*Am. J. Hum. Genet.* 61:40–50, 1997

# Genomic Variation and Gene Conversion in Spinal Muscular Atrophy: Implications for Disease Process and Clinical Phenotype

Louise Campbell,[1] Allyson Potter,[1] Jaakko Ignatius,[2] Victor Dubowitz,[3] and Kay Davies[1]

[1]Genetics Unit, Department of Biochemistry, University of Oxford, Oxford; [2]Family Federation of Finland, Department of Medical Genetics, Helsinki; and [3]Neuromuscular Unit, Department of Paediatrics, Royal Postgraduate Medical School, Hammersmith Hospital, London

## Summary

**Autosomal recessive spinal muscular atrophy (SMA) is classified, on the basis of age at onset and severity, into three types: type I, severe; type II, intermediate; and type III, mild. The critical region in 5q13 contains an inverted repeat harboring several genes, including the survival motor neuron (SMN) gene, the neuronal apoptosis inhibitory protein (NAIP) gene, and the p44 gene, which encodes a transcription-factor subunit. Deletion of NAIP and p44 is observed more often in severe SMA, but there is no evidence that these genes play a role in the pathology of the disease. In >90% of all SMA patients, exons 7 and 8 of the telomeric SMN gene (SMNtel) are not detectable, and this is also observed in some normal siblings and parents. Point mutations and gene conversions in SMNtel suggest that it plays a major role in the disease. To define a correlation between genotype and phenotype, we mapped deletions, using pulsed-field gel electrophoresis. Surprisingly, our data show that mutations in SMA types II and III, previously classed as deletions, are in fact due to gene-conversion events in which SMNtel is replaced by its centromeric counterpart, SMNcen. This results in a greater number of SMNcen copies in type II and type III patients compared with type I patients and enables a genotype/phenotype correlation to be made. We also demonstrate individual DNA-content variations of several hundred kilobases, even in a relatively isolated population from Finland. This explains why no consensus map of this region has been produced. This DNA variation may be due to a midisatellite repeat array, which would promote the observed high deletion and gene-conversion rate.**

Received February 26, 1997; accepted for publication April 22, 1997.

Address for correspondence and reprints: Dr. Kay E. Davies, Genetics Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom. E-mail: kdavies@biochem.ox.ac.uk


0002-9297/97/6101-0009$02.00

## Introduction

Spinal muscular atrophy (SMA) is an autosomal recessive disorder with variable clinical severity and an overall incidence of 1/6,000–1/10,000 (Dubowitz 1995). The disease primarily affects the anterior horn cells of the spinal cord, degeneration of which results in proximal muscle weakness. SMA is conventionally classified, on the basis of age at onset and severity, into three types (Munsat and Davies 1992; Dubowitz 1995). Werdnig-Hoffman disease, or type I SMA, is the most severe form, with onset either in utero or during the first few months of life. Affected children cannot sit unsupported, and death usually occurs at <2 years of age. Type II SMA usually manifests within the 1st year of life, and, although affected children may sit unaided, they do not achieve the ability to stand or walk independently. The survival of type II individuals depends on the degree of respiratory complications. Kugelberg-Welander disease, or type III SMA, is a less severe form, characterized by a later age at onset. Affected individuals have variable severity, walk independently, and may have a normal life expectancy.

All three forms of childhood SMA map to the same region on chromosome 5q13 (Brzustowicz et al. 1990; Gilliam et al. 1990; Melki et al. 1990*a*, 1990*b*). A number of YAC physical maps spanning the SMA region have been published, but it has not been possible to generate a consensus map of the region (Francis et al. 1993; Kleyn et al. 1993; Carpten et al. 1994; Melki et al. 1994; Roy et al. 1995*b*). This may be due to the fact that the critical region consists of many repeated sequences, pseudogenes, and retrotransposable elements, which may make the region unstable (Francis et al. 1993, 1995; Sargent et al. 1994; Theodosiou et al. 1994; Selig et al. 1995). It also harbors a large inverted duplication containing the survival motor neuron gene (SMN) (Lefebvre et al. 1995) and the neuronal apoptosis inhibitory-protein gene (NAIP) (Roy et al. 1995*a*) (see fig. 1*A*). The SMN gene encodes a novel protein with a proposed role in mRNA processing (Liu and Dreyfuss 1996). The telomeric SMN gene (SMNtel) differs from its centromeric copy (SMNcen) by five nucleotide changes, two of which occur in exons 7 and 8 and enable the two genes to be distinguished either by SSCP analysis

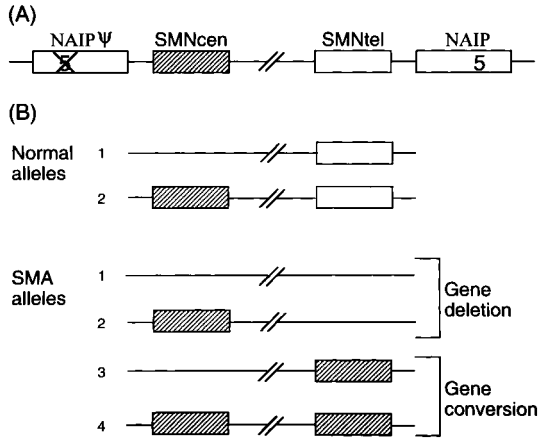


**Figure 1** *A,* Schematic representation of the duplicated region on chromosome 5q13, containing the SMN and NAIP genes. NAIPΨ = NAIP-pseudogene copy. The p44 gene is not represented, since it is not considered a candidate gene for SMA. *B,* Illustration of the types of normal and mutant SMN alleles proposed. Gray-shaded and unshaded boxes represent SMNcen and SMNtel gene copies, respectively. A gray-shaded box in the SMNtel position represents a gene-conversion event: SMNtel converted to SMNcen.

(Lefebvre et al. 1995) or by enzyme digestion of PCR products (van der Steege et al. 1995). Although both genes are transcribed, SMNtel is believed to be the important functional copy, since some exons of this gene are not detectable in SMA patients (Hahnen et al. 1995; Lefebvre et al. 1995; Rodrigues et al. 1995, 1996; Matthijs et al. 1996; Velasco et al. 1996). However, SMNcen exons 7 and 8 are also undetectable in a minority of normal individuals (Hahnen et al. 1995; Lefebvre et al. 1995; Rodrigues et al. 1995; Matthijs et al. 1996). The telomeric NAIP gene can be distinguished from its centromeric pseudogene counterpart by a PCR test for the presence or absence of exon 5, which only exists within the telomeric functional gene. Also contained within the inverted duplicated region is the gene encoding p44, a subunit of the basal transcription factor TFIIH (Humbert et al. 1994). Deletion or interruption of this gene is observed in 73% of type I SMA patients, but the observation of deletions in parents and controls suggests that this gene does not play a role in the disease pathology. In addition, the protein structure and function appear normal in patients homozygously deleted for the telomeric p44 gene (Bürglen et al. 1997).

Although lack of amplification of SMNtel exons 7 and 8 is observed by PCR-based methods in >90% of all SMA patients, there has as yet been no correlation between these apparent deletions and the severity of the disease, most likely because the extent of deletion cannot easily be established, because of the presence of SMNcen. Nonamplification of NAIP exon 5 is observed more frequently in type I SMA than in types II and III, suggesting the association of larger deletions with type I (Hahnen et al. 1995; Roy et al. 1995*a;* Wirth et al. 1995; Burlet et al. 1996; Rodrigues et al. 1996; Velasco et al. 1996).

It has been proposed that the SMA phenotype can be modified by the presence of differing numbers of copies of SMNcen. For example, studies have been carried out on the parents of SMA patients to determine the SMNcen:SMNtel dose ratio by densitometry of SSCP bands (Velasco et al. 1996). As obligate carriers, the parents were assumed to possess only one SMNtel gene. The SMNcen:SMNtel dose ratio therefore reflected the number of SMNcen copies. It was found that parents of type II and type III patients carried more copies of SMNcen than were carried by parents of type I patients. However, in control individuals the total number of SMNcen copies could not be assessed definitively, since a ratio of 1:1 could represent either two telomeric and two centromeric genes or one copy of each. The SMNcen:SMNtel copy ratio was also investigated by solid-phase minisequencing (Schwartz et al. 1997). The authors identified six predominant haplotypes, three for normal chromosomes and three for SMA chromosomes, and proposed a disease model involving compound heterozygosity of these alleles. Although the genotype containing three SMNcen copies was observed in SMA types II and III, the authors had insufficient evidence to correlate an increase in SMNcen copy number with a less severe phenotype. It has also been demonstrated that the number of alleles of the multicopy marker C272 (or Ag1CA) is an indicator of severity (DiDonato et al. 1994; Melki et al. 1994; Lefebvre et al. 1995; Wirth et al. 1995). Since the C272 marker lies at the 5′ end of each SMN gene, it should reflect the number of gene copies. Alleles of C272 are deleted more often in type I SMA than in types II and III, suggesting that fewer SMN copies remain in type I SMA. However, although SMNcen copy number is implicated in the variable severity observed in SMA, none of the methods employed to date enable a definitive assessment of SMN copy number to be made both in SMA patients and in control individuals, and no method can distinguish between the number of copies of SMNcen and the number of copies of SMNtel.

In ~5% of SMA patients, exon 7 of SMNtel is not detectable whereas exon 8 remains intact. (Cobben et al. 1995; Hahnen et al. 1995; Rodrigues et al. 1995, 1996). SMN genes in such individuals are the result of a gene-conversion event, resulting in chimeric genes consisting of SMNcen exon 7 linked to SMNtel exon 8 (Bussaglia et al. 1995; Lefebvre et al. 1995; Devriend et al. 1996; Hahnen et al. 1996; van der Steege et al. 1996; Velasco et al. 1996; DiDonato et al. 1997; Talbot et al., in press). The partial gene conversion, not involving the entire gene, that occurs in this minority of patients has been observed in all SMA severities. Compound heterozygosity of multi-

ple alleles has been proposed to account for the differences in clinical expression of the disease (Wirth et al. 1995; Talbot et al. 1996; DiDonato et al. 1997), and it may be that combinations of SMN gene deletions leading to severe alleles and gene conversions leading to mild alleles will result in variability of severity.

Nonamplification of exons of SMNtel has also been described in asymptomatic siblings and parents of affected individuals (Cobben et al. 1995; Hahnen et al. 1995; Wang et al. 1996). This observation has cast some doubt on the role of SMNtel as the causative gene for SMA. We propose that gene-conversion events may have led to the replacement of SMNtel by SMNcen in such individuals, thereby leading to the existence of four copies of SMNcen. Since SMNcen is thought to produce a small level of fully functional protein, four copies of this gene may produce enough functional protein to result in a normal phenotype (see Discussion).

In summary, the SMA region is highly complex, discrepancies exist in the YAC maps from different groups, and the exact copy number of the genes in the region cannot be reliably determined. Although the PCR-based deletion analysis of SMN exons is an extremely useful indicator of affected status, it cannot detect the extent of deletion. We therefore chose to investigate gene copy number and the extent of DNA deletion in SMA patients, using pulsed-field gel electrophoresis (PFGE). This also allowed us to test the hypothesis that the variability in the physical maps might be due to genomic variation in the normal population.

In this paper we demonstrate the direct assessment of SMN copy number by PFGE, enabling the identification of those bands representing telomeric and centromeric copies. We show the presence of a high degree of variation specific to the SMA region, in which fragment size can vary by as much as 400 kb between individuals. This variation is inherited in a Mendelian fashion and exists even in a highly inbred population. Its existence explains the lack of correlation between different physical maps of the region and suggests a high rate of change in this region of the genome. The data obtained allow us to make a hypothesis about the relationship between genotype and the severity of the SMA phenotype.

## Material and Methods

### *DNA Resources*

All patients conformed to internationally agreed diagnostic criteria (Munsat and Davies 1992). Lymphoblastoid cell lines (LCLs) were available from 15 unrelated SMA patients, of whom 5 were type I (1 from a consanguineous family), 7 were type II, and 3 were type III (1 consanguineous). Affected siblings and other family members were available from some families. LCL control samples were available from two non-SMA families. In order to study the genomic variation of the SMA region, blood was obtained from 7 unrelated Caucasian control individuals, as well as from 10 healthy blood donors originating from northern Finland. Informed consent was obtained from all subjects prior to sampling.

### *SMN and NAIP PCR Deletion Analysis*

SMN exon 7 and 8 deletion analysis was performed by the PCR and enzyme-digestion method described by van der Steege et al. (1995). PCR was performed by use of primers spanning SMN exons 7 and 8, and the products were digested with *Dra*I and *Dde*I, respectively. Digested products were run on 2.5% agarose gels, and patients were scored as to their deletion status. NAIP exon 5 deletion analysis was performed by the method described by Roy et al. (1995*a*). Duplex PCR was performed by use of primers specific to NAIP exons 5 and 13, and patients were scored as positive or negative for NAIP exon 5, exon 13 being an internal control.

### *PFGE*

Genomic DNA was prepared from lymphocytes embedded in 0.6% low-gelling-temperature agarose blocks at a concentration of $10^7$ cells/ml. After cell lysis and protein degradation in 0.5 M EDTA (pH 8.0) containing 1% SDS and 0.25 mg proteinase K,/ml at 50°C for 48 h, blocks were incubated at 37°C for 2 h in Tris-EDTA containing 2 mM Pefabloc (Boehringer Mannheim), to inactivate proteinase K. Blocks were digested overnight with 40 U restriction enzyme, and the DNA wasanalyzed on 1.2% pulsed-field agarose gels for 40 h (pulse time 70 s), by use of an LKB 2015 Pulsaphor system apparatus (Pharmacia). DNA was transferred to Hybond $N^+$ membrane (Amersham) by capillary transfer in 0.4 M NaOH. Probes used in the hybridization of pulsed-field gels were prepared by PCR amplification using primers to exons 3 and 8 of the SMN gene and to exon 5 of the NAIP gene. Amplified products were gel-purified (Geneclean; Bio 101) and then were labeled to high specific activity by random priming (Pharmacia). Hybridization was performed at 65°C in Church buffer (0.5 M $NaHPO_4$ [pH 7.2]/1 mM EDTA/7% SDS) overnight. Filters were washed to a final stringency of 0.5 × SSC/0.1% SDS at 65°C.

## Results

### *Polymorphism in the SMA Region*

In order to investigate the degree of variability in the SMA region, probes corresponding to exon 3 of SMN (SMN3) and to exon 5 of NAIP (NAIP5) were used for hybridization of genomic DNA digested with rare-cutting restriction enzymes. Figure 2 shows the analysis of seven unrelated Caucasian normal individuals whose

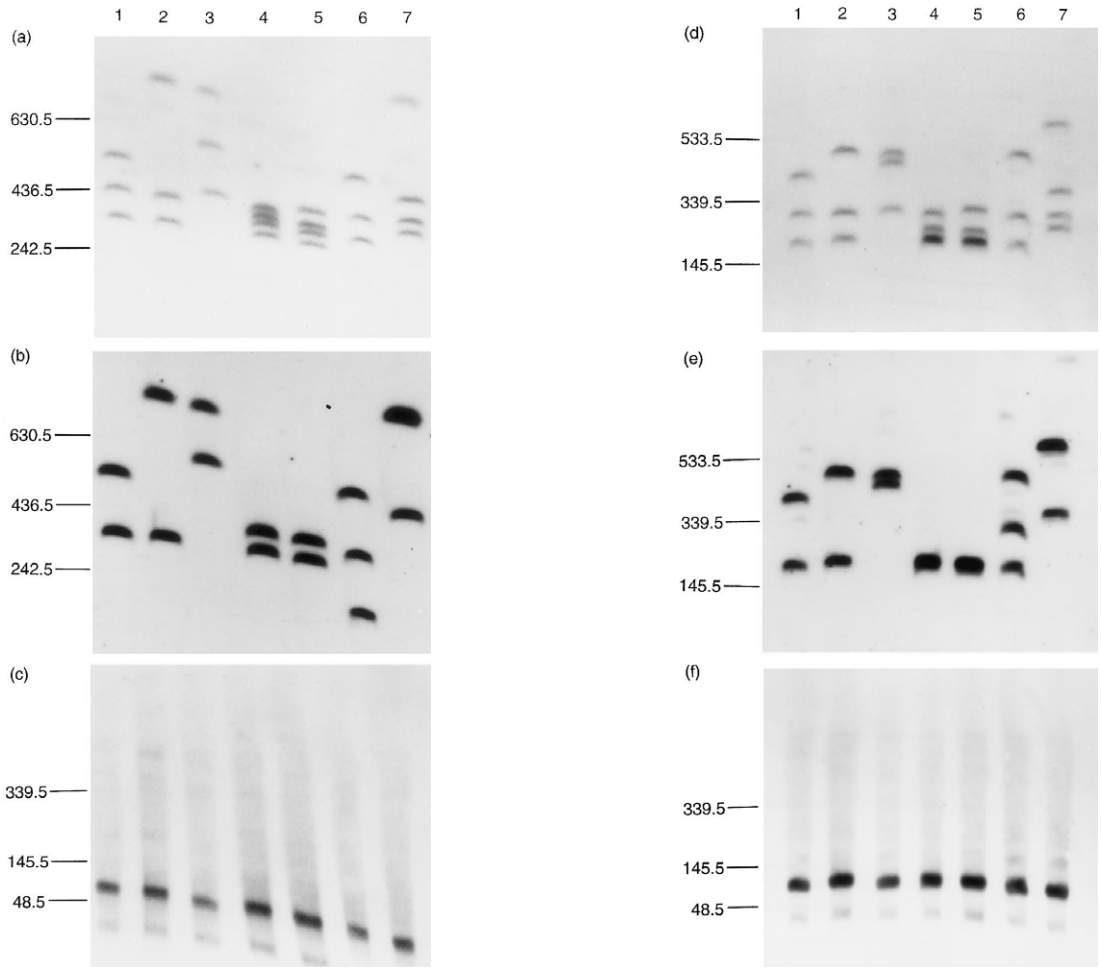


**Figure 2** Analysis of normal individuals. *Eag*I (*a–c*) and *Bss*HII (*d–f*) digests of DNA of seven unrelated Caucasian normal individuals, hybridized with SMN3 (*a* and *d*), NAIP5 (*b* and *e*), and utrophin (*c* and *f*) probes. Sizes (in kbp) relate to lambda-ladder PFG markers.

DNA was digested with either *Bss*HII (fig. 2*a–c*) or *Eag*I (fig. 2*d–f*). SMN exon 3 contains neither a *Bss*HII nor an *Eag*I restriction site. Each band observed must therefore relate to at least one copy of the gene. In some cases, where the band intensity appears greater than normal, it may be that more than one band of the same size is present, as is likely to be the case in lanes 4 and 5 of figure 2*d* and *e*. Thus, hybridization with SMN3 shows that the total number of SMN gene copies among normal individuals varies from three to four (fig. 2*a* and *d*). A high degree of variability was observed in the size of hybridizing bands, the *Eag*I bands varying in size from ~200 kb to ~600 kb in unrelated individuals. The same filters were subsequently hybridized with NAIP5, which is contained only within the telomeric region of the SMA locus and which is lacking in the centromeric NAIP pseudogene.(fig. 2*b* and *e*) Exon 5 of the NAIP gene contains an *Eag*I restriction site that we have found to be methylated in genomic DNA (data not shown). The same exon does not contain any *Bss*HII sites. Each band again relates to at least one copy of the gene. Thus, it is possible to identify the SMN and NAIP cohybridizing bands as telomeric copies. The number of SMNtel genes in the individuals shown is generally two. The individual whose DNA is shown in lane 6 has three bands hybridizing to NAIP5 but still has two SMNtel copies, since only two of the bands cohybridize with SMN3. This is the case with both *Eag*I and *Bss*HII. In the case of *Eag*I, the bands appear to be very similar in size, but, when autoradiographs of both hybridizations were overlaid, one of the three bands was clearly of a different size. Many NAIP pseudogenes consisting of various exons exist in the genome, and the third band may represent such a sequence (Roy et al. 1995*a*). We have subsequently observed three NAIP5 bands in 1/10 other control individuals. The pattern of bands observed with either *Eag*I or *Bss*HII is very similar within each individual, although the larger size range with *Bss*HII indicates that this enzyme cuts outside the *Eag*I sites. These data suggest that restriction sites for both enzymes

occur on either side of each duplicated region containing the SMN and NAIP genes. Hybridization of a probe specific to the utrophin gene on chromosome 6 (fig. 2*c* and *f*) shows a single band of the same size in all individuals, indicating that the variability seen with SMN and NAIP probes is not generated by partial digestion.

### *Mendelian Inheritance of SMA-Region Polymorphism*

In order to determine whether the instability observed is heritable, we analyzed eight members of a non-SMA family, the results of which are presented in figure 3. This *Bss*HII digest was hybridized with the SMN3 (fig. 3*a*) and NAIP5 probes (fig. 3*b*). Seven members of the family have four SMN copies, two of which, because of cohybridization of SMN3 and NAIP5, are designated telomeric. The individual whose DNA is shown in lane 5 shows only two SMN copies, one of which is telomeric, but he is believed to have inherited the same-size bands from both parents. When this filter was hybridized with the utrophin probe, a band of lesser intensity was observed in this lane, both indicating that less DNA is present and accounting for the lack of dosage difference (data not shown). These data show that the variation in band size is inherited in a Mendelian fashion.

### *Variability within an Inbred Population*

In view of the high degree of variation in the normal population, it was of interest to determine whether a highly inbred population, in which only a few founder mutations might be expected to exist, would show a lower degree of fragment-size variability. Since a number of the individuals in our patient database originate from Finland, we chose to investigate the size variability in controls from one region of this inbred population.

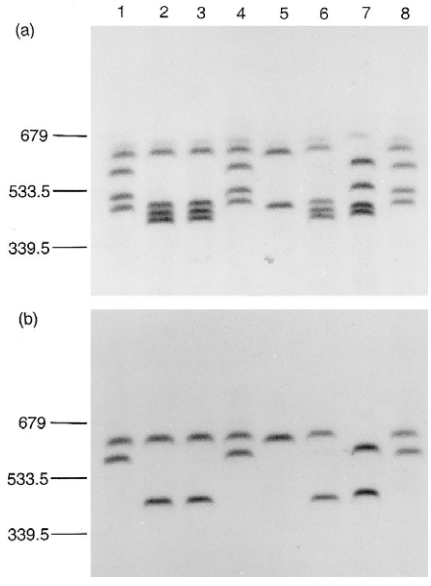


**Figure 3** Mendelian inheritance of the variation in a non-SMA family. *Bss*HII digests of DNA of eight members of a non-SMA family, hybridized with SMN3 (*a*) and NAIP5 (*b*) probes. Lane 1, Father. Lane 2, Mother. Lanes 3–8, Offspring. Sizes (in kbp) relate to lambda-ladder PFG markers.

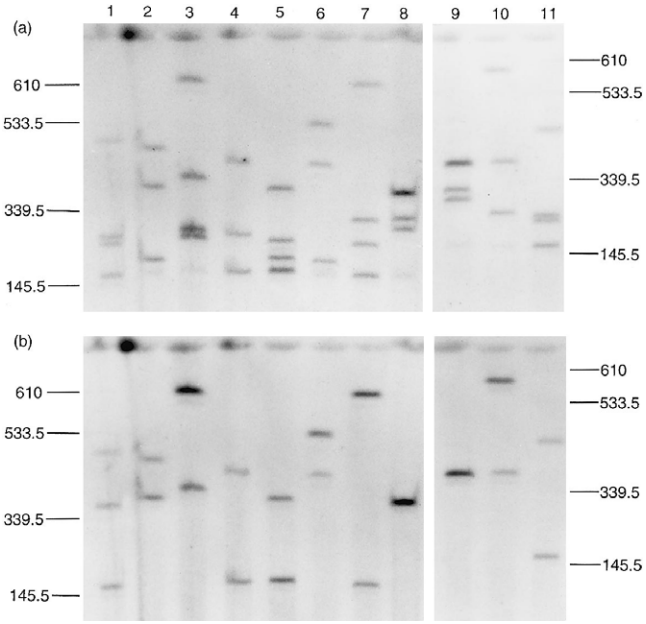


**Figure 4** Analysis of variation in northern Finnish control individuals: *Eag*I digests of DNA of 10 unrelated control individuals, hybridized with SMN3 (*a*) and NAIP5 (*b*) probes. Lanes 8 and 9 show DNA from the same individual, run on different gels to allow comparison. Sizes (in kbp) relate to lambda-ladder or yeast-chromosome PFG markers.

Figure 4 shows the results obtained when *Eag*I digests of DNA from 10 northern Finnish non-SMA individuals were hybridized to the SMN3 (fig. 4*a*) and NAIP5 probes (fig. 4*b*). The same type of variability was observed as was seen in the seven non-Finnish control samples, and the total number of SMN copies again varied from three to four. SMNtel copies in these controls varied between one and three. The individual whose DNA is shown in lanes 8 and 9 (DNA of the same individual is present in both lanes), who has only one NAIP5-hybridizing band, may be a carrier for the disease, having only one SMNtel gene, or may have more than one fragment of this size. The size of *Eag*I bands in Finnish control individuals varies between ~180 kb and ~600 kb, a size range similar to that observed in the non-Finnish controls (see fig. 2*d* and *e*). This amount of variability in such an inbred population is indicative of a high degree of instability within the SMA region.

### *PCR Deletion Analysis of SMN and NAIP*

Deletion analysis of the SMN and NAIP genes was performed on the patients to be analyzed by PFGE, as

described in the Methods section. Patients were scored as to their deletion status for SMN exons 7 and 8 independently. Patients were scored as positive or negative for NAIP exon 5, the telomeric-gene marker. A summary of these results is presented in table 1.

### *SMN Copy Number in SMA Type I Patients*

The copy number of SMN was investigated in the same five SMA type I patients as were analyzed by the PCR deletion method (see table 1). A blot containing *Eag*I digests of the DNA of five SMA type I patients was hybridized to the SMN3 probe (fig. 5*a*). The individual whose DNA is shown in lane 5 is from a consanguineous family. The total number of SMN gene copies was assessed directly from the number of bands visible. Subsequent hybridization to the NAIP5 probe allowed identification of cohybridizing bands as telomeric copies (fig. 5*b*). Type I individuals carry one or two SMN copies, of the centromeric type only, since none show cohybridization of NAIP5. This agrees with the PCR deletion data (see table 1). So far, it had been assumed that any SMN band not cohybridizing with NAIP exon 5 would be centromeric. To confirm this, a probe specific to SMN exon 8 was hybridized to the same filter, and hybridization in the affected individuals was indeed observed only to the bands previously believed to be centromeric (fig. 5*c*). The control individual whose DNA is shown in figure 5 showed four SMN bands, on hybridization to SMN3 and SMN8. Two of these SMN copies were designated telomeric, because of cohybridization with NAIP5.

In conclusion, type I chromosomes appear to result from the deletion of a large genomic region, encompassing SMN exon 3, SMN exon 8, and NAIP exon 5. Since the distance between SMNtel and the telomeric NAIP gene has been estimated to be ~20 kb (Lefebvre et al. 1995), the minimum size of such a deletion would be expected to be ~70 kb.

### *SMN Copy Number in SMA Type II and Type III Patients*

SMN copy number was investigated in the same seven type II and three type III patients as were analyzed by the PCR deletion method (see table 1). Southern blots of *Eag*I digests of DNA of these patients were hybridized to the SMN3 probe. The seven type II individuals and two of the three type III SMA individuals carry two or three SMN copies, one of which must exist within the telomeric duplicated region, because of cohybridization of NAIP5 (fig. 6*a* and *b*). The individual whose DNA is shown in lane 1 (SMA type III) is a patient from a consanguineous family, who shows only one SMN band, although it is of increased intensity, suggesting more than one copy of the same size. Hybridization of other probes to this filter indicated a slight increase in the amount of DNA in this lane compared with the

**Table 1**

**Deletion Analysis of SMN and NAIP in SMA Patients**

| | Result In[b] | | | | |
|---|---|---|---|---|---|
| | SMN Exon 7 | | SMN Exon 8 | | |
| Patient[a] | Centromere | Telomere | Centromere | Telomere | NAIP: Exon 5 |
| I-1 | + | − | + | − | − |
| I-2 | + | − | + | − | − |
| I-3 | + | − | + | − | − |
| I-4 | + | − | + | − | − |
| I-5 | + | − | + | − | − |
| I/II-1 | + | − | + | − | + |
| II-2 | + | − | + | − | + |
| II-3 | + | − | + | − | + |
| II-4 | + | − | + | + | + |
| II-5 | + | − | + | − | + |
| II-6 | + | − | + | − | + |
| II-7 | + | − | + | − | + |
| III-1 | + | − | + | − | − |
| III-2 | + | − | + | − | + |
| III-3 | + | − | + | − | + |

[a] Designation consists of SMA type (roman numeral) and number assigned to individual (arabic numeral) and correlates with that given in figures 5 and 6.

[b] A plus sign (+) denotes presence of a PCR product; and a minus sign (−) denotes absence of a PCR product.

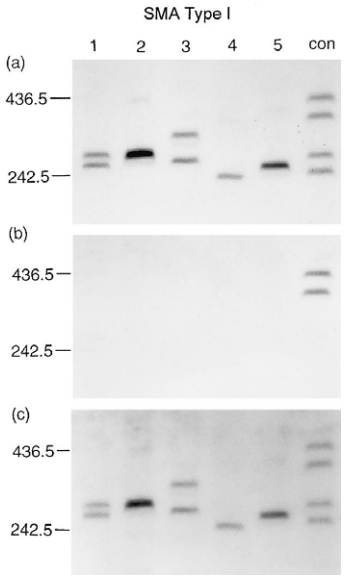


**Figure 5** Analysis of SMN gene copy number in SMA type I patients: *Eag*I digests of DNA of five SMA type I patients, hybridized with SMN3 (*a*), NAIP5 (*b*), and SMN8 (*c*) probes. Lane 5 shows DNA from a patient from a consanguineous family. con = control individual. Sizes (in kbp) relate to lambda-ladder PFG markers.

others. However, this increase does not fully account for the extra intensity observed on SMN exon 3 and exon 8 hybridization, and we conclude that this band contains at least two SMNcen copies. This patient does not carry a telomeric band, and it cannot be determined whether the functional NAIP gene has been lost as a result of a deletion or as a result of a gene-conversion event. Thus, in all type II and type III patients except the patient whose DNA is shown in lane 1 (SMA type III), both hybridization and PCR deletion analysis suggest the presence of NAIP exon 5. If deletion of the telomeric duplicated region has occurred in these patients, it is not as large a deletion as is seen in type I SMA, since NAIP exon 5 remains. In conclusion, the total number of SMN copies in type II and type III individuals appears to be greater than that in type I.

### *Gene Conversion of SMNtel to SMNcen in Type II and Type III Chromosomes*

SMN PCR deletion analysis showed an absence of amplification of SMNtel exons 7 and 8 in six of seven type II patients and in all three type III patients, suggesting deletion of these exons (see table 1). It was therefore expected that the SMN exon 8–specific probe would hybridize only to the SMNcen bands in these patients. As can be seen in figure 6*c*, this is not the case; the SMN8 probe hybridizes to the telomeric band as well as to the centromeric bands. This result conflicts with the PCR deletion data, suggesting that exon 8 of SMNtel is not deleted. This discrepancy in the data could be explained if the SMNtel gene was not deleted but, rather, was replaced by a copy of SMNcen by a mechanism such as gene conversion (see Discussion). Absence of SMNtel exon 7 but presence of SMNtel exon 8 was detected for the patient whose DNA is shown in lane 4 of figure 6, suggesting that a partial gene conversion has occurred.

## Discussion

The data presented in this paper demonstrate that a high degree of variability between individuals exists in the SMA genomic region. In SMA types II and III, the PCR-detected mutations previously classed as deletions are likely to be the result of gene-conversion events, and this results in an increase in SMNcen copies, SMNtel having undergone gene conversion to SMNcen, which correlates with severity of SMA phenotype.

We have shown that a genomic size variation in the SMA gene region occurs in the normal population as well as in SMA patients and that it is inherited in a Mendelian fashion. The Finnish control samples used in this study originate from the province of Oulu in northern Finland, where several autosomal recessive diseases belonging to the Finnish disease heritage are known to be enriched. These include congenital chloride diarrhea, lethal congenital contracture syndrome, nonketotic hyperglycinemia, and northern epilepsy (Leisti et al. 1990;

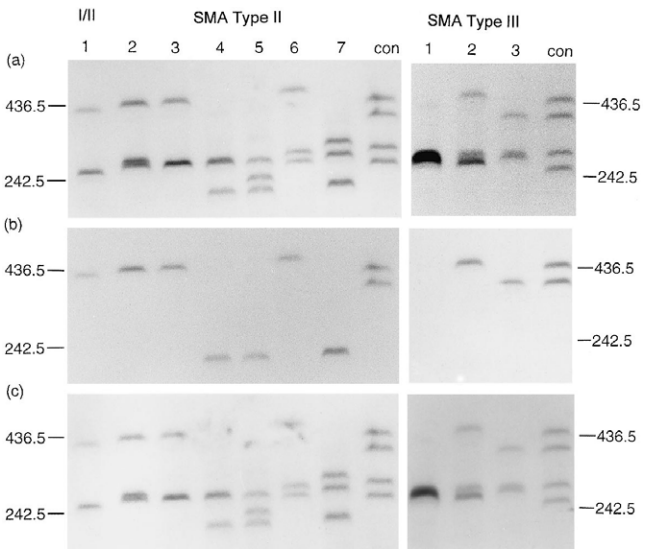


**Figure 6** Analysis of SMN gene copy number in SMA type II and type III patients: *Eag*I digests of DNA of seven SMA type II individuals and three SMA type III individuals, hybridized with SMN3 (*a*), NAIP5 (*b*), and SMN8 (*c*) probes. Lane 1 (SMA type II) shows DNA of a patient whose phenotype is considered borderline between type I and type II. Lane 1 (SMA type III) shows DNA from a patient from a consanguineous family. con = control individual (the same individual as in fig. 5). Sizes (in kbp) relate to lambda-ladder PFG markers.

de la Chapelle 1993). A high frequency of childhood SMA is also observed in this region, the birth incidence of SMA with onset at age <12 mo being 1/7,100 and the estimated minimum carrier frequency being 1/42 (Ignatius 1992). The observation of genomic variability in the SMA region of this highly inbred Finnish population indicates a high degree of instability within this region. In view of these data, it is possible to speculate on the existence of a midisatellite type of repeated sequence, the expansion and contraction of which, by a method such as unequal crossing-over, could result in the formation of alleles differing in size by multiples of the repeat size. One such sequence has been reported to exist at a single locus on chromosome 1q, and PFGE has shown the presence of allelic polymorphic fragments varying in size by as much as 250 kb, depending on the enzymes used (Nakamura et al. 1987). Two such alleles were observed in each individual. If a copy of a midisatellite sequence exists associated with each SMN gene, we would expect to see four such variable alleles, as indeed were observed when we used SMN gene probes. The presence of such a sequence would greatly enhance the instability of the SMA region and could lead to an increase in the mutation rate, since the repeat could mediate mutations such as gene-deletion and -conversion events, resulting in the disease. Although the mechanism of gene conversion is not known, the presence of repeat units may facilitate alignment of nonhomologous strands, allowing sequence conversion, when one repeat unit is used as template for DNA repair following strand breakage.

Our data substantiate those of other groups and suggest that the majority of type I SMA cases do indeed result from large-scale deletions, extending over $\geq$70 kb and encompassing the SMN and NAIP genes. Our data also show that type I SMA is characterized by the presence of one or two SMNcen genes only. The lack of hybridization of the NAIP exon 5 probe to any band suggests that no telomeric genes exist in the type I SMA families examined.

Two possibilities exist to explain the role of NAIP in the disease. It may be that, as a result of its proximity to this gene, NAIP is coincidentally deleted along with SMNtel and that its presence or absence has no effect on the disease phenotype. Deletion of NAIP exon 5 is observed in 2% of SMA carriers and so can be associated with a normal phenotype (Roy et al. 1995*a*). Alternatively, NAIP deletion may contribute to the severity of the phenotype, by generating an effect additive to that of SMNtel deletion. Indeed, deletion of NAIP is associated more often with severe SMA than with intermediate or mild forms of the disease. In this study, the presence of extra loci containing NAIP exon 5 was suggested in 2/17 control individuals. Such a finding, even in a small number of samples, implies that the deletion rate for this gene could be higher than that previously predicted from the NAIP exon 5 deletion data.

We have shown that nonconsanguineous type II and type III SMA patients carry two or three copies of the SMN gene, one of which, by cohybridization with the NAIP exon 5 probe, appeared to be telomeric in origin. PCR deletion analysis had revealed a lack of amplification of SMNtel exons 7 and 8, but hybridization of an SMN exon 8 probe to the telomeric band in these patients suggested that this duplicated region still possessed a copy of SMN exon 8. Since the SMNtel and SMNcen exons 7 and 8 are easily distinguishable by the PCR and enzyme-digestion method used in the deletion analysis, we suggest that the exon 8 in question is indeed centromeric in origin. We propose that type II and type III SMA mutant chromosomes occur as a result of gene-conversion events in which SMNtel is converted to SMNcen. In this way, patients do not possess a functional SMNtel gene, although no physical DNA deletion has occurred. Other genome regions containing duplicated genes show a degree of gene-sequence conversion in those patients in whom gene deletion is not observed (Collier et al. 1989).

The number of copies of SMNcen is greater in types II and III than in type I, since SMNtel copies are not deleted but are converted to SMNcen. The greater number of SMNcen copies may serve to ameliorate the SMA phenotype in the less severe phenotypes, types II and III. Several groups have tested this hypothesis, using various techniques, including investigation of the deletion of alleles of the multicopy marker C272, which is contained within the 5′ end of each SMN gene (DiDonato et al. 1994; Melki et al. 1994; Lefebvre et al. 1995; Wirth et al. 1995). Correlation of the presence of fewer C272 alleles in type I SMA compared with types II and III suggested that decreased copy number of C272 alleles—and, therefore, of SMN—led to an increase in phenotypic severity. Wirth et al. also proposed a model based on the existence of mild and severe disease alleles and on the suggestion that compound heterozygosity of these alleles results in varying severities of the disease. Velasco et al. (1996) employed densitometry of SSCP bands to observe an increased number of SMNcen copies in parents of type II and type III patients, compared with parents of type I patients, and to suggest that this correlated with a disease of lower severity in type II and type III offspring. In a further study, solid-phase minisequencing was used to obtain an SMNtel:SMNcen ratio of gene copies (Schwartz et al. 1997). Although proposing a model in which a number of normal and SMA haplotypes exist, which, in combination, give rise to variable disease severity, those authors were unable to correlate the number of SMNcen copies to severity. In addition, it has been proposed that compound heterozygosity of different disease alleles gives rise to the observation of

different SMA severities within one family (Talbot et al. 1996).

The data presented in this study suggest the existence of four SMA alleles, consisting of deletion of SMNtel in addition to presence of one or no copies of SMNcen, or conversion of SMNtel to SMNcen in addition to presence of one or no copies of SMNcen. Figure 1*B* shows a schematic representation of these alleles. Combinations of these disease alleles would give rise to SMA phenotypes of varying severities; type I, for example, would be the result of two chromosomes in which SMNtel has been deleted. On the basis of this model, for instance, this situation would be represented by a genotype consisting of SMA alleles 1,2 or 2,2 (see fig. 1*B*). SMA type I patients homozygous for the most severe allele, consisting of deletion of SMNtel and the absence of SMNcen gene (genotype 1,1) are not observed, suggesting that this combination would be lethal in utero. A similar model recently has been proposed by DiDonato et al. (1997), who observed a strong association between the milder forms of SMA (types II and III) and partial gene-conversion events not involving exon 8. However, it should be noted that other groups have observed these events also in type I SMA (Hahnen et al. 1996; van der Steege et al. 1996; Velasco et al. 1996; Talbot et al., in press). The observation of one NAIP5-specific band in all but one of the type II and type III patients examined, in contrast to the two bands present in most normal individuals, suggests that these patients are compound heterozygotes with one deleted allele and one gene-converted allele. Although we were unable to include in this study any type I patients who retain an intact NAIP gene, it would be of interest to examine such patients, in order to determine whether the lack of amplification of SMNtel exons in such a case would be due to deletion or to gene conversion. The model proposed by Emery (1991), in which one normal allele exists and combinations of at least four mutant alleles could account for SMA types I–III, would have to be modified to include two normal alleles, consisting of one SMNtel gene and either one or no SMNcen genes (see fig. 1*B*).

Lefebvre et al. (1995) showed that transcription of SMNtel results in production of a full-length transcript, thereby encoding a fully functional protein. SMNcen transcription, on the other hand, results in an alternative transcript lacking exon 7 and in a significant decrease in levels of the full-length transcript. The alternative transcript is spliced in-frame from exon 6 to exon 8, generating a truncated protein. The importance of exons 6 and 7 has been suggested recently by Talbot et al. (1997), who have proposed that a highly conserved tyrosine-glycine (Y-G) dodecapeptide motif in the region of the protein encoded by exons 6 and 7 is crucial to the correct functioning of the protein and that mutations affecting this region result in SMA. Patients have been described with mutations predicted to disrupt this region and that would therefore be expected to disrupt this motif (Lefebvre et al. 1995; Talbot et al. 1997). Other mutations result in premature truncation of the SMN protein (Bussaglia et al. 1995; Brahe et al. 1996; Parsons et al. 1996). Thus, SMA patients with either deletion of SMNtel or gene conversion of SMNtel to SMNcen will produce a smaller-than-normal amount of functional SMN protein, the amount being dependent on the number of SMNcen genes present.

It has been shown that the level of SMN protein detectable on a western blot is much reduced in SMA type I (Lefebvre et al. 1996). This result is in accordance with our observation that a minimal number (one or two) of SMNcen copies are carried by SMA type I individuals. In contrast, an almost normal amount of SMN protein is produced in type III patients, indicating a correlation between the amount of SMN protein and the clinical expression of the disease. The size difference between the proteins translated from the SMNcen full-length and truncated transcripts is only 17 amino acids. It is not possible by conventional western analysis to distinguish between the two resultant proteins. The observation of an almost normal level of SMN protein in type III suggests that the truncated protein is indeed translated. It appears that, because of the presence of a greater number of SMNcen copies, an increased amount of fully functional protein is produced as the severity of the disease decreases.

In this study we have highlighted a difference, in the mechanism of formation, between SMA type I chromosomes and those of types II and III. We have shown that gene conversion plays a major role in type II and type III SMA. We propose that an increase in SMNcen copy number leads to a decrease in severity of the disease, by the production of increased levels of functional SMN protein, low levels of which are encoded by this gene. If this proves to be the case, as is also indicated by the demonstration of the importance of SMNcen copy number in severity when the marker C272 (or Ag1CA) is used, then the up-regulation of this centromeric gene could hold the key to future therapeutic advances in SMA.

## Acknowledgments

We wish to thank Kevin Talbot, Nanda Rodrigues, Jon Tinsley, Francesco Muntoni and Nicholas Owen for helpful discussions, Steve Phelps for initial help with tissue culture, and Dr. Peter Harris for normal lymphoblastoid cell lines. We are grateful to the SMA families who kindly donated samples for this research. This work was supported by the Muscular Dystrophy Association (USA), the Muscular Dystrophy Group (UK) and the Medical Research Council (UK).

## References


Brahe C, Clermont O, Zappata S, Tiziano F, Melki J, Neri G (1996) Frameshift mutation in the survival motor neuron gene in a severe case of SMA type I. Hum Mol Genet 5: 1971–1976

Brzustowicz LM, Lehner T, Castilla LH, Penchaszadeh GK, Wilhelmsen KC, Daniels R, Davies KE, et al (1990) Genetic mapping of chronic childhood-onset spinal muscular atrophy to 5q11.2-13.3. Nature 344:540–541

Bürglen L, Seroz T, Miniou P, Lefebvre S, Burlet P, Munnich A, Pequignot EV, et al (1997) The gene encoding p44, a subunit of the transcription factor TFIIH, is involved in large-scale deletions associated with Werdnig-Hoffman disease. Am J Hum Genet 60:72–79

Burlet P, Bürglen L, Clermont O, Lefebvre S, Viollet L, Munnich A, Melki J (1996) Large scale deletions of the 5q13 region are specific to Werdnig-Hoffmann disease. J Med Genet 33:281–283

Bussaglia E, Clermont O, Tizzano E, Lefebvre S, Bürglen L, Cruaud C, Urtizberea JA, et al (1995) A frame-shift deletion in the survival motor neuron gene in Spanish spinal muscular atrophy patients. Nat Genet 11:335–337

Carpten JD, DiDonato CJ, Ingraham SE, Wagner MC, Nieuwenhuijsen BW, Wasmuth JJ, Burghes AH (1994) A YAC contig of the region containing the spinal muscular atrophy gene (SMA): identification of an unstable region. Genomics 24:351–356

Cobben JM, van der Steege G, Grootscholten P, de Visser M, Scheffer H, Buys CHCM (1995) Deletions of the survival motor neuron gene in unaffected siblings of patients with spinal muscular atrophy. Am J Hum Genet 57:805–808

Collier S, Sinnott PJ, Dyer PA, Price DA, Harris R, Strachan T (1989) Pulsed field gel electrophoresis identifies a high degree of variability in the number of tandem 21-hydroxylase and complement C4 gene repeats in 21-hydroxylase deficiency haplotypes. EMBO J 8:1393–1402

de la Chapelle A (1993) Disease gene mapping in isolated human populations: the example of Finland. J Med Genet 30:857–865

Devriendt K, Lammens M, Schollen E, Van Hole C, Dom R, Devlieger H, Cassiman J-J, et al (1996) Clinical and molecular genetic features of congenital spinal muscular atrophy. Ann Neurol 40:731–738

DiDonato CJ, Ingraham SE, Mendell JR, Prior TW, Lenard S, Moxley RT III, Florence J, et al (1997) Deletion and conversion in spinal muscular atrophy patients: is there a relationship to severity? Ann Neurol 41:230–237

DiDonato CJ, Morgan K, Carpten JD, Fuerst P, Ingraham SE, Prescott G, McPherson JD, et al (1994) Association between Ag1-CA alleles and severity of autosomal recessive proximal spinal muscular atrophy. Am J Hum Genet 55:1218–1229

Dubowitz V (1995) Muscle disorders in childhood, 2d ed. WB Saunders, London

Emery AEH (1991) Clinical and genetic heterogeneity in spinal muscular atrophy—the multiple allele model. Neuromuscul Disord 1:307–308

Francis MJ, Morrison KE, Campbell L, Grewal PK, Christodoulou Z, Daniels RJ, Monaco AP, et al (1993) A contig of non-chimaeric YACs containing the spinal muscular atrophy gene in 5q13. Hum Mol Genet 2:1161–1167

Francis MJ, Nesbit MA, Theodosiou AM, Rodrigues NR, Campbell L, Christodoulou Z, Qureshi SJ, et al (1995) Mapping of retrotransposon sequences in the unstable region surrounding the spinal muscular atrophy locus in 5q13. Genomics 27:366–369

Gilliam TC, Brzustowicz lM, Castilla LH, Lehner T, Penchaszadeh R, Daniels RJ, Byth BC, et al (1990) Genetic homogeneity between acute and chronic forms of spinal muscular atrophy. Nature 345:823–825

Hahnen E, Forkert R, Marke C, Rudnik-Schöneborn S, Schönling J, Zerres K, Wirth B (1995) Molecular analysis of candidate genes on chromosome 5q13 in autosomal recessive spinal muscular atrophy: evidence of homozygous deletions of the SMN gene in unaffected individuals. Hum Mol Genet 4:1927–1933

Hahnen E, Schönling J, Rudnik-Schöneborn S, Zerres K, Wirth B (1996) Hybrid survival motor neuron genes in patients with autosomal recessive spinal muscular atrophy: new insights into molecular mechanisms responsible for the disease. Am J Hum Genet 59:1057–1065

Humbert S, van Vuuren H, Lutz Y, Hoeijmakers JHJ, Egly JM, Moncollin V (1994) p44 and p34 subunits of the BTF2/TFIIH transcription factor have homologies with SSL1, a yeast protein involved in DNA repair. EMBO J 13:2393–2398

Ignatius J (1992) Childhood spinal muscular atrophy in Finland. PhD thesis, University of Tampere. Acta Univ Tamperensis A, vol 358. Tampere, Finland

Kleyn PW, Wang CH, Lien LL, Vitale E, Pan J, Ross BM, Grunn A, et al (1993) Construction of a yeast artificial chromosome contig spanning the spinal muscular atrophy disease gene region. Proc Natl Acad Sci USA 90:6801–6805

Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, et al (1995) Identification and characterization of a spinal muscular atrophy-determining gene. Cell 80:155–165

Lefebvre S, Burlet P, Bertrandy S, Clermont O, Bürglen L, Viollet L, Munnich A, et al (1996) Strong correlation between the survival motor neuron (SMN) protein level and SMA phenotypes. Am J Hum Genet Suppl 59:A268

Leisti J, Jouppila P, Mustonen A, Kahkonen M, Herva R, Ruokonen A, Kirkinen P (1990) Prenatal diagnosis of single gene disorders in northern Finland. Ann Med 22:123–129

Liu Q, Dreyfuss G (1996) A novel nuclear structure containing the survival of motor neurons protein. EMBO J 15:3555–3565

Matthijs G, Schollen E, Legius E, Devriendt K, Goemans N, Kayserili H, Apak MY, et al (1996) Unusual molecular findings in autosomal recessive spinal muscular atrophy. J Med Genet 33:469–474

Melki J, Abdelhak S, Sheth P, Bachelot MF, Burlet P, Marcadet A, Aicardi J, et al (1990*a*) Gene for chronic proximal spinal muscular atrophies maps to chromosome 5q. Nature 344: 767–768

Melki J, Lefebvre S, Bürglen L, Burlet P, Clermont O, Millasseau P, Reboullet S, et al (1994) De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. Science 264:1474–1477

Melki J, Sheth P, Abdelhak S, Burlet P, Bachelot MF, Lathrop MG, Frezal J, et al (1990*b*) Mapping of acute (type I) spinal muscular atrophy to chromosome 5q12-q14. Lancet 336: 271–273

Munsat TL, Davies KE (1992) Meeting report: International SMA consortium meeting. Neuromuscul Disord 2:423–428

Nakamura Y, Julier C, Wolff R, Holm T, O'Connell P, Leppert M, White R (1987) Characterization of a human 'midisatellite' sequence. Nucleic Acids Res 15:2537–2547

Parsons WD, McAndrew P, Monani U, Mendell J, Burghes A, Prior T (1996) An 11 base pair duplication in exon 6 of the SMN gene produces a type I spinal muscular atrophy (SMA) phenotype: further evidence for SMN as the primary SMA determining gene. Hum Mol Genet 5:1727–1732

Rodrigues NR, Owen N, Talbot K, Ignatius J, Dubowitz V, Davies KE (1995) Deletions in the survival motor neuron gene on 5q13 in autosomal recessive spinal muscular atrophy. Hum Mol Genet 4:631–634

Rodrigues NR, Owen N, Talbot K, Patel S, Muntoni F, Ignatius J, Dubowitz V, et al (1996) Gene deletions in spinal muscular atrophy. J Med Genet 33:93–96

Roy N, Mahadevan MS, McLean M, Shutler G, Yaraghi Z, Farahani R, Baird S, et al (1995*a*) The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. Cell 80:167–178

Roy N, McLean MD, Besner JA, Lefebvre C, Salih M, Carpten JD, Burghes AH, et al (1995*b*) Refined physical map of the spinal muscular atrophy gene (SMA) region at 5q13 based on YAC and cosmid contiguous arrays. Genomics 26:451–460

Sargent CA, Chalmers IJ, Leversha M, Affara NA (1994) A rearrangement on chromosome 5 of an expressed human betaglucuronidase pseudogene. Mamm Genome 5:791–796

Schwartz M, Sorensen N, Hansen FJ, Hertz JM, Norby S, Tranebjaerg L, Skovby F (1997) Quantification, by solid-phase minisequencing, of the telomeric and centromeric copies of the survival motor neuron gene in families with spinal muscular atrophy. Hum Mol Genet 6:99–104

Selig S, Bruno S, Scharf JM, Wang CH, Vitale E, Gilliam TC, Kunkel LM (1995) Expressed cadherin pseudogenes are localized to the critical region of the spinal muscular atrophy gene. Proc Natl Acad Sci USA 92:3702–3706

Talbot K, Ponting CP, Theodosiou AM, Rodrigues NR, Surtees R, Mountford R, Davies KE (1997) Missense mutation clustering in the survival motor neuron gene: a role for a conserved tyrosine and glycine rich region of the protein in RNA metabolism? Hum Mol Genet 6:497–500

Talbot K, Rodrigues N, Bernert G, Bittner R, Davies K (1996) Evidence for compound heterozygosity causing mild and severe forms of autosomal recessive spinal muscular atrophy. J Med Genet 33:1019–1021

Talbot K, Rodrigues NR, Ignatius J, Muntoni F, Davies KE. Gene conversion at the SMN locus in autosomal recessive spinal muscular atrophy does not predict a mild phenotype. Neuromuscul Disord (in press)

Theodosiou AM, Morrison KE, Nesbit AM, Daniels RJ, Campbell L, Francis MJ, Christodoulou Z, et al (1994) Complex repetitive arrangements of gene sequence in the candidate region of the spinal muscular atrophy gene in 5q13. Am J Hum Genet 55:1209–1217

van der Steege G, Grootscholten PM, Cobben JM, Zappata S, Scheffer H, den Dunnen JT, van Ommen G-JB, et al (1996) Apparent gene conversion involving the SMN gene in the region of the spinal muscular atrophy locus on chromosome 5. Am J Hum Genet 59:834–838

van der Steege G, Grootscholten PM, van der Vlies P, Draaijers TG, Osinga J, Cobben JM, Scheffer H, et al (1995) PCR-based DNA test to confirm clinical diagnosis of autosomal recessive spinal muscular atrophy. Lancet 345:985–986

Velasco E, Valero C, Valero A, Moreno F, Hernandez CC (1996) Molecular analysis of the SMN and NAIP genes in Spanish spinal muscular atrophy (SMA) families and correlation between number of copies of cBCD541 and SMA phenotype. Hum Mol Genet 5:257–263

Wang CH, Xu J, Carter TA, Ross BM, Dominski MK, Bellcross CA, Penchaszadeh GK, et al (1996) Characterization of survival notor neuron ($SMN^T$) gene deletions in asymptomatic carriers of spinal muscular atrophy. Hum Mol Genet 5:359–365

Wirth B, Hahnen E, Morgan K, DiDonato CJ, Dadze A, Rudnik SS, Simard LR, et al (1995) Allelic association and deletions in autosomal recessive proximal spinal muscular atrophy: association of marker genotype with disease severity and candidate cDNAs. Hum Mol Genet 4:1273–1284