGATCA-3′) and *541C770* (5′-TAAGGAATGTGAGCACCTTCCTTC-3′). To amplify genomic DNA containing divergent exon 8 sequences, we used oligonucleotides *541C960* (5′-GTAATAACCAAATGCAATGTGAA-3′) and *541C1120* (5′-CTACAACACCCTTCTCACAG-3′).

RNA Procedures

Poly(A)$^+$ RNA from lymphoblastoid cell lines of controls and SMA patients was isolated using the QuickPrep mRNA purification kit (Pharmacia) and electrophoresed on a 1.5% Seakem agarose gel containing 10 mM methylmercury hydroxide in a 5 mM tetraborate buffer, 10 mM sodium sulfate, 1 mM EDTA, and 50 mM boric acid (pH 8.2), transferred to a positively charged membrane in 20× SSC, and baked for 2 hr at 80°C. After hybridization, the membranes were washed to a final stringency of 0.1× SSEP, 0.1% SDS at 65°C for 10 min. Autoradiography was at −80°C with intensifying screens and using Kodak XAR films for 2–10 days. The β-actin cDNA was used as an internal control probe, and the autoradiographs were scanned at 600 nm using a computerized densitometer (Hoeffer Scientific Instruments). A Northern blot of poly(A)$^+$ RNA from several human tissues was purchased from Clonetech.

First-strand synthesis of cDNA from poly(A)$^+$ RNA was carried out using random hexanucleotide primers and AMV reverse transcriptase (Promega). The single-stranded cDNAs were consequently amplified by PCR using 2 pmol of forward (*541C618*, 5′-CTCCCATATGTCCAGATTCTCTTG-3′) and reverse (*541C1040*, 5′-ACTGCCTCACCACCGTGCTGG-3′) primers in a final volume of 20 μl containing 1 U of Taq polymerase (Cetus). Amplification conditions were as follows: denaturation at 94°C, annealing at 55°C, and extension at 72°C for 1 min each for 30 cycles. The PCR products were separated, excised after polyacrylamide gel electrophoresis, and eluted in 100 μl of TE buffer. The diluted fragments were reamplified with the same primers prior to direct sequencing with both primers using the DyeDeoxy terminator cycle sequencing kit protocol (Applied Biosystems).

Acknowledgments

We thank the patients, families, and doctors who have contributed to this project and without whom this study would not have been possible. We thank V. Raclin, N. Gigarel, M. Berthelon (INSERM Unité 393), and J. F. Prudhomme (Genethon) for technical help. We are grateful to C. Petit and S. Strautnieks for advice and useful comments. We thank the members of the International SMA Consortium for stimulating discussions. This work was supported by the Association Française contre les Myopathies (AFM), the Groupement de Recherches et d'Etudes sur les Génomes, the Fondation de France, the Fondation Paribas, the Société d'Etudes et de Soins pour les Enfants Paralysés et Polymalformés, the Fonds d'Etudes et de Recherche du Corps Medical des Hôpitaux de Paris, and the Assistance Publique, Hôpitaux de Paris. L. B. and S. L. are recipients of AFM fellowships.

Received October 31, 1994; revised December 15, 1994.

References

Bottema, C. D. K., Ketterling, R. P., Yoon, H. S., and Sommer, S. (1990). The pattern of factor IX germ-line mutation in Asians is similar to that of Caucasians. Am. J. Hum. Genet. *47*, 835–841.

Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmsen, K. C., Daniels, R. J., Davies, K. E., Leppert, M., Ziter, F., Wood, D., Dubowitz, V., Zerres, K., Hausmanova-Petrusewics, I., Ott, J., Munsat, T. L., and Gilliam, T. C. (1990). Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-q13.3. Nature *344*, 540–541.

Burghes, A. H. M., Ingraham, S. E., McLean, M., Thompson, T. G., McPherson, J. D., Kote-Jarai, Z., Carpten, J. D., DiDonato, C. J., Ikeda, J.-E., Surh, L., Wirth, B., Sargent, C. A., Ferguson-Smith, M. A., Fuerst, P., Moysis, R. K., Grady, D. L., Zerres, K., Korneluk, R., MacKenzie, A., and Wasmuth, J. J. (1994). A multicopy dinucleotide marker that maps close to the spinal muscular atrophy gene. Genomics *21*, 394–402.

Carlock, L. R., Skarecky, D., Dana, S. L., and, Wasmuth, J. J. (1985). Deletion mapping of human chromosome 5 using chromosome-specific DNA probes. Am. J. Hum. Genet. *37*, 839–852.

Clermont, O., Burlet, P., Bürglen, L., Lefebvre, S., Pascal, F., McPherson, J., Wasmuth, J. J., Cohen, D., Le Paslier, D., Weissenbach, J., Lathrop, M., Munnich, A., and Melki, J. (1994). Use of genetic and physical mapping to locate the spinal muscular atrophy locus between two new highly polymorphic DNA markers. Am. J. Hum. Genet. *54*, 687–694.

Czeizel, A., and Hamula, J. (1989). A Hungarian study on Werdnig–Hoffmann disease. J. Med. Genet. *26*, 761–763.

Dessen, P., Fondrat, C., Valencien, C., and Mugnier, C. (1990). BISANCE: a French service for access to biomolecular sequence databases. Comput. Appl. Biosci. *6*, 355–356.

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., Mendell, J. R., Burghes, A. H. M., and Simard, L. (1994). Association between Ag1-CA alleles and severity of autosomal recessive proximal spinal muscular atrophy. Am. J. Hum. Genet. *55*, 1218–1229.

Fournier, B., Saudubray, J. M., Benichou, B., Lyonnet, S., Munnich, A., Clevers, H., and Tien Poll-The, B. (1994). Large deletion of the peroxysomal Acyl-CoA oxydase gene in pseudoneonatal adrenoleukodystrophy. J. Clin. Invest. *94*, 526–531.

Francis, M. J., Morrisson, K. E., Campbell, L., Grewal, P. K., Christodoulou, Z., Daniels, R. J., Monaco, A. P., Frischauf, A. M., McPherson, J., Wasmuth, J. J., and Davies, K. E. (1993). A contig of non-chimaeric YACs containing the spinal muscular atrophy gene in 5q13. Hum. Mol. Genet. *2*, 1161–1167.

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. L., Ott, J., and Davies, K. E. (1990). Genetic homogeneity between acute and chronic forms of spinal muscular atrophy. Nature *345*, 823–825.

Green, E. D., and Olson, M. V. (1990). Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci. USA *87*, 1213–1217.

Hazan, J., Dubay, C., Pankowiak, M. P., Becuwe, N., and Weissenbach, J. (1992). A genetic linkage map of human chromosome 20 composed entirely of microsatellite markers. Genomics *12*, 183–189.

Herrmann, B. G., Barlow, D. P., and Lehrach, H. (1987). A large inverted duplication allows homologous recombination between chromosomes heterozygous for the proximal T complex inversion. Cell *48*, 813–825.

Higashi, Y., Tanae, A., Inoue, H., and Fujii-Kuriyama, Y. (1988). Evidence for frequent gene conversion in the steroid 21-hydroxylase P-450 (C21) gene: implications for steroid 21-hydroxylase deficiency. Am. J. Hum. Genet. *42*, 17–25.

Hoffmann, J. (1900). Über die hereditare progressive spinale Muskelatrophie im Kindesalter. München Med. Wschr. *47*, 1649–1651.

Kleyn, P. W., Wang, C. H., Lien, L. L., Vitale, E., Pan, J., Ross, B. M., Grunn, A., Palmer, D. A., Warburton, D., Brzustowicz, L. M., Kunkel, L. M., and Gilliam, T. C. (1993). Construction of a yeast artificial chromosome contig spanning the SMA disease gene region. Proc. Natl. Acad. Sci. USA *90*, 6801–6805.

Kugelberg, E., and Welander L. (1956). Heredo-familial juvenile muscular atrophy simulating muscular dystrophy. Arch. Neurol. Psych. *75*, 500–509.

Lien, L. L., Boyce, F. M., Kleyn, P., Brzustowicz, L. M., Menninger, J., Ward, D. C., Gilliam, T. C., and Kunkel, L. M. (1991). Mapping of human microtubule-associated protein 1B in proximity to the spinal muscular atrophy locus at 5q13. Proc. Natl. Acad. Sci. USA *88*, 7873–7876.

McLean, M. D., Roy, N., MacKenzie, A. E., Salih, M., Burghes, A. H. M., Simard, L., Korneluk, R. G., Ikeda, J.-E., and Surh, L. (1994). Two 5q13 simple tandem repeat loci are in linkage desequilibrium with type 1 spinal muscular atrophy. Hum. Mol. Genet. *3*, 1951–1956.

Melki, J., Abdelhak, S., Sheth, P., Bachelot, M. F., Burlet, P., Marcadet, A., Aicardi, J., Barois, A., Carriere, J. P., Fardeau, M., Fontan, D., Ponsot, G., Billette, T., Angelini, C., Barbosa, C., Ferriere, G., Lanzi, G., Ottolini, A., Babron, M. C., Cohen, D., Hanauer, A., Clerget-Darpoux, F., Lathrop, M., Munnich, A., and Frézal, J. (1990a). Gene for proximal spinal muscular atrophies maps to chromosome 5q. Nature *344*, 767–768.

Melki, J., Sheth, P., Abdelhak, S., Burlet, P., Bachelot, M. F., Lathrop, M., Frézal, J., Munnich, A., and the French Spinal Muscular Atrophy Investigators (1990b). Mapping of acute (type I) spinal muscular atrophy to chromosome 5q12-q14. Lancet *336*, 271–273.

Melki, J., Burlet, P., Clermont, O., Pascal, F., Paul, B., Abdelhak, S., Sherrington, R., Gurling, H., Nakamura, Y., Weissenbach, J., Lathrop, M., and Munnich, A. (1993). Refined linkage map of chromosome 5 in the region of the spinal muscular atrophy gene. Genomics *15*, 521–524.

Melki, J., Lefebvre, S., Bürglen, L., Burlet, P., Clermont, O., Millasseau, P., Reboullet, S., Bénichou, B., Zeviani, M., Le Paslier, D., Cohen, D., Weissenbach, J., and Munnich, A. (1994). *De novo* and inherited deletions of the 5q13 region in spinal muscular atrophies. Science *264*, 1474–1477.

Morrisson, K. E., Daniels, R. J., Suthers, G. K., Flynn, G. A., Francis, M. J., Buckle, V. J., and Davies, K. E. (1992). High-resolution genetic map around the spinal muscular atrophy (SMA) locus on chromosome 5. Am. J. Hum. Genet. *50*, 520–527.

Munsat, T. L. (1991). Workshop report: international SMA collaboration. Neuromusc. Disord. *1*, 81.

Murru, S., Loudianos, G., Deiana, M., Camaschella, C., Sciarratta, G. V., Agosti, S., Parodi, M. I., Cerruti, P., Cao, A., and Pirastu, M. (1991). Molecular characterization of β-thalassemia intermedia in patients of Italian descent and identification of three novel β-thalassemia mutations. Blood *77*, 1342–1347.

Orkin, S. (1987). Disorders of hemoglobin: synthesis: the thalassemias. In The Molecular Basis of Blood Diseases, G. Stamatoyannopoulos, A. W. Nienhuis, P. Leder, and P. W. Majerus, eds. (Philadelphia: W. B. Saunders Company), pp. 106–126.

Pearn, J. (1973). The gene frequency of acute Werdnig–Hoffmann disease (SMA type I): a total population survey in north-east England. J. Med. Genet. *10*, 260–265.

Pearn, J. (1978). Incidence, prevalence, and gene frequency studies of chronic childhood spinal muscular atrophy. J. Med. Genet. *15*, 409–413.

Roberts, D. F., Chavez, J., and Court, S. D. M. (1970). The genetic component in child mortality. Arch. Dis. Child. *45*, 33–38.

Roy, N., Mahadevan, M. S., McLean, M., Shutler, G., Yaraghi, Z., Farahani, R., Baird, S., Besner-Johnston, A., Lefebrve, C., Kang, X., Salih, M., Aubry, H., Tamai, K., Guan, X., Ioannou, P., Crawford, T. O., de Jong, P. J., Surh, L., Ikeda, J.-E., Korneluk, R. G., and MacKenzie, A. (1995). The gene for NAIP, a novel protein with homology to baculoviral inhibitors of apoptosis, is partially deleted in individuals with spinal muscular atrophy. Cell *80*, this issue.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Sheth, P., Abdelhak, S., Bachelot, M. F., Burlet, P., Masset, P., Hillaire, D., Clerget-Darpoux, F., Frézal, J., Lathrop, M., Munnich, A., and Melki, J. (1991). Linkage analysis in spinal muscular atrophy, by six closely flanking markers on chromosome 5. Am. J. Hum. Genet. *48*, 764–768.

Soares, V. M., Brzustowicz, L. M., Kleyn, P. W., Knowles, J. A., Palmer, D. A., Asokan, S., Penchaszedeh, G. K., Munsat, T. L., and Gilliam, T. C. (1993). Refinement of the spinal muscular atrophy locus to the interval between D5S435 and MAP-1B. Genomics *15*, 365–371.

Werdnig, G. (1894). Die frühinfantile progressive spinale Amyotrophie. Arch. Psychiat. *26*, 706–744.

GenBank Accession Number

The accession number for the sequence reported in this paper is U18423.