



- (51) **International Patent Classification:**
C12Q 1/68 (2006.01)
- (21) **International Application Number:**
PCT/US2012/041406
- (22) **International Filing Date:**
7 June 2012 (07.06.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/494,282 7 June 2011 (07.06.2011) US
- (71) **Applicant (for all designated States except US):** MOUNT SINAI SCHOOL OF MEDICINE [US/US]; One Gustave L. Levy Place, Box 1675, New York, NY 10029-6574 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** EDELMANN, Lisa [US/US]; Office of Technology & Business Development, One Gustave L. Levy Place, Box 1497, New York, NY 10029-6574 (US). DESNICK, Robert, J. [US/US]; Office of Technology & Business Development, One Gustave L. Levy Place, Box 1497, New York, NY 10029-6574 (US).
- (74) **Agent:** MERKEL, William, K.; Marshall, Gerstein & Borun LLP, 233 S. Wacker Drive, 6300 Willis Tower, Chicago, IL 60606-6357 (US).

(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** MATERIALS AND METHOD FOR IDENTIFYING SPINAL MUSCULAR ATROPHY CARRIERS

(57) **Abstract:** Materials and methods for identifying carriers of genetic determinants of spinal muscular atrophy are disclosed. In particular, polymorphisms in linkage disequilibrium are associated as markers of spinal muscular atrophy alleles detectable by various techniques, including multiplex ligation-dependent probe analysis, sequence analysis, and RFLP detection. The materials and methods of the disclosure are particularly useful in identifying silent (2+0) carriers of spinal muscular atrophy in which two copies of the SMN1 gene are located on a single human chromosome 5 and no copies of the gene are located on the chromosome 5 homo-



MATERIALS AND METHOD FOR IDENTIFYING SPINAL MUSCULAR ATROPHY CARRIERS

FIELD

[0001] The disclosed technology generally relates to the field of genetic counseling and, more particularly, to methods of identifying carriers of genetic determinants associated with deleterious conditions.

BACKGROUND

[0002] Spinal muscular atrophy is one of the most common and severe autosomal recessive diseases with an overall incidence of about 1 in 6,000 to 10,000 live births and a carrier frequency of 1 in 35 to 1 in 117, depending on ethnicity. The disease is characterized by the progressive degeneration and loss of anterior horn cells in the spinal cord and brain stem nuclei causing symmetric muscle weakness and atrophy, and the clinical subtypes are primarily based on age at onset. Type I disease (Werdnig-Hoffmann disease, MIM# 253300), which occurs in 60-70% of patients, is characterized by onset of respiratory insufficiency at birth or before six months of age leading to death within two years. Type I patients never sit or walk. Onset of Type II disease (MIM# 253550) typically occurs after 6 months of age; these infants can sit, but never walk unaided, and have significantly reduced life expectancy. Patients with Type III disease (Kugelberg-Welander syndrome, MIM# 253400) present after 18 months of age, stand and walk, but often become wheelchair-bound during childhood or early adulthood. Type IV disease (MIM# 271150) is characterized by adult onset (mean age of 35 years) and slow disease progression.

[0003] Homozygous mutations of the *SMN1* gene on chromosome 5q13.2 are the major cause of SMA (Lefebvre et al., Cell 80:155-165, 1995). In approximately 95-98% of SMA patients, both copies of *SMN1* exons 7 and 8 are either deleted or are rendered non-functional due to gene conversion of *SMN1* to *SMN2* (Ogino et al., J. Hum. Genet. 12:1015-1023 (2004)). The remaining 2-5% of patients are compound heterozygotes that carry an intragenic mutation(s) on one allele and either a deletion or gene conversion mutation on the other allele (Wirth et al., Am J. Hum. Genet. 64:1340-1356, 1999). *SMN1* and *SMN2* are separated by approximately 1.4 Megabases and comprise the telomeric and centromeric members, respectively, of a set of genes, including *NAIP*, present within a segmental duplication on 5q13.2. The *SMN1* (NCBI Reference Sequence NG_008691) and *SMN2* genes have high sequence similarity even in the promoter regions and there are no encoded amino acid differences. A single base change affecting a putative splice enhancer in exon 7 (840C>T; position 27006 of SEQ ID NO:1), however, accounts for splicing differences such

that the majority of *SMN2* transcripts lack exon 7 (Lorson et al., Proc. Natl. Acad. Sci. (USA) 96:6307-6311 (1999)). Although some genotype/phenotype correlations have been established among patients that carry *SMN1* point mutations, the presence of additional copies of *SMN2* positively modifies clinical prognosis (Wirth et al., 1999).

[0004] The majority of mutations causing all four SMA types (i.e., Types I-IV) involve *SMN1* copy number loss. Mutation detection typically involves PCR amplification of *SMN1* exon 7, which is homozygously absent in most affected individuals. Carrier screening, however, is performed by dosage-sensitive but location-insensitive methods that can distinguish *SMN1* and *SMN2*, but that provide no information on gene location. These carrier screens are currently performed using quantitative PCR (qPCR), Multiplex Ligation-Dependent Probe Amplification (MLPA), and/or Taqman quantitative technology. These methods cannot determine the number of *SMN1* copies present on individual chromosomes. Given the tight linkage of the structurally related *SMN1* and *SMN2* genes and the opportunity for gene conversion as well as gene duplication or deletion, it is unsurprising that these methods exhibit detectability varying from 71-94% due to the inability to identify individuals that are SMA silent carriers, such as silent (2+0) carriers. Information on gene location is urgently needed to allow genetic counseling to consider the effects of segregation in germ cell production and its implications for offspring. Individuals with two *SMN1* copies on one chromosome (a duplication allele) and no copies on the other (a deletion allele), are referred to as silent (2+0) carriers, as most individuals with two intact *SMN1* copies, one copy on each chromosome (1+1), are not carriers (Fig. 1). Thus, SMA carrier detection by current techniques that are location-insensitive generates false negative results.

[0005] The frequency of silent (2+0) carriers varies based on ethnicity and is directly proportional to the product of the frequency of deletion and duplication alleles in a given population. Among Ashkenazi Jews, the SMA carrier frequency has been reported at 1 in 41 with a detectability of approximately 90%. Of the remaining 10%, about 8% are silent (2+0) carriers and the rest carry intragenic mutations. The frequency of *SMN1* silent (2+0) carriers in any population modifies the risk of being a carrier after a negative screening result. For example, in the African American (AA) population, the frequency of duplication (2+1) individuals is relatively high (47%), and consequently there are more silent (2+0) carriers, such that the carrier detection rate is only 71% with a residual risk of 1 in 121 after a negative result.

[0006] Accordingly, a need exists in the art for screening methods for SMA and SMA carriers that are location-sensitive in addition to being sensitive to copy number, e.g., copy

numbers of *SMN1* and/or *SMN2* alleles. An accurate assessment of the risk of producing offspring with SMA requires knowledge of both the copy numbers of *SMN1* and/or *SMN2* and of the location of those alleles that are found in a given genome. Such information is vital to competent genetic counseling of prospective parents.

SUMMARY

[0007] The technology disclosed herein solves at least one of the aforementioned problems in the art. More specifically, the disclosure provides methods of identifying carriers of SMA, such as silent SMA carriers, as exemplified by silent (2+0) carriers of SMA. The ability to identify silent (2+0) carriers would significantly improve carrier detection and, therefore, efforts were directed to identifying ethnic-specific *SMN1* founder deletion and/or duplication alleles by detecting a genotype unique to either the deletion or duplication alleles present in silent (2+0) carriers.

[0008] The disclosure provides AJ founder alleles for *SMN1*, including one that is present in approximately half of all AJ *SMN1* duplication (2+1) individuals. This duplication haplotype can be identified by two tightly linked, highly specific polymorphisms in *SMN1* (SEQ ID NO:1), g.27134T>G (position 27134 of SEQ ID NO:1) in intron 7 and g.27706_27707delAT (positions 27706 and 27707 of SEQ ID NO:1) in exon 8. These polymorphisms, and other genetic features noted in this disclosure, bear the numbering used for the reference *SMN1* sequence submitted to Genbank under Accession No. NG_008691.1 (cognate mRNA reference sequence NM_000344.3; encoded amino acid reference sequence NP_000335.1), all incorporated herein by reference in their entireties. These polymorphisms can be used in conjunction with dosage-sensitive methods to detect silent (2+0) carriers and improve the overall carrier detection rate in a particular ethnic group, such as the Ashkenazi Jewish group, and in other ethnic or racial groups, such as Caucasians, Asians, African-Americans and Hispanics. For example, either or both of these polymorphisms can be detected using any known technique, including but not limited to, probe hybridization under stringent conditions, wherein the probe sequence is perfectly complementary to a sequence including the polymorphic site, a PCR amplification coupled to sequence analysis, restriction fragment length polymorphism, and the like. In view of the polymorphisms disclosed herein, moreover, it is contemplated that these relative hotspots might yield other forms of localized nucleic acid sequence changes, such as deletions of 1-5 or 1-10 nucleotides, insertions of 1-5 or 1-10 nucleotides and substitutions at position 27134 of SEQ ID NO:1 wherein any conventional nucleotide replaces the wild type T/U and/or the wild type AT at positions 27706-27707 of SEQ ID NO:1.

[0009] Particular aspects of the disclosure are described in the following enumerated paragraphs.

[0010] 1. A method of identifying a human subject as a carrier of a *SMN1* duplication allele comprising:

- (a) obtaining a nucleic acid sample from a human subject;
- (b) screening the nucleic acid by determining
 - (i) the identity of the nucleotide corresponding to position 11678 of SEQ ID NO:1; or
 - (ii) the identity of the nucleotide corresponding to position 15774 of SEQ ID NO:1; or
 - (iii) the identity of the nucleotide corresponding to position 22804 of SEQ ID NO:1; or
 - (iv) the identity of the nucleotide corresponding to position 26190 of SEQ ID NO:1; or
 - (v) the identity of the nucleotide corresponding to position 27134 of SEQ ID NO:1; or
 - (vi) whether the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 has been deleted; and
- (c) identifying the human subject as a carrier of a *SMN1* duplication allele if
 - (i) the nucleotide corresponding to position 11678 of SEQ ID NO:1 is not G or there is a nucleotide inserted between nucleotides corresponding to positions 11678 and 11679 of SEQ ID NO:1;
 - (ii) the nucleotide corresponding to position 15774 of SEQ ID NO:1 is not G or there is a nucleotide inserted between nucleotides corresponding to positions 15774 and 15775 of SEQ ID NO:1;

- (iii) the nucleotide corresponding to position 22804 of SEQ ID NO:1 is not G or there is a nucleotide inserted between nucleotides corresponding to positions 22804 and 22805 of SEQ ID NO:1;
- (iv) the nucleotide corresponding to position 26190 of SEQ ID NO:1 is not A or there is a nucleotide inserted between nucleotides corresponding to positions 26190 and 26191 of SEQ ID NO:1;
- (v) the nucleotide corresponding to position 27134 of SEQ ID NO:1 is not a T or there is at least one nucleotide inserted between nucleotides corresponding to positions 27134 and 27125 of SEQ ID NO:1; or
- (vi) the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 is deleted.

[0011] 2. The method according to paragraph 1 wherein the human subject is identified as a carrier of a *SMN1* duplication allele based on the presence of a nucleotide structure selected from the group consisting of a T corresponding to the nucleotide at position 11678 or 11679 of SEQ ID NO:1, an A corresponding to the nucleotide at position 15774 or 15775 of SEQ ID NO:1, an A corresponding to position 22804 or 22805 of SEQ ID NO:1, a G corresponding to the nucleotide at position 26190 or 26191 of SEQ ID NO:1, a G corresponding to the nucleotide at position 27134 or 27135 of SEQ ID NO:1 and an AT deletion corresponding to the nucleotides at positions 27706-27707 of SEQ ID NO:1.

[0012] 3. A method of identifying a human Spinal Muscular Atrophy (SMA) silent (2+0) carrier comprising:

- (a) screening nucleic acid from a human subject by determining
 - (i) the identity of the nucleotide corresponding to position 27134 of SEQ ID NO:1, or
 - (ii) whether the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 has been deleted; and
- (b) identifying an individual as a silent (2+0) carrier of SMA if the nucleotide corresponding to position 27134 of SEQ ID NO:1 is not a T or the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 is deleted, or both.

[0013] 4. The method according to paragraph 3 further comprising obtaining a sample of nucleic acid from the human subject.

[0014] 5. The method according to paragraph 3 or 4 further comprising providing genetic counseling to the human subject based on the results of the method of identifying a human Spinal Muscular Atrophy (SMA) silent (2+0) carrier.

[0015] 6. A method of identifying a human Spinal Muscular Atrophy (SMA) silent (2+0) carrier comprising:

- (a) screening nucleic acid from a human subject by determining the identity of the nucleotide corresponding to position 27134 of SEQ ID NO:1;
- (b) further determining if the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 has been deleted; and
- (c) identifying an individual as a silent (2+0) carrier of SMA if the nucleotide corresponding to position 27134 of SEQ ID NO:1 is G and the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 is deleted.

[0016] 7. The method according to paragraph 6 further comprising obtaining a sample of nucleic acid from the human subject.

[0017] 8. The method according to paragraph 6 or 7 further comprising providing genetic counseling to the human subject based on the results of the method of identifying a human Spinal Muscular Atrophy (SMA) silent (2+0) carrier.

[0018] 9. The method according to paragraph 6 or 7 wherein the identity of the nucleotide corresponding to position 27134 of SEQ ID NO:1, and the identities of the nucleotides corresponding to positions 27706 and 27707 of SEQ ID NO:1 are determined by restriction fragment length polymorphism.

[0019] 10. The method according to paragraph 6 or 7 wherein a G is identified at the position corresponding to position 27134 of SEQ ID NO:1 by restriction fragment length polymorphism using restriction endonuclease HpyCH4III.

[0020] 11. A method of identifying a human Spinal Muscular Atrophy (SMA) silent carrier comprising screening the nucleic acid from a human subject for a *SMN1* (SEQ ID

NO:1) haplotype that correlates with *SMN1* deletion and increased occurrence of SMA in a human population, wherein the SMA haplotype comprises:

- (i) a polymorphism selected from the group consisting of D5S681 having allele 2, D5S435 having allele 1 or allele 5, MS1 having allele 4, and D5S610 having allele 5, wherein the presence of the *SMN1* haplotype in the nucleic acid identifies the subject as having an *SMN1* deletion and being an SMA silent carrier, and wherein the absence of the *SMN1* haplotype in the nucleic acid identifies the subject as not being an SMA silent carrier.

[0021] 12. The method according to paragraph 11 further comprising obtaining a sample of nucleic acid from the human subject.

[0022] 13. The method according to paragraph 11 or 12 wherein the silent carrier is a silent (2+0) carrier.

[0023] 14. The method according to paragraph 13 further comprising providing genetic counseling to the human subject based on the results of the method of identifying a human Spinal Muscular Atrophy (SMA) silent carrier.

[0024] 15. The method according to paragraph 13 wherein the SMA haplotype for D5S681-D5S435-MS1-D5S610 is 2-5-4-5.

[0025] 16. The method according to paragraph 13 wherein the SMA haplotype for D5S681-D5S435-MS1-D5S610 is 2-1-4-5.

[0026] 17. A method of identifying a human Spinal Muscular Atrophy (SMA) carrier comprising: screening the nucleic acid from the human subject for a *SMN1* (SEQ ID NO:1) haplotype that correlates with *SMN1* deletion and increased occurrence of SMA in a human population, wherein the SMA haplotype comprises polymorphism MS1 having allele 6.

[0027] 18. The method according to paragraph 17 further comprising obtaining a sample of nucleic acid from the human subject.

[0028] 19. A method of identifying a human Spinal Muscular Atrophy (SMA) carrier comprising: screening the nucleic acid from the human subject for a *SMN1* (SEQ ID NO:1) haplotype that correlates with *SMN1* duplication and increased occurrence of SMA in a human population, wherein the SMA haplotype is selected from the group consisting of

- (i) polymorphism D5S681 having allele 2, D5S435 having allele 5, MS1 having allele 4, and D5S610 having allele 5;
- (ii) polymorphism D5S681 having allele 2; D5S435 having allele 5, MS1 having allele 9, and D5S610 having allele 4;
- (iii) polymorphism D5S681 having allele 2, D5S435 having allele 5, MS1 having allele 6; and
- (iv) polymorphism D5S681 having allele 2, D5S435 having allele 5, MS1 having allele 6, and D5S610 having allele 10,

wherein the presence of the *SMN1* haplotype in the nucleic acid identifies the subject as having an *SMN1* duplication and being an SMA carrier, and wherein the absence of the *SMN1* haplotype in the nucleic acid identifies the subject as not being an SMA carrier.

[0029] 20. The method according to paragraph 19 further comprising obtaining a sample of nucleic acid from the human subject.

[0030] 21. An isolated nucleic acid comprising a sequence selected from the group consisting of sequences of at least 17 nucleotides comprising a nucleotide corresponding to position 27134 of SEQ ID NO:1, sequences of at least 17 nucleotides comprising a dinucleotide corresponding to position 27705 and 27708 of SEQ ID NO:1, sequences of at least 17 nucleotides comprising the sequence set forth at positions 69-148 of SEQ ID NO:2, sequences of at least 17 nucleotides comprising the sequence set forth at positions 61-127 of SEQ ID NO:3, and sequences of at least 17 nucleotides comprising the sequence set forth at positions 61-216 of SEQ ID NO:4.

[0031] 22. The isolated nucleic acid according to paragraph 21 wherein the nucleic acid is about 10 to 200 nucleotides in length.

[0032] 23. A kit for diagnosing a Spinal Muscular Atrophy (SMA) carrier comprising the isolated nucleic acid according to paragraph 21 and a protocol for the use thereof in diagnosis of a SMA carrier.

[0033] Other features and advantages of the present disclosure will become apparent from the following detailed description, including the drawing. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments, are provided for illustration only, because various changes and modifications

within the spirit and scope of the invention will become apparent to those skilled in the art from the detailed description.

BRIEF DESCRIPTION OF THE DRAWING

[0034] Figure 1. A schematic of *SMN1* and *SMN2* alleles on chromosome 5q13 is provided. (A) Wild-type with one copy of *SMN1* and *SMN2* on each chromosome 5 (1+1) (B) SMA carrier with one copy of *SMN1* on one chromosome 5 and loss of *SMN1* on the other chromosome (1+0). (C) Duplication with two copies of *SMN1* one on one chromosome 5 and one copy on the other chromosome (2+1) (D) SMA silent carrier with two copies of *SMN1* on one chromosome 5 and loss of *SMN1* on the other chromosome (2+0).

[0035] Figure 2. A schematic of SMA Locus and Flanking Microsatellite Markers at 5q13.2 are shown. At the top of the figure are microsatellite markers used for the analysis of AJ individuals with deletions or duplications of *SMN1* (genomic coordinates obtained from the UCSC Genome Browser, NCBI36/HG18). The SMA locus, including the low copy repeats with *SMN2* in the proximal copy and *SMN1* in the distal copy, are shown below from genomic coordinates 68.9 Mb to 70.36 Mb on 5q13.2. The orientation of each of the genes is indicated by the direction of the arrows and was taken from genomic contig NT_006713.14.

[0036] Figure 3. The g.27134T>G and g.27706_27707delAT polymorphisms in *SMN1* (Genbank Acc. No. NG_008691.1). (A) Partial genomic sequence of *SMN1*. Exons are shaded. Primer sequences used for amplification and sequencing are denoted by *. Polymorphisms are denoted by dbSNP designation or by genomic coordinates and are indicated above the sequence and denoted by asterisks above (forward primers) or below (reverse primers). (B) Sequence traces that correspond to (a) g.27134T homozygote, (c) g.27134T/G heterozygote, (e) g.27134G homozygote, and g.27706_27707delAT (b) wild-type, (d) g.27706_27707delAT heterozygote, and (f) g.27706_27707delAT homozygote. (C) Gain of an hpyCH4III restriction site created in g.27134T>G. Lanes 10, 11 and 12 depict results with g.27134T (371 bp), lanes 2-7 depict results with individuals heterozygous for g.27134T/G (371bp, 218 bp and 153 bp) and lanes 1 and 9 show results for individuals homozygous for g.27134G (218 bp and 153 bp). M is the marker lane with a 50 bp ladder shown.

DETAILED DESCRIPTION

[0037] Spinal Muscular Atrophy (SMA) is the most common lethal inherited disease in children. Approximately 94% of affected human individuals show a bi-allelic loss of *SMN1* exon 7, and such losses are known to be the result of intragenic point mutations altering

splicing patterns of *SMN1* mRNA, intragenic deletions eliminating at least part of exon 7 as well as loss of the *SMN1* coding region via deletion. Disposed on human chromosome 5 is *SMN1* (SEQ ID NO:1) and *SMN2* (differing from the sequence of *SMN1* set forth in SEQ ID NO:1 by C>T at position 27006 of SEQ ID NO:1). These two genes are highly homologous, with a telomeric relative location for *SMN1* and a centromeric relative location for *SMN2*. The *SMN2* gene sequence differs from the *SMN1* sequence in exon 7 due to a point mutation, and this difference alone is sufficient to result in SMA. Given the relative proximity of *SMN1* and *SMN2*, and given the presence of repeats in this localized region of chromosome 5 (*i.e.*, 5q13.2), the genetics of SMA become complicated by the possibility of gene conversion as well as duplications and/or deletions. Although current methodologies can identify individual mutations in terms of sequence variation and can even provide a measure of *SMN1* copy number, these methods cannot discriminate beyond copy number to identify SMA carriers having, for example, excessive copies of the *SMN1* gene on one chromosome and no copies of *SMN1* on the other chromosome. In particular examples, the wild-type copy number of two for *SMN1* is present in the genome of SMA carriers because of an imbalance of *SMN1* on the chromosome 5 homologs (*i.e.*, both copies of *SMN1* on one member of the chromosome 5 pair). The disclosure advances efforts to identify SMA carriers by providing methods of identifying silent carriers of Spinal Muscular Atrophy (SMA) that contain a chromosomally unbalanced load of *SMN1*, such as the silent (2+0) carriers of SMA.

[0038] To aid in understanding the detailed description of the compositions and methods according to the disclosure, a few express definitions are provided to facilitate an unambiguous disclosure of the various aspects of the disclosure.

[0039] A “polymorphism” is a difference in DNA or RNA sequence among individuals, groups, or populations that gives rise to different alleles. The alleles may be alleles of a gene encoding a gene product, such as *SMN1*, and the polymorphism may involve a sequence change (relative to wild type sequence) in the coding region, in the transcribed but untranslated region associated with a gene, in the expression control region of a gene, in the proximal nucleic acid environment of a gene or located at some distance from the gene. Typically, polymorphisms of interest in identifying SMA carriers will be genetically linked to the *SMN1* gene. Exemplary polymorphisms include substitutions of one or more (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) nucleotides, deletions of a polynucleotide region comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 150, 200, 250, 500, 1,000, or more nucleotides, and insertions of nucleotides into a polynucleotide region wherein the insertion is of a length defined above in the context of addressing deletions.

[0040] A “SNP” is a single-nucleotide polymorphism, or single nucleotide difference in the nucleic acid sequence relative to the wild type sequence.

[0041] As used herein, a “haplotype” is a partial genotype of at least one determinant containing at least one polymorphism, such as single-nucleotide polymorphism (SNP), a deletion or an insertion, on a chromosome. For haplotypes comprising more than one polymorphism, the individual polymorphisms exhibit statistically significant linkage disequilibrium. Exemplary polymorphisms are single- or multiple-nucleotide substitutions, insertions or deletions and each polymorphism may be localized to a determinant that is a gene recognized in the art, such as *SMN1*, a new gene, or an extragenic region of a chromosome.

[0042] “Copy number” refers to the number of physical copies of a genetic determinant, such as a gene, or region of the genome of an organism.

[0043] A “carrier” or genetic “carrier” is an individual containing at least one copy of an allele of a genetic determinant that is involved in elaborating a given phenotype, such as SMA, provided that the individual containing the copy or copies of the determinant does not exhibit the phenotype.

[0044] A “silent carrier” is a carrier that cannot be detected using a copy number-based diagnostic technique conventional in the art.

[0045] SMA is a pan-ethnic disease, and the frequency of single *SMN1* exon 7 copy number in various ethnic groups are known. In particular, the single-copy exon 7 genotype frequencies range from 2.7% (1 in 37) in the Caucasian population, 2.2% (1 in 46) in the Ashkenazi Jewish population, 1.8% (1 in 56) in the Asian population, 1.1% (1 in 92) in the African American population, down to 0.8% (1 in 125) in the Hispanic population. Hendrickson et al., J. Med. Genet. 46:641-644 (2009), incorporated herein by reference in its entirety. Thus, the disclosure is not limited to the exemplified application of identifying carriers in the Ashkenazi Jewish population, but rather extends to identification of SMA carriers such as silent (2+0) SMA carriers in any ethnic population, including the populations characterized herein.

[0046] Identification of deletion/duplication founder alleles provides a general approach to identify silent carriers and improve carrier detection in various ethnic/racial groups, as exemplified by the data disclosed herein that relates to the Ashkenazi Jewish (AJ) population.. The data identify founder alleles in the Ashkenazi Jewish (AJ) population. Carrier screening of 692 healthy AJ adults was conducted by Multiplex Ligation-dependent

Probe Analysis (MLPA), which identified 1 in 46 (2.2%) deletion (1+0) carriers, and 1 in 7 duplication (2+1) individuals. These data indicate that carrier detectability using only *SMN1* gene dosage in this population is around 90%. Microsatellite analyses of markers flanking the SMA locus identified two deletion haplotypes and one major duplication founder haplotype. Importantly, two polymorphisms tightly linked to *SMN1*, *i.e.*, g.27134T>G in intron 7 (position 27006 of SEQ ID NO:1, *see* Genbank Acc. No. NG_008691.1) and g.27706_27707delAT in exon 8 (positions 27706-27707 of SEQ ID NO:1, *see* Genbank Acc. No. NG_008691.1), were detected on the major AJ duplication allele, but not in 351 AJ individuals with two *SMN1* copies, making the haplotype highly specific for duplication alleles and effectively increasing the accuracy of carrier detection. It is expected that the greatest diagnostic accuracy will be obtained when assaying for both of the above-identified polymorphisms, but the disclosure comprehends diagnostic methods assaying for either polymorphism alone or in conjunction with one or more other polymorphisms. Identification of specific markers for duplication or deletion alleles is expected to improve carrier detectability and decrease the current residual risk due to the inability to detect silent (2+0) carriers in any population.

[0047] In 2008, The American College of Medical Genetics (ACMG) recommended offering carrier screening for SMA to all couples regardless of race or ethnicity. Preconceptual genetic counseling for SMA is complicated by the nature of the mutations, which involve copy number loss of *SMN1* by deletion or gene conversion with the highly homologous, centromeric copy, *SMN2*, as well as the high rate of *de novo* mutations at the locus. Since most mutations involve copy number loss, carrier detectability is limited by the false negative rate that varies from 4% to 27%, based on ethnicity, due to the inability to detect silent (2+0) carriers with two copies of *SMN1* on one chromosome and none on the other.

[0048] To overcome this limitation, and to increase the detectability of SMA carriers in the AJ population, founder alleles with unique polymorphisms were sought to distinguish the silent (2+0) carriers from wild-type (1+1) individuals. Screening of 692 AJ individuals by MLPA established the frequencies of normal, deleted, and duplicated *SMN1* alleles and identified deletion (1+0) and duplication (2+1) individuals to seek unique deletion- and/or duplication-allele haplotypes surrounding the *SMN1* locus. Haplotype studies of the AJ population were aided by the availability of founder mutations identified in shared haplotype blocks, which facilitated carrier detection for recessive disorders prevalent among AJ individuals.

[0049] To search for potential founder alleles, microsatellite analyses were performed with markers that flanked the *SMN1* locus and marker trait association studies were performed with four of the most tightly linked markers (D5S681-D5S435-MS1-D5S610; SEQ ID NOS:5, 6, 2, and 7, respectively), as described in the following examples. D5S681 was reported in Morrison et al., Hum Mol Genet. 2(10):1753 (1993); Soares et al., Genomics. 15(2):365-71 (1993) reported D5S435; for D5S610, see Velasco et al., Eur J Hum Genet. 3(2):96-101 (1995). MS1 is a newly identified microsatellite marker, as are MS2 (SEQ ID NO:3) and MS3 (SEQ ID NO:4). Microsatellites are also called simple sequence repeats, and constitute a class of genetic polymorphisms commonly used for mapping, linkage analysis and the tracing of inheritance patterns. Typically, microsatellites are tandemly repeated sequences. The number of times the unit is repeated in a given microsatellite can be highly variable, a characteristic that makes them useful as genetic markers.

[0050] A comparison of the haplotype frequencies in deletion (1+0) and duplication (2+1) individuals to those in AJ non-carrier controls identified two deletion- and three duplication-specific haplotypes with p values ≤ 0.001 (Table 3). Of these haplotypes, one was present in approximately 46.4% (32/69) of the duplication alleles tested, but was absent in all 78 AJ control individuals making it highly specific.

[0051] The following examples illustrate embodiments of the disclosure. Example 1 discloses materials and methods used in the studies disclosed herein. Example 2 provides the results of genetic screening of the AJ population for deletions and insertions associated with *SMN1*, along with a determination of the frequency of silent (2+0) carriers of *SMN1* mutations associated with SMA. Disclosed in Example 3 is the results of genetic analyses of the deletion (1+0) and duplication (2+1) mutation groups in the AJ population. Example 4 discloses haplotype reconstruction in individuals with deletions (1+0) or duplications (2+1). Example 5 describes the sequencing of the major duplication allele in the AJ population, identification of an additional two polymorphisms, and characterization of these polymorphisms in several human populations. Example 6 provides the results of RFLP analyses based on the g.27134T>G SNP from *SMN1* exon 7. Example 7 discloses data and analyses thereof that establish the improved identification of SMA carriers, e.g., SMA silent (2+0) carriers, in various racial and ethnic groups.

Example 1

Specimen Collection and DNA Isolation

[0052] Genomic DNA was obtained with informed consent from the peripheral blood specimens of anonymous individuals of Ashkenazi Jewish, African American, Hispanic and

European ancestry living in the Greater New York Metropolitan area. DNA was isolated using the Puregene[®] DNA Purification kit (Qiagen, Valencia, CA) according to the manufacturer's instructions.

Multiplex Ligation-Dependent Probe Amplification (MLPA)

[0053] A total of 200 ng of genomic DNA from each specimen was used for multiplex ligation-dependent probe analysis (MLPA) analysis with the Salsa MLPA SMA P021 kit (MRC-Holland, Amsterdam, Netherlands) according to the manufacturer's instructions. The MLPA PCR products were separated by capillary electrophoresis using the ABI-3130XL Genetic Analyzer (Applied Biosystems Inc, Foster City, CA). The results were imported as html files into GeneMarker software (Softgenetics, LLC., State College, PA) for MLPA data analysis using population normalization with the MLPA ratio as the analysis method and peak height as the quantification method. The peak-heights for each of the 38 MLPA probe ligation products in the test sample were compared to the mean peak-heights of at least three normal controls, and the test/wild-type peak-height ratios for each probe ligation product were determined. The peak-height ratios less than 0.8 and/or greater than 1.3 would indicate the loss or gain of the copy number for each probe, respectively.

Microsatellite Analysis

[0054] Microsatellite analysis was performed with previously described markers, D5S681, D5S435, D5S610 that flank the *SMN1* and *SMN2* genes on 5q13 (Scheffer et al., Eur. J. Hum. Genet. 9:484-494 (2001)). Three novel markers, MS1 (SEQ ID NO:2), MS2 (SEQ ID NO:3) and MS3 (SEQ ID NO:4), were based on publicly available human genomic sequence using the UCSC genome database. The following primer pairs were designed to amplify MS1: MS1F (5'-TAAATGTCAAATTTATGTATGGG-3'; SEQ ID NO:12) and MS1R (5'-CTGTGATTGAAAC AAAGACACCT-3'; SEQ ID NO:13); MS2: MS2F (5'-TCCATGGATACGGAGAGCTG-3'; SEQ ID NO:16) and MS2R (5'-TGATGGCACCACAACCATGC-3'; SEQ ID NO:17) and MS3: MS3F (5'-TCAGGACCCTCCTCATCATC-3'; SEQ ID NO:18) and MS3R (5'-CGGGATCTCTTCTCCACAGA-3'; SEQ ID NO:19). All forward primers were 6-carboxy Fluorescein-labeled or 6-FAM-labeled (Invitrogen). Each primer pair was amplified separately in a total volume of 25 µl using 200 ng of genome DNA, 1X PCR buffer (Invitrogen), 1.5 mM MgCl₂, 0.3 µM forward and reverse primers and 0.3 µl of platinum Taq polymerase (Invitrogen). The amplifications were initiated with 10 minutes at 94°C, followed by 35 cycles of 30 seconds at 94°C, 30 seconds at 58°C and 30 seconds at 72°C, with a final extension for 10 minutes at 72°C. PCR products were diluted 1:10 and run on an

ABI 3730 DNA analyzer (Applied Biosystems) and the results were analyzed with GeneMapper software v3.7 (Applied Biosystems).

LongRange PCR and Sequencing

[0055] Long-range PCR spanning the fragment between exon 7 and exon 8 (1025 bp) of the *SMN1* gene in AJ individuals identified with three copies of the *SMN1* gene by MLPA was performed using LongRange PCR reagents (Qiagen) with primers *SMN1*-E7F (5'-TCCTTTATTTTCCTTACAGGGTTTC-3'; SEQ ID NO.:20) and *SMN1*-E8R (5'-CAATGAACAG CCATGTCCAC-3'; SEQ ID NO.:21). Primer *SMN1*-E7F was designed to preferentially amplify *SMN1* by the inclusion of c.840C of Exon 7 at the 3' end. PCR amplification was performed according to the manufacturer's instructions in 50 µl of reaction mixture containing 500 mM dNTP, 1X PCR buffer, 2.5 mM MgCl₂, 2 units (U) of long-range enzyme, 400 nM of each primer, and 250 ng of purified DNA. After 3 minutes of initial denaturation at 93°C, 35 cycles of amplification were performed as follows: 15 seconds at 93°C, 30 seconds at 59°C, and 1.5 minutes at 68°C.

[0056] The PCR amplicons were purified using QIAquick PCR Purification Kit (Qiagen). The fragment was sequenced using the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) and the products were run on the ABI 3730 automated sequencer (Applied Biosystems). The same primers used for PCR amplification were also used in the sequencing reactions. Additional primers *SMN1*-I7F2 (5'-AGGAAATGCTGGCATAGAGC-3'; SEQ ID NO.:22) and *SMN*-E8Ra (5'-CTGCTCTATGCCAGCATTTC-3'; SEQ ID NO.:23) were designed to confirm sequencing results in both directions. Sequence traces were analyzed using both Mutation Surveyor software v3.01 (SoftGenetics) and SequencherTM version 4.8 (Gene Codes, Ann Arbor, MI) and by visual inspection.

TA Cloning

[0057] LongRange PCR fragments (1025 bp) from five AJ major duplication (2+1) alleles (see below) were subjected to TA-cloning. In brief, amplicons were gel-purified (Qiagen) and cloned into the PCR vector from the TOPO TA Cloning® Kit (Invitrogen) according to the manufacturer's instructions. Twelve clones from each of these duplication carrier individuals were isolated for PCR screening. Clones with the expected insert size (1025 bp) were subsequently sub-cultured, and plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen) and sequenced as described above using M13 primers (Fisher Scientific, Pittsburgh, PA).

Restriction Fragment Length Polymorphism Analysis (RFLP)

[0058] PCR reactions were performed to amplify a 372 bp fragment containing the g.27134T>G single nucleotide alteration with the primers *SMN1*-E7F (above) and *SMN1*-I7R1 (5' AAAAGTCTGCTGGTCTGCCTAC-3'; SEQ ID NO.:24). The amplification was carried out in a 25 µl reaction mixture with 400 nM of each primer, 200 µM dNTP, 1.5 mM MgCl₂ and 1 U of Platinum[®]Taq DNA Polymerase (Invitrogen, Carlsbad, CA). Cycling conditions were as follows: 94°C for 5 minutes, followed by 30 cycles of 94°C for 30 seconds, 59°C for 30 seconds, 72°C for 1 minutes and a final extension of 72°C for 10 minutes. The g.27134T>G single nucleotide alteration resulted in a gain of a restriction site for enzyme *Hyp*CH4III. Following amplification, 5 µl of PCR product were digested with 4 U of *Hyp*CH4III (New England Biolabs, Ipswich, MA) at 37°C for 4 hours and were resolved by either 6% polyacrylamide or 1.5% agarose gel electrophoresis, visualized with a UV transilluminator and photographed.

Statistical Analysis

[0059] Genotype frequencies for each microsatellite marker were tested for Hardy-Weinberg equilibrium (HWE) using the χ^2 test. Haplotypes of the markers, D5S681-D5S435-MS1-D5S610, were estimated for AJ controls and AJ *SMN1* deletion (1+0), duplication (2+1), and silent (2+0) carriers using SAS/Genetics version 9.2. Lengths of the various alleles for each microsatellite marker are provided in Table 8. Putative haplotypes were inferred from their frequency among all examined individuals and reconstructed based on the most likely combination of alleles. Differences in haplotype frequencies between AJ control cohort versus the AJ cohort of individuals with *SMN1* deletion or duplication were tested using the χ^2 test. P-values for overall and pair-wise comparisons were calculated for the four marker haplotypes with the threshold of 0.05 considered statistically significant. Improvement in the detection rate with the addition of new markers was estimated using the Bayesian approach.

Example 2

AJ Population Screening for *SMN1* Deletions and Duplications

[0060] To determine the frequency of *SMN1* deletion (1+0) and duplication (2+1) individuals in the AJ population, 692 individuals of 100% AJ ancestry (with four AJ grandparents) were screened by an *SMN1*- and *SMN2*-specific MLPA assay, which detects the copy number of 38 different genomic regions in a single reaction. Specifically, this assay determined the individual *SMN1* and *SMN2* exon 7 and exon 8 copy numbers for each of *SMN1* and *SMN2* as well as the combined *SMN1* and *SMN2* copy number of their respective exons 1, 4, 6 and 8. Fifteen *SMN1* deletion (1+0) carriers (1 in 46 or 2.1%) were identified,

which is similar to the frequency reported in other European populations. However, 99 *SMN1* duplication (2+1) individuals (1 in 7 or 14%) were identified, which is twice the frequency in Europeans. No individuals with four *SMN1* copies (i.e., two duplication alleles, 2+2) were identified, although the expected frequency of (2+2) individuals was 0.5% (Table 2). The observed frequencies did not deviate significantly from Hardy-Weinberg equilibrium (HWE) (exact $p > 0.05$) and are shown in Table 1 along with the expected frequencies.

Frequency of AJ Silent (2+0) Carriers

[0061] Using the frequency of deletions (1+0) (2.2%) and duplications (2+1) (14.3%), assuming Hardy-Weinberg Equilibrium for each of three alleles (*SMN1* = 0, 1 or 2 copies), silent (2+0) carriers were estimated to occur at a frequency of $2pr$, (where p is the frequency of 1 copy of the wild-type (WT) allele and r is the frequency of 0 copy or deletion alleles) which was 1 in 531 AJ individuals. Based on this calculation, the overall carrier frequency for SMA among AJ individuals was approximately 1 in 41.1 (2.4%), which includes carriers identified by MLPA with loss of one *SMN1* (1+0) copy, silent (2+0) carriers not detected by dosage-sensitive, location-insensitive methods, and the small fraction of individuals that carry intragenic mutations (0.00047). These values assume that r^2 is 0.00014 (Pearn, J Med Genet 15:409-413 (1978)) and that 2% of affected individuals have an intragenic mutation, as has been previously reported (Wirth et al., 1999). The carrier detection rate in the AJ population by MLPA assays was about 90% based on this analysis with about 8% silent (2+0) carriers and about 2% intragenic mutation carriers. Therefore, the residual risk of being a carrier after a negative screening result was 1/402 (Table 2).

Example 3

Mutation Groups Among AJ with Deletions (1+0) or Duplications (2+1)

[0062] A total of 23 AJ deletion (1+0) carriers were identified for these studies, including 15 from the above-described population screening and eight that were subsequently identified by MLPA. Four major mutation groups were encountered: 1del) *SMN1* deletion only; 2del) *SMN1* and *SMN2* deletions; 3del) *SMN1* exon 7 and 8 deletion and *SMN2* deletion, or an *SMN1* deletion and *SMN2* exon 7 and 8 deletion, and 4del) *SMN1* exon 7 and 8 deletion only (Table 5). Interestingly, 32% of AJ individuals with two copies of *SMN1* carried a deletion of *SMN2* indicating that loss of *SMN2* is a common polymorphism in the AJ population and that there may be significant overlap between groups 1del and 2del.

[0063] For the gain-of-copy-number duplication group, a total of 72 AJ duplication (2+1) individuals were analyzed by MLPA, including three with four *SMN1* copies (2+2), and three

major groups were identified: 1dup) three *SMN1* copies with no *SMN2* copy; 2dup) three *SMN1* copies with one *SMN2* copy, and 3dup) three *SMN1* copies with two *SMN2* copies (Table 6). Of the 72 duplication (2+1) individuals examined, 26 (36.1%) had no *SMN2* copies, and an additional 32 (44.4%) had one *SMN2* copy, indicating that the duplication in these 58 individuals arose by a gene conversion event replacing *SMN2* sequences with *SMN1* sequences. As noted herein, there may be significant overlap between groups 1dup and 2dup due to the high incidence of *SMN2* deletions. Of the *SMN1* duplication (2+1) individuals, 11 (15.3%) had two *SMN2* copies and, without wishing to be bound by theory, it is expected that those duplications arose by a different mechanism, such as non-allelic homologous recombination (NAHR) or a gene conversion event in individuals with an *SMN2* duplication. *SMN2* duplications are not uncommon, and were observed by MLPA in the AJ population at a frequency of 1 in 38 (2.6%).

Example 4

Haplotype Reconstruction in AJ with Deletions (1+0) or Duplications (2+1)

[0064] To determine whether founder haplotypes were present for *SMN1* deletion (1+0) alleles, 20 carriers from groups 1del, 2del and 3del (see Example 3) that were identified by MLPA were genotyped with six microsatellite markers that span 3.9 Mb on 5q13.2 surrounding the *SMN1* locus. They included D5S681 (Genbank NT_006713.15; SEQ ID NO:5), D5S435 (Genbank NT_006713.15; SEQ ID NO:6) and D5S610 (NT_006713.15; SEQ ID NO:7) that flank the *SMN1* and *SMN2* genes (Scheffer et al., 2001), and three novel markers, MS1 (Genbank NT006713.15; SEQ ID NO:2), MS2 (Genbank NT006713.15; SEQ ID NO:3) and MS3 (Genbank NT006713.15; SEQ ID NO:4) that were created from the publicly available human genomic reference sequence using the UCSC genome database browser. A schematic representation of this region of the human genome is present in Fig. 2. Using these markers, 78 AJ control individuals with two (1+1) *SMN1* copies were genotyped. Based on these results, the four consecutive loci chosen for marker association studies were D5S681-D5S435-MS1-D5S610. Their most likely marker configurations are shown in Table 3. Haplotypes for these markers were constructed for the 20 *SMN1* deletion (1+0) carriers, including all individuals from groups 1del, 2del, and 3del (Table 5). Group 4del was excluded because all three individuals only had loss of exon 7 and 8 sequences.

[0065] Two putative founder haplotypes were significantly enriched ($p \leq 0.001$) in the deletion (1+0) carriers as compared to controls. The most common allele combination among the 20 (1+0) carriers examined, 2-5-4-5, was assigned in 22.5% of carriers and was possible in eight additional (40%) carriers. The 2-5-4-5 haplotype also was assigned on only one

chromosome of 78 controls (0.6%) indicating that this haplotype was not specific to deletion alleles. The other carrier-specific allele haplotype, 2-1-4-5, was not assigned in any control individuals; however, two control individuals could potentially carry it. One of the group 4del carriers with the *SMN1* exon 7 and 8 deletion only could possibly have the 2-5-4-5 haplotype indicating that it may not be specific to *SMN1* whole gene copy loss alleles or that it may be the haplotype of the wild-type allele in carrier no. 23 (Table 5).

[0066] To analyze the duplication (2+1) individuals for unique shared haplotypes, the six microsatellite markers were genotyped in 69 of the 99 AJ individuals identified with three copies of *SMN1* and in three additional AJ individuals that were subsequently identified with four copies of *SMN1* (Table 6). The same four marker genotypes, D5S681-D5S435-MS1-D5S610, were used in haplotype reconstruction for 42 duplication (2+1) individuals from groups 1dup and 2dup. Of these, 26 individuals had three *SMN1* copies and no *SMN2* copies, and 16 had three *SMN1* copies and one *SMN2* copy; 31 of these duplication (2+1) individuals chosen for the analysis had MS1 allele 6 (Table 8). A separate analysis was performed for the nine group 3dup individuals that carried three *SMN1* copies and two *SMN2* copies, as they presumably represented a separate duplication group.

[0067] The most likely marker haplotypes for the three groups are shown in Table 3. One major duplication marker haplotype, 2-5-6-10, was significantly enriched in Groups 1dup and 2dup (19%). This haplotype was completely absent from the control population and was highly specific for the duplication carriers ($p < 0.0001$). Of the 58 group 1dup and 2dup individuals, 23 (39.7%) had the full four-marker haplotype and eight (13.8%) had the 2-5-6 partial haplotype. In addition, one of the group 3dup individuals (no. 59, Table 6) had the 2-5-6-10 haplotype indicating that this individual most likely carried an *SMN2* duplication on the wild-type allele. The major duplication haplotype was also positive in two of the individuals with four *SMN1* copies, one of which was homozygous for the 5-6-10 haplotype (no. 70, Table 6). There were two haplotypes that were significantly enriched in group 3dup individuals, 2-5-4-5 and 2-5-9-4, as compared to controls ($p < 0.0001$). The 2-5-4-5 haplotype was also the major deletion haplotype, which was also present in a small fraction of the controls.

Example 5

Sequencing of the Major AJ Duplication Allele

[0068] Several putative AJ founder haplotypes were computationally inferred in the *SMN1* deletion and duplication alleles. Only the major duplication haplotype, 2-5-6-10, or partial haplotype, 2-5-6, was present in the majority (51.7%) of AJ duplication (2+1) individuals

with 0 or 1 copy of *SMN2* genotyped (Table 6) and was absent from the AJ control population. The major duplication allele was further scrutinized by sequencing. An *SMN1*-specific 1025 bp fragment between exons 7 and 8 was amplified from five different AJ individuals with the major duplication allele using primer *SMN1*-E7F, which includes c.840C at the 3' end, and primer *SMN1*-E8R (Fig. 3A).

[0069] Two sequence variants in the *SMN1* gene were identified, g.27134T>G in intron 7 and g.27706_27707delAT in exon 8 (*see* Genbank Acc. No. NG_008691.1; Fig. 3B). Neither of these polymorphisms has been previously reported in the NCBI ENTREZ (SNP) database. Further sequence analysis indicated that both polymorphisms were present in all individuals that carried the major duplication haplotype (2-5-6-10) or partial haplotype (2-5-6), and were absent from all other duplication individuals. Thus, each of these polymorphic sequences alone is diagnostic of duplication, as is the combination of both polymorphism sequences. Using a combination of LongRange PCR and TA-cloning analysis, the presence of the two polymorphisms within the same clone in the 12 tested duplication (2+1) individuals indicated that these polymorphisms resided in the same *SMN1* allele. No other sequence variants specific to the major duplication allele were identified within the 1025 bp that were analyzed.

[0070] Sequencing 1025 bp between *SMN1* intron 7 and exon 8 in individuals that carried the major duplication haplotype identified two tightly linked polymorphisms, *i.e.*, g.27134T>G in intron 7 and g.27706_27707delAT in exon 8. Neither of these polymorphisms had been described in the literature or in the NCBI ENTREZ (SNP) database as being found in or associated with either the *SMN1* or *SMN2* genes. The g.27134T>G transversion introduced an hpyCH4III restriction site that was present in most of the duplication (2+1) individuals, but not in any of the 315 healthy (1+1) AJ controls tested. Thus, an AJ individual with two copies of *SMN1* as determined by MLPA (or another dosage-sensitive method) who is positive for the g.27134T>G generated hpyCH4III cleavage site and its linked g.27706_27707delAT polymorphism, is highly likely to be an *SMN1* silent (2+0) carrier. Therefore, detection of g.27134T>G and/or g.27706_27707delAT would increase SMA carrier detection in the AJ population from 90% to 94% and reduce the residual risk after a negative carrier screening result from 1 in about 400 to 1 in about 700 (Table 2). Since 83.5% (Table 1) of AJ individuals will have a negative carrier screening result (having two copies of *SMN1*), this assay can be routinely performed to detect the silent (2+0) carriers among them, thereby increasing SMA carrier detectability. One of skill will recognize, moreover, that convenient RFLP analyses to detect SMA carriers can be conducted with any restriction endonuclease having a site dependent upon the presence or

absence of a marker allele spanning that site. Thus, HpyCH4III will distinguish wild type alleles from alleles harboring the g.27134T>G transversion, as will any other restriction endonuclease requiring either the “T” or the “G” at position 27134. All such embodiments of RFLP-based screening or diagnostic-based assays are contemplated in the present disclosure.

[0071] Interestingly, the two linked polymorphisms were also present at different frequencies among AA, AS and HI, but not in the EU population. In the AA, AS, and HI populations, the g.27134T>G and g.27706_27707delAT polymorphisms were more frequent in duplication (2+1) individuals than in those with 1 or 2 *SMN1* copies. For example, in the AS population, the linked polymorphisms were exclusive to individuals with *SMN1* duplications although they were only present in about 11.5% (3 of 26) of the duplication alleles examined. Testing for these polymorphisms would increase SMA carrier detection by approximately 0.7% in the AS population and would increase carrier detectability to about 93.3% (Table 2). For the AA and HI populations, the linked polymorphisms were not specific to the duplication alleles and were present in a minority of *SMN1* single-copy alleles. The AA population has the lowest rate of SMA carrier detection (71%) due to the high prevalence of duplication alleles. Both g.27134T>G and g.27706_27707delAT were present in the majority of AA duplication alleles (83.7%) indicating that they likely arose from an ancient allele that was involved in the original gene conversion/recombination event that led to the *SMN1* duplication, which is present in the AJ, AA, AS and HI populations. None of the (2+1) duplication individuals examined in the AA, HI or AS populations who had the g.27134T>G and g.27706_27707delAT polymorphisms had the marker MS1 allele 6, which was present in all AJ individuals with the major duplication haplotype. Examination of the extended *SMN1* SNP haplotype in individuals that carry g.27134T>G and g.27706_27707delAT is expected to permit estimation of the date of origin of the duplication allele and confirm whether it is of similar origin in other populations. Expansion of these studies to additional ethnic or demographic groups will be useful in finding specific haplotypes that demarcate the *SMN1* deletions or duplications and facilitate identification of silent (2+0) carriers that currently escape detection.

Example 6

RFLP Analysis for the g.27134T>G (Intron7) Polymorphism

[0072] Digestion of the PCR amplicons containing g.27134T>G polymorphism with hpyCH4III revealed a restriction site (Fig. 3C) gain attributable to the polymorphism. RFLP analysis of 315 AJ control individuals with two *SMN1* copies revealed that none had the hpyCH4III restriction site, indicating that the polymorphism was specific to the major AJ

duplication allele. Interestingly, individual no. 69 (Table 6), who was homozygous for the 2-5-6-10 allele combination (Table 6) with four *SMN1* copies and no *SMN2* copies, was heterozygous for the restriction site, yielding both undigested (391 bp) and digested (153 bp and 218 bp) fragments upon exposure to *hypCH4III*, consistent with the polymorphism being present in only one copy of the *SMN1* duplication. This finding also confirmed the heterozygosity observed for the g.27134T>G and g.27706_27707delAT polymorphisms among all AJ duplication (2+1) individuals that were sequenced. A total of 49 of the 99 duplication (2+1) individuals identified during the population screening carried the 27134T>G polymorphism as determined by RFLP analysis. Therefore, testing for this polymorphism in the AJ population would increase SMA carrier detection by approximately 4%, identifying about half of the silent (2+0) carriers, for a total detection rate of 94%. Moreover, the residual risk of being an SMA carrier after a negative screening result would decrease from 1 in about 400 to 1 in about 700.

[0073] To determine whether the g.27134T>G polymorphism was also present on *SMN1* duplication alleles in other ethnic groups, MLPA and RFLP analyses were performed on genomic DNAs from 276 African American (AA), 250 Asian (AS), 262 Hispanic (HI) and 137 European (EU) individuals. The results, summarized in Table 4, indicate that g.27134T>G was present on duplication alleles in all populations examined except for the EU individuals. In the AA population, the RFLP was observed in individuals with 1, 2, 3 or 4 copies of *SMN1* indicating that it is not specific to their duplication alleles. The majority (81%) of AA duplication (2+1) individuals, however, were heterozygous for g.27134T>G, while only 21% of individuals with two copies (1+1) of *SMN1* were positive for the polymorphism. Of note, one of the five AA deletion carriers was positive for the RFLP, presumably on the wild-type allele. Similarly, the RFLP was present in about half of HI duplication (2+1) individuals, but it also was present in approximately 5.5% of wild-type alleles. Conversely, in the AS population, the RFLP was not found in wild-type alleles, but was present only in about 14% of the duplication (2+1) individuals. Sequencing was performed on 142 AA, 2 AS and 5 HI individuals that were positive for the RFLP to confirm that both g.27134T>G and g.27706_27707delAT variants were present on the RFLP-positive alleles. In all cases examined, the two polymorphisms occurred together within the same *SMN1* allele, regardless of how many copies of *SMN1* were present.

Example 7

Improved Identification of SMA Carriers in Various Racial and Ethnic Groups

[0074] The promising results disclosed in the preceding examples established that the method of detecting chromosomally unbalanced *SMN1* loads, such as is found in silent (2+0) SMA carriers, can appreciably improve the identification of SMA carriers in ethnic and racial groups, and in particular, in the Ashkenazi Jewish (AJ) population. The data further showed that SMA carrier identification was improved in Asians, and gave some indication that the method had an even broader range of improving the detection of chromosomally unbalanced *SMN1* loads across various ethnic and racial groups. The data provided in Table 9 reveals the residual risks of being a carrier after screening positive and negative for g.27134T>G in intron 7 and g.27706_27707delAT in exon 8. Apparent from Table 9 is that the identification method disclosed herein improves the detection of chromosomally unbalanced *SMN1* loads, leading to increased detection of SMA carriers, such as silent (2+0) SMA carriers, in a variety of ethnic and racial groups, including AJ, Asians, African Americans and Hispanics.

[0075] The implications of these results for genetic counseling and family planning are significant. With the method disclosed herein, a more accurate assessment can be made of the modified risk of having a child with SMA if one partner is a known carrier and the other has two copies of *SMN1*, as disclosed in Table 10. The data show that adding the dosage-sensitive, location-sensitive method of identifying SMA carriers leads to a significant reduction in potentially devastating false negative outcomes, in which potential parents mistakenly believe they can reproduce without significant risk of SMA offspring. The data in Table 10 establish that the benefit of reducing the frequency of false negative outcomes attributable to the method disclosed herein spans ethnic and racial groups and is, therefore, of widespread value to human populations.

[0076] Beyond the polymorphisms disclosed in the examples above, additional polymorphisms are expected to be useful in assessing the residual risk of SMA due to the carrier status of one or both parents. As catalogued in Table 11, these novel polymorphisms can occur throughout the *SMN1* gene, including within introns 1, 2a, 6, 7 and 8. With these polymorphisms and the polymorphisms disclosed above, the risk of erring in identifying SMA carrier status is minimized, allowing potential parents, guided by genetic counselors, to be better informed in making family planning decisions.

Table 1. *SMN1* Carrier Screening of 692 Ashkenazi Jewish Individuals

SMN1 copy number	Genotype	n	%	Expected* (n)	Expected* (%)
------------------	----------	---	---	---------------	---------------

0	0+0	0	0	0.1	0.01
1	1+0	15	2	14	2
2	1+1	578	84	583	84
	2+0	-	-	1	0.2
3	1+2	99	14	91	13
4	2+2	0	0.0	4	0.5

*Expected frequencies were calculated using the observed wild-type (p^2 with 2 *SMN1* copies) and the observed carrier frequency ($2pq$ with 1 *SMN1* copy) and the equation $p^2+2pq+q^2+2qr+2pr+r^2=1$

Table 2 SMA Carrier Detection and Residual Risk with addition of *SMN1* g.27134T>G and g.27706_27707delAT

Ethnicity	Carrier Frequency	Detection Rate	Residual risk after negative carrier screen	Detection rate with addition of <i>SMN1</i> g.27134T>G and g.27706_27707delAT	Residual risk after negative result for <i>SMN1</i> g.27134T>G and g.27706_27707delAT
Ashkenazi Jewish	1 in 41.1	90%	1 in 402	94%	1 in 683
Asian	1 in 53*	92.6%*	1 in 628*	93.3%	1 in 765

* Taken from Hendrickson BC, Donohoe C, Akmaev VR, Sugarman EA, Labrousse P, Boguslavskiy L, Flynn K, Rohlf EM, Walker A, Allitto B, Sears C, Scholl T Differences in *SMN1* allele frequencies among ethnic groups within North America. *J. Med. Genet.* 2009;**46**:641-44.

Table 3 Haplotype Analysis for SMA Locus.

Marker	D5S681	D5S435	MS1	D5S610	Haplotype frequency in WT, % (N=78)	Haplotype frequency in Carriers, %	P-value
<i>SMN1</i>	2	1	4	4	0	7.5	1.0
Deletion	2	1	4	5	0	10.0	0.001
Carriers (N=20)	2	5	4	5	0.6	22.5	<.0001
	2	5	7	5	1.4	5.0	0.19
Overall P-value <.0001							
<i>SMN1</i>	2	3	7	5	0	4.7	0.01
Duplication	2	5	3	4	2.6	3.6	0.12
Groups 1dup & 2dup (N=42)	2	5	6	10	0	19.0	<.0001
	2	5	9	2	1.8	4.8	0.19
	4	3	4	2	0	3.4	0.03
	4	5	4	2	0.8	3.6	0.003
Overall P-value <.0001							
<i>SMN1</i>	2	4	9	5	0	5.6	0.035
Duplication	2	5	4	5	0.6	11.1	<.0001
Group 3dup	2	5	7	2	0	5.6	1.0

(N=9)	2	5	9	4	0	16.7	<.0001
	4	1	4	10	0	5.6	1.0
	4	5	4	9	1.8	11.1	0.02
	6	4	4	5	0	5.6	0.025
	7	3	2	5	0	5.6	0.003

Overall P-value <.0001

Haplotypes shown in boldface represent significance with comparison to wild-type (WT) (2 copies) frequencies. Only haplotypes with a frequency above 3% in carriers are shown.

Table 4 *SMN1* Population Screening Data for g.27134T>G RFLP

Ethnicity	SMN1 copy number	RFLP Negative	RFLP Heterozygous	RFLP Homozygous	Total
Ashkenazi Jewish	2	315	0	0	315
	1	4	0	1	5
African American	2	108	26	2	136
	3	21	86	4	111
	4	1	23	0	24
	1	2	0	0	2
Asian	2	222	0	0	222
	3	20	2	0	22
	4	1	1	0	2
	1	1	0	0	1
Hispanic	2	206	12	1	219
	3	20	20	0	40
	4	0	2	0	2
	1	12	0	0	12
European	2	413	2	0	415
	3	23	4	0	27
	4	2	2	0	4

Table 5 Microsatellite Analysis of AJ Deletion Carriers

MARKER	D5S6 81	D5S4 35	MS 1	SMN	D5S6 10	MS 2	MS 3
PHYSICAL LOC (Mb)	67.64	68.13	68.7 6	68.9- 70.36	71.02	71.2 9	71.5 1

MLPA copy number

Group 1del	Sample #							
SMN1/SMN2(1/1)	1	2,4	2,5	4,5		2,5	1,4	1,2
	2	2,2	1,6	3,4		2,5	2,4	2,2
	3	2,2	1,5	4,10		2,4	4,10	1,2

	4	2,4	1,3	4,4		2,5	4,6	2,4
	5	2,4	4,5	1,4		5,5	5,5	3,4
Group 2del								
SMN1/SMN2(1/2)	6	2,4	1,1	4,8		5,11	4,5	2,2
	7	2,5	1,5	3,8		2,3	5,5	1,2
	8	1,2	1,5	3,7		4,5	4,5	1,3
	9	2,2	1,5	4,5		1,5	4,5	2,2
	10	2,2	5,5	4,7		5,5	1,4	2,3
	11	2,4	1,3	4,11		5,5	2,4	1,2
	12	2,4	1,4	4,4		4,9	2,10	3,6
	13	2,4	1,3	4,5		4,9	4,6	2,2
	14	2,9	1,5	4,8		5,7	1,4	2,4
	15	1,4	3,5	5,15		4,4	4,5	1,2
	16	2,6	4,5	1,4		5,12	5,5	1,3
	17	2,2	1,5	4,13		4,5	5,5	1,3
Group 3del								
SMN1/SMN2 (1/del ex7,ex8) or (del ex7,ex8/1)								
	18	3,6	3,5	3,4		2,10	2,5	1,3
	19	2,6	1,5	4,4		4,5	4,6	2,2
	20	2,2	5,5	4,7		5,9	4,4	1,3
Group 4del								
SMN1/SMN2(del ex7,ex8/2)								
	21	2,4	3,5	4,9		2,9	4,5	2,2
	22	3,5	4,5	2,4		4,5	3,6	1,2
	23	2,4	1,5	4,4		5,6	4,5	2,3

Table 6 Microsatellite Analysis of AJ Duplication Carriers

MARKER	D5S68 1	D5S43 5	MS1 68.7	SMN 68.9- 70.36	D5S61 0	MS2 71.2	MS3 71.5
PHYSICAL LOC (Mb)	67.64	68.13	6		71.02	9	1

MLPA copy number

Group
1dupSMN1/SMN2(3/
0)

Sample #							
1	2,4	4,5	6,8		6,10	2,4	1,1
2	2,3	3,5	4,6		2,10	2,5	1,1
3	2,3	5,5	6,9		2,10	2,4	1,4
4	2,4	5,5	6,9		4,10	2,2	2,2
5	2,2	3,5	4,6		5,10	4,5	1,2
6	2,4	5,6	4,6		5,10	4,4	1,2
7	2,4	4,5	6,9		9,10	4,4	1,1
8	2,4	4,5	6,9		5,10	2,4	2,3
9	2,2	3,5	2,6		5,10	2	1,2
10	2,2	5,5	4,6		9,10	4,4	1,6
11	2,2	4,5	5,6		4,10	2,5	1,2

Group
2dup

SMN1/SMN2(3/
1)

12	10,11	2,5	5,6		4,10	2,4	1,2
13	2,6	5,5	3,6		4,9	4,10	1,3
14	2,6	3,5	4,6		2,5	2,4	1,2
15	2,3	1,5	4,6		4,11	3,5	2,2
16	2,10	2,5	6,17		2,2	2,4	1,2
17	2,2	2,5	6,12		2,2	2,4	1,2
18	2,4	1,5	3,5		4,4	2,4	1,2
19	2,4	5,5	3,5		2,6	4,5	1,1
20	2,2	1,4	4,4		6,9	4,5	1,1
21	4,6	5,5	4,5		6,9	5,5	1,2
22	2,6	1,4	4,9		2,10	5,6	1,2
23	2,4	3,3	4,7		2,5	2,6	4,6
24	2,4	3,3	7,7		5,5	2,2	2,6
25	2,6	3,5	4,7		5,5	2,6	2,6
26	2,4	4,5	4,4		6,7	3,4	1,2
27	2,7	3,5	5,6		8,10	10,5	2,3
28	2,2	1,5	8,6		10,11	4,5	2,4
29	2,4	5,5	4,6		2,10	4,5	1,2
30	4,4	3,5	4,6		2,10	2,4	2,5
31	2,4	5,5	5,6		2,10	2	1,2
32	2,3	1,5	8,6		9,10	4	2,4
33	2,8	4,5	4,6		5,10	2,6	1,2
34	2,2	4,5	4,6		4,10	2,10	2,2
35	2,2	3,5	4,6		2,10	2,4	2,2
36	2,4	5,5	6,8		6,10	2,5	1,5
37	2,4	1,5	4,6		2,5	4,4	1,3
38	2,4	3,5	4,6		2,9	4,4	2,4
39	2,2	5,5	13,6		5,9	2,5	1,3
40	2,2	4,5	4,6		5,9	6,10	3,4
41	2,4	5,5	5,9		2,6	5,7	1,2
42	2,4	2,5	5,9		2,6	5,7	1,2
43	2,6	5,5	3,4		4,9	4,4	3,4
44	2,7	5,5	3,4		4,11	4,4	2,4
45	2,5	1,5	3,5		4,4	4,6	1,2
46	2,4	3,5	3,5		2,4	2,4	2,4
47	2,2	5,5	4,5		9,11	8,5	1,2
48	2,2	4,4	4,4		4,5	2,8	1,5
49	4,6	5,7	1,4		2,5	2,4	1,3
50	2,6	5,5	7,9		2,5	2,6	6,1
51	2,2	1,5	7,9		2,5	2,2	6,4
52	2,4	3,8	5,7		5,9	2,5	6,6
53	2,2	1,3	3,7		2,5	2,5	1,2
54	2,2	3,5	4,7		5,5	2,4	6,3
55	3,4	5,8	5,7		5,9	2,5	6,6
56	2,4	3,10	7,9		1,2	2,5	1,2
57	2,4	3,4	2,10		5,5	2,4	1,3
58	2,5	3,5	5,9		2,8	4,10	1,2

Group
3dup

SMN1/SMN2(3/2)								
	59	2,4	3,5	3,6		4,10	2,5	2,4
	60	2,7	3,5	2,9		4,5	2,4	2,9
	61	2,2	4,5	9		4,5	4	3,9
	62	2,4	5,5	4,9		4,9	4,5	1,3
	63	2,2	5,5	4,7		2,5	2,4	1,3
	64	4,6	3,3	2,8		5,9	2,5	1,2
	65	2,6	1,1	3,3		5,11	2,4	1,2
	66	2,4	1,5	4,4		5,10	2,5	3,5
	67	4,6	4,5	4,4		5,9	9,10	1,3
	68	2,6	3,5	2,10		2,5	NA	NA
	69	2,10	3,4	3,5		5,5	4,9	1,3
SMN1/SMN2(4/0)								
	70	2,5	5,5	6,6		10,10	2,4	2,4
	71	3,7	7,7	4,9		9,12	3,5	2,2
	72	2,5	2,5	6,9		2,10	4,4	1,1

NA=Not Analyzed

Table 7 Microsatellite Analysis of AJ Controls - 2 copies of *SMN1*

MARKER		D5S681	D5S435	MS1	SMN	D5S610	MS2	MS3
PHYSICAL LOC (Mb)		67.64	68.13	68.76	68.9-70.36	71.02	71.29	71.51
	Sample #							
	1	2,4	4,5	4,5		2,5	2,4	2,3
	2	1,2	5,5	3,4		4,5	4,4	2,2
	3	2,4	5,5	4,4		5,5	4,4	3,3
	4	4,5	5,7	4,6		5,7	4,5	1,2
	5	2,4	5,5	3,4		5,11	5,10	1,6
	6	4,4	5,9	4,11		2,5	2,2	2,2
	7	2,4	1,5	3,4		4,5	2,7	1,3
	8	1,2	4,4	3,4		5,9	4,6	1,2
	9	3,3	1,3	4,7		4,5	2,4	1,1
	10	1,1	7,8	4,5		5,11	2,4	1,3
	11	2,5	1,4	3,5		2,10	4,5	2,5
	12	2,4	4,5	4,9		2,4	2,4	3,3
	13	2,4	5,5	3,3		2,5	2,5	1,2
	14	4,4	5,5	4,7		2,11	5,5	1,2
	15	2,5	1,3	3,4		1,3	5,5	2,2
	16	2,2	1,5	1,5		4,5	5,6	3,3
	17	2,4	1,1	3,6		2,6	1,10	2,4
	18	2,6	3,4	4,5		7,8	5,10	3,4
	19	2,6	4,5	4,5		9,10	2,4	1,1
	20	1,2	1,1	3,8		3,11	8,8	1,2
	21	2,3	1,4	4,6		5,9	4,9	1,2
	22	3,7	4,4	4,7		2,7	3,6	1,2
	23	4,5	4,5	4,4		4,7	5,10	3,3

24	2,7	4,4	4,5		9,9	5,5	1,1
25	2,4	1,5	4,6		2,11	4,5	1,2
26	4,6	1,4	4,9		2,2	3,5	1,2
27	2,4	4,5	7,10		2,5	4,5	2,3
28	2,4	4,5	9,9		5,9	2,5	1,2
29	4,4	3,10	3,5		2,4	2,4	1,4
30	2,2	1,5	3,10		11,11	4,5	1,2
31	2,4	1,3	5,11		2,11	2,4	2,5
32	2,3	5,5	7,10		2,5	2,4	2,2
33	2,6	3,5	4,5		2,4	4,5	2,2
34	5,6	5,5	4,7		5,5	2,6	2,4
35	6,10	3,4	4,9		2,5	4,9	1,4
36	2,9	1,3	4,8		2,7	4,4	2,4
37	2,4	3,5	4,6		2,5	4,10	2,2
38	2,4	4,5	3,5		4,9	2,4	1,4
39	2,4	3,5	5,7		2,2	4,9	1,1
40	3,5	5,5	4,7		2,5	2,2	2,4
41	1,4	5,5	4,7		2,6	1,10	1,1
42	2,2	3,5	2,3		2,5	2,10	1,2
43	4,5	3,5	2,9		5,5	2,2	2,2
44	4,4	5,5	4,4		5,9	3,5	2,3
45	2,4	3,5	5,5		2,2	2,9	2,4
46	4,4	4,5	4,9		2,11	4,5	1,2
47	2,2	5,5	3,4		4,10	4,5	2,3
48	2,4	1,5	4,6		2,9	1,4	1,1
49	4,4	3,4	4,4		4,9	2,10	3,3
50	2,2	1,3	4,5		5,9	5,6	1,3
51	2,2	5,5	3,4		5,6	2,6	1,2
52	2,4	1,5	3,7		2,4	4,5	2,2
53	2,4	4, 5	5,9		2,5	5,5	1,2
54	2,7	5,5	3,4		10,10	3,4	1,2
55	1,4	2,8	4,5		2,4	3,4	1,3
56	2,2	3,5	3,5		2,3	2,2	3,4
57	2,3	5,5	9,9		2,12	2,4	1,5
58	2,4	4,5	5,9		2,9	4,6	1,1
59	2,2	5,5	3,3		8,8	2,2	3,3
60	2,4	5,5	4,5		2,4	2,4	1,2
61	2,2	5,5	3,8		4,9	5,5	1,3
62	4,5	3,5	4,14		2,5	3,4	1,3
63	2,2	5,5	3,8		5,7	5,5	1,1
64	4,4	3,4	4,5		2,6	2,4	3,4
65	2,4	3,5	4,4		2,9	1,2	1,3
66	3,4	3,4	3,4		2,6	2,5	3,4
67	3,4	5,5	7,8		5,6	2,5	2,3
68	3,6	1,5	4,4		2,9	3,6	1,1
69	2,4	1,5	3,5		4,9	4,6	1,2
70	2,2	3,5	3,5		2,2	2,9	3,4
71	2,4	5,5	3,7		4,5	4,10	3,4
72	3,8	1,1	4,8		9,10	2,4	2,2
73	2,2	3,5	5,7		5,11	5,10	1,2

74	2,4	5,5	3,3		4,5	5,5	2,3
75	2,6	3,8	5,5		2,11	2,2	3,3
76	4,5	5,5	4,9		2,11	3,5	2,6
77	2,6	1,5	3,4		2,5	2,4	1,3
78	2,4	4,5	4,4		4,7	5,5	2,2

Table 8 Allele marker lengths

PHYSICAL LOC (Mb)	67.64	68.13	68.76	71.02
Alleles for Marker	D5S681	D5S435	MS1	D5S610
1	143	122	98	124
2	145	124	114	128
3	149	130	116	132
4	151	132	118	139
5	155	134	120	141
6	157	138	122	134
7	141	126	124	136
8	147	128	126	143
9	153	140	128	126
10	137	136	130	117
11	128	116	132	130
12			136	147
13			112	
14			134	
15			102	
16			106	
17			108	

Table 9 SMA Carrier Detection and Residual Risk with addition of *SMN1* g.27134T>G and g.27706_27707delAT

Ethnicity	Carrier Frequency	Detection Rate	Residual risk after negative carrier screen with 2copies of <i>SMN1</i>	Detection rate with addition of <i>SMN1</i> g. 27134T>G and g.27706_27707 delAT with 2copies of <i>SMN1</i>	Residual risk after negative result for <i>SMN1</i> g.27134T>G and g.27706_27707delAT with 2copies of <i>SMN1</i>	Residual risk after positive result for <i>SMN1</i> g.27134T>G and g.27706_27707delAT with 2copies of <i>SMN1</i>
Ashkenazi Jewish	1 in 41.1	90%	1 in 345	94%	1 in 580	carrier
Asian	1 in 53 ¹	92.6% ¹	1 in 628 ¹	93.3%	1 in 701.8	carrier
African American	1 in 66 ¹	71.1% ¹	1 in 121 ¹	N/A	1 in 395.7	1 in 33.5
Hispanic	1 in 117 ¹	90.6% ¹	1 in 1061 ¹	N/A	1 in 1762	1 in 139.6
Caucasian	1 in 35 ¹	94.9% ¹	1 in 632 ¹	N/A	1 in 769.3	1 in 28.6

N/A - not applicable or cannot be calculated, ¹ Hendrickson et al., 2009

Table 10 Risk of having a child with SMA if one member of a couple is a known carrier and the other has two copies of *SMN1*

Ethnicity of partner with 2 copies of <i>SMN1</i>	Negative result by dosage sensitive methods	Negative result for <i>SMN1</i> g.27134T>G and g.27706_27707delAT	Positive result for <i>SMN1</i> g.27134T>G and g.27706_27707delAT
Ashkenazi Jewish	1 in 1380	1 in 2320	1 in 4
Asian	1 in 2512	1 in 2807	1 in 4
African American	1 in 484	1 in 1583	1 in 134
Hispanic	1 in 4244	1 in 7048	1 in 558.4
Caucasian	1 in 2528	1 in 3185	1 in 114

Table 11 List of novel sequence variants identified on the AJ and AA RFLP+ duplication alleles

Ethnicity				AJ	AJ	AJ	AA	AA	AA	Seq. Method
SMN1 duplication RFLP status				dup/RFLP_het	dup/RFLP_wt	non dup/RFLP_wt	dup/RFLP_homo	dup/RFLP_het	dup/RFLP_wt	
Chr. 5 Position	SMN1 Gen. Position	Exon/ Intron	Ref. Nuc.	Seq.	Seq.	Seq.	Seq.	Seq.	Seq.	
70232445	11678	Intron 1	G	<u>G</u> T	G	G	<u>T</u>	<u>G</u> T	G	NGS w/Sanger
70236541	15774	Intron2a	G	<u>G</u> A	G	G	<u>A</u>	<u>G</u> A	G	NGS w/Sanger
70243571	22804	Intron 6	G	<u>G</u> A	G	G	<u>A</u>	<u>G</u> A	G	Sanger
70246957	26190	Intron 6	A	<u>A</u> G	A	A	<u>G</u>	<u>A</u> G	A	Sanger
70247901	27134*	Intron 7	T	<u>T</u> G	T	T	<u>G</u>	<u>T</u> G	T	Sanger
70248473-4	27706-27707	Exon 8	AT	AT/ <u>del</u> AT	AT	AT	[<u>del</u> AT]	AT/ <u>del</u> AT	AT	Sanger

* g.27134T>G, HypCH4III RFLP, AJ: Ashkenazi Jewish, AA: African American; last two rows represent the g.27134T>G and g.27706_27707delAT; variant residues are underscored

[0077] From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

Claims

What is Claimed is:

1. A method of identifying a human subject as a carrier of a *SMN1* duplication allele comprising:
 - (a) obtaining a nucleic acid sample from a human subject;
 - (b) screening the nucleic acid by determining
 - (i) the identity of the nucleotide corresponding to position 11678 of SEQ ID NO:1; or
 - (ii) the identity of the nucleotide corresponding to position 15774 of SEQ ID NO:1; or
 - (iii) the identity of the nucleotide corresponding to position 22804 of SEQ ID NO:1; or
 - (iv) the identity of the nucleotide corresponding to position 26190 of SEQ ID NO:1; or
 - (v) the identity of the nucleotide corresponding to position 27134 of SEQ ID NO:1, or
 - (vi) whether the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 has been deleted; and
 - (c) identifying the human subject as a carrier of a *SMN1* duplication allele if
 - (i) the nucleotide corresponding to position 11678 of SEQ ID NO:1 is not G or there is a nucleotide inserted between nucleotides corresponding to positions 11678 and 11679 of SEQ ID NO:1;
 - (ii) the nucleotide corresponding to position 15774 of SEQ ID NO:1 is not G or there is a nucleotide inserted between nucleotides corresponding to positions 15774 and 15775 of SEQ ID NO:1;

- (iii) the nucleotide corresponding to position 22804 of SEQ ID NO:1 is not G or there is a nucleotide inserted between nucleotides corresponding to positions 22804 and 22805 of SEQ ID NO:1;
- (iv) the nucleotide corresponding to position 26190 of SEQ ID NO:1 is not A or there is a nucleotide inserted between nucleotides corresponding to positions 26190 and 26191 of SEQ ID NO:1;
- (v) the nucleotide corresponding to position 27134 of SEQ ID NO:1 is not a T or there is at least one nucleotide inserted between nucleotides corresponding to positions 27134 and 27125 of SEQ ID NO:1; or
- (vi) the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 is deleted.

2. The method according to claim 1 wherein the human subject is identified as a carrier of a *SMN1* duplication allele based on the presence of a nucleotide structure selected from the group consisting of a T corresponding to the nucleotide at position 11678 or 11679 of SEQ ID NO:1, an A corresponding to the nucleotide at position 15774 or 15775 of SEQ ID NO:1, an A corresponding to position 22804 or 22805 of SEQ ID NO:1, a G corresponding to the nucleotide at position 26190 or 26191 of SEQ ID NO:1, a G corresponding to the nucleotide at position 27134 or 27135 of SEQ ID NO:1 and an AT deletion corresponding to the nucleotides at positions 27706-27707 of SEQ ID NO:1.

3. A method of identifying a human Spinal Muscular Atrophy (SMA) silent (2+0) carrier comprising:

- (a) screening nucleic acid from a human subject by determining
 - (i) the identity of the nucleotide corresponding to position 27134 of SEQ ID NO:1, or
 - (ii) whether the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 has been deleted; and

- (b) identifying an individual as a silent (2+0) carrier of SMA if the nucleotide corresponding to position 27134 of SEQ ID NO:1 is not a T or the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 is deleted, or both.
- 4. The method according to claim 3 further comprising obtaining a sample of nucleic acid from the human subject.
- 5. The method according to claim 3 or 4 further comprising providing genetic counseling to the human subject based on the results of the method of identifying a human Spinal Muscular Atrophy (SMA) silent (2+0) carrier.
- 6. A method of identifying a human Spinal Muscular Atrophy (SMA) silent (2+0) carrier comprising:
 - (a) screening nucleic acid from a human subject by determining the identity of the nucleotide corresponding to position 27134 of SEQ ID NO:1;
 - (b) further determining if the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 has been deleted; and
 - (c) identifying an individual as a silent (2+0) carrier of SMA if the nucleotide corresponding to position 27134 of SEQ ID NO:1 is G and the A-T dinucleotide corresponding to positions 27706-27707 of SEQ ID NO:1 is deleted.
- 7. The method according to claim 6 further comprising obtaining a sample of nucleic acid from the human subject.
- 8. The method according to claim 6 or 7 further comprising providing genetic counseling to the human subject based on the results of the method of identifying a human Spinal Muscular Atrophy (SMA) silent (2+0) carrier.
- 9. The method according to claim 6 or 7 wherein the identity of the nucleotide corresponding to position 27134 of SEQ ID NO:1, and the identities of the nucleotides corresponding to positions 27706 and 27707 of SEQ ID NO:1 are determined by restriction fragment length polymorphism.

10. The method according to claim 6 or 7 wherein a G is identified at the position corresponding to position 27134 of SEQ ID NO:1 by restriction fragment length polymorphism using restriction endonuclease HpyCH4III.
11. A method of identifying a human Spinal Muscular Atrophy (SMA) silent carrier comprising screening the nucleic acid from a human subject for a *SMN1* (SEQ ID NO:1) haplotype that correlates with *SMN1* deletion and increased occurrence of SMA in a human population, wherein the SMA haplotype comprises:
 - (i) a polymorphism selected from the group consisting of D5S681 having allele 2, D5S435 having allele 1 or allele 5, MS1 having allele 4, and D5S610 having allele 5, wherein the presence of the *SMN1* haplotype in the nucleic acid identifies the subject as having an *SMN1* deletion and being an SMA silent carrier, and wherein the absence of the *SMN1* haplotype in the nucleic acid identifies the subject as not being an SMA silent carrier.
12. The method according to claim 11 further comprising obtaining a sample of nucleic acid from the human subject.
13. The method according to claim 11 or 12 wherein the silent carrier is a silent (2+0) carrier.
14. The method according to claim 13 further comprising providing genetic counseling to the human subject based on the results of the method of identifying a human Spinal Muscular Atrophy (SMA) silent carrier.
15. The method according to claim 13 wherein the SMA haplotype for D5S681-D5S435-MS1-D5S610 is 2-5-4-5.
16. The method according to claim 13 wherein the SMA haplotype for D5S681-D5S435-MS1-D5S610 is 2-1-4-5.
17. A method of identifying a human Spinal Muscular Atrophy (SMA) carrier comprising:

screening the nucleic acid from the human subject for a *SMN1* (SEQ ID NO:1) haplotype that correlates with *SMN1* deletion and increased occurrence of SMA in a human population, wherein the SMA haplotype comprises polymorphism MS1 having allele 6.

18. The method according to claim 17 further comprising obtaining a sample of nucleic acid from the human subject.

19. A method of identifying a human Spinal Muscular Atrophy (SMA) carrier comprising:

screening the nucleic acid from the human subject for a *SMN1* (SEQ ID NO:1) haplotype that correlates with *SMN1* duplication and increased occurrence of SMA in a human population, wherein the SMA haplotype is selected from the group consisting of

(i) polymorphism D5S681 having allele 2, D5S435 having allele 5, MS1 having allele 4, and D5S610 having allele 5;

(ii) polymorphism D5S681 having allele 2; D5S435 having allele 5, MS1 having allele 9, and D5S610 having allele 4;

(iii) polymorphism D5S681 having allele 2, D5S435 having allele 5, MS1 having allele 6; and

(iv) polymorphism D5S681 having allele 2, D5S435 having allele 5, MS1 having allele 6, and D5S610 having allele 10,

wherein the presence of the *SMN1* haplotype in the nucleic acid identifies the subject as having an *SMN1* duplication and being an SMA carrier, and wherein the absence of the *SMN1* haplotype in the nucleic acid identifies the subject as not being an SMA carrier.

20. The method according to claim 19 further comprising obtaining a sample of nucleic acid from the human subject.

21. An isolated nucleic acid comprising a sequence selected from the group consisting of sequences of at least 17 nucleotides comprising a nucleotide corresponding to position 27134 of SEQ ID NO:1, sequences of at

least 17 nucleotides comprising a dinucleotide corresponding to position 27705 and 27708 of SEQ ID NO:1, sequences of at least 17 nucleotides comprising the sequence set forth at positions 69-148 of SEQ ID NO:2, sequences of at least 17 nucleotides comprising the sequence set forth at positions 61-127 of SEQ ID NO:3, and sequences of at least 17 nucleotides comprising the sequence set forth at positions 61-216 of SEQ ID NO:4.

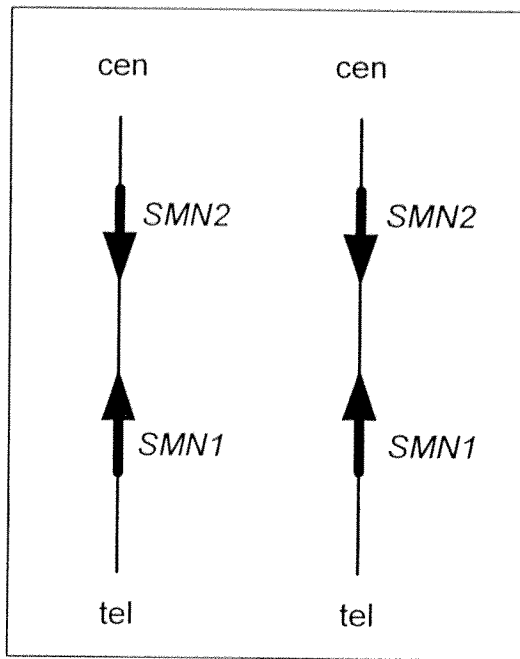
22. The isolated nucleic acid according to claim 21 wherein the nucleic acid is about 10 to 200 nucleotides in length.

23. A kit for diagnosing a Spinal Muscular Atrophy (SMA) carrier comprising the isolated nucleic acid according to claim 21 and a protocol for the use thereof in diagnosis of a SMA carrier.

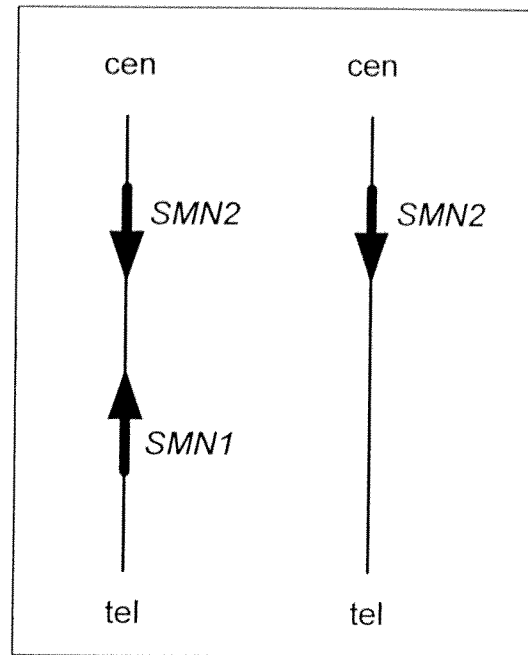
1/5

Figure 1

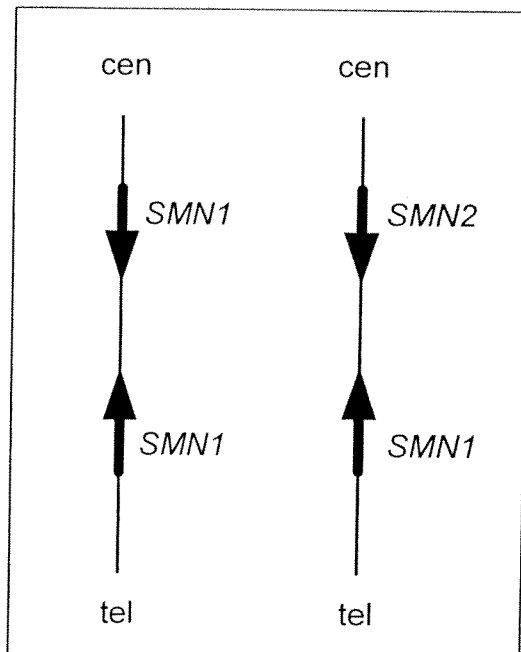
A.



B.



C.



D.

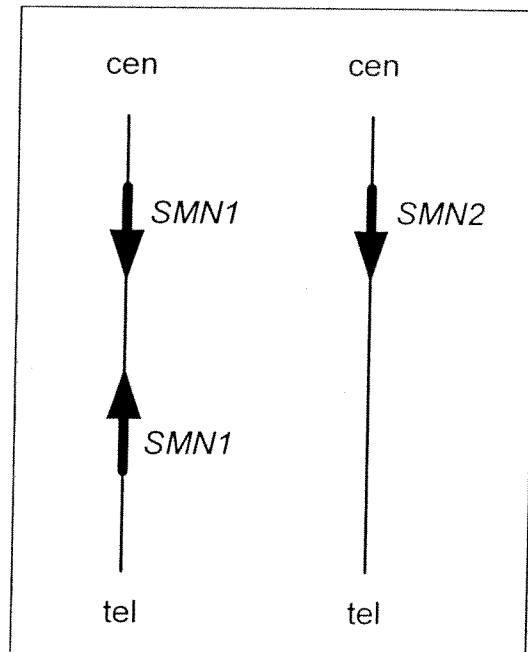
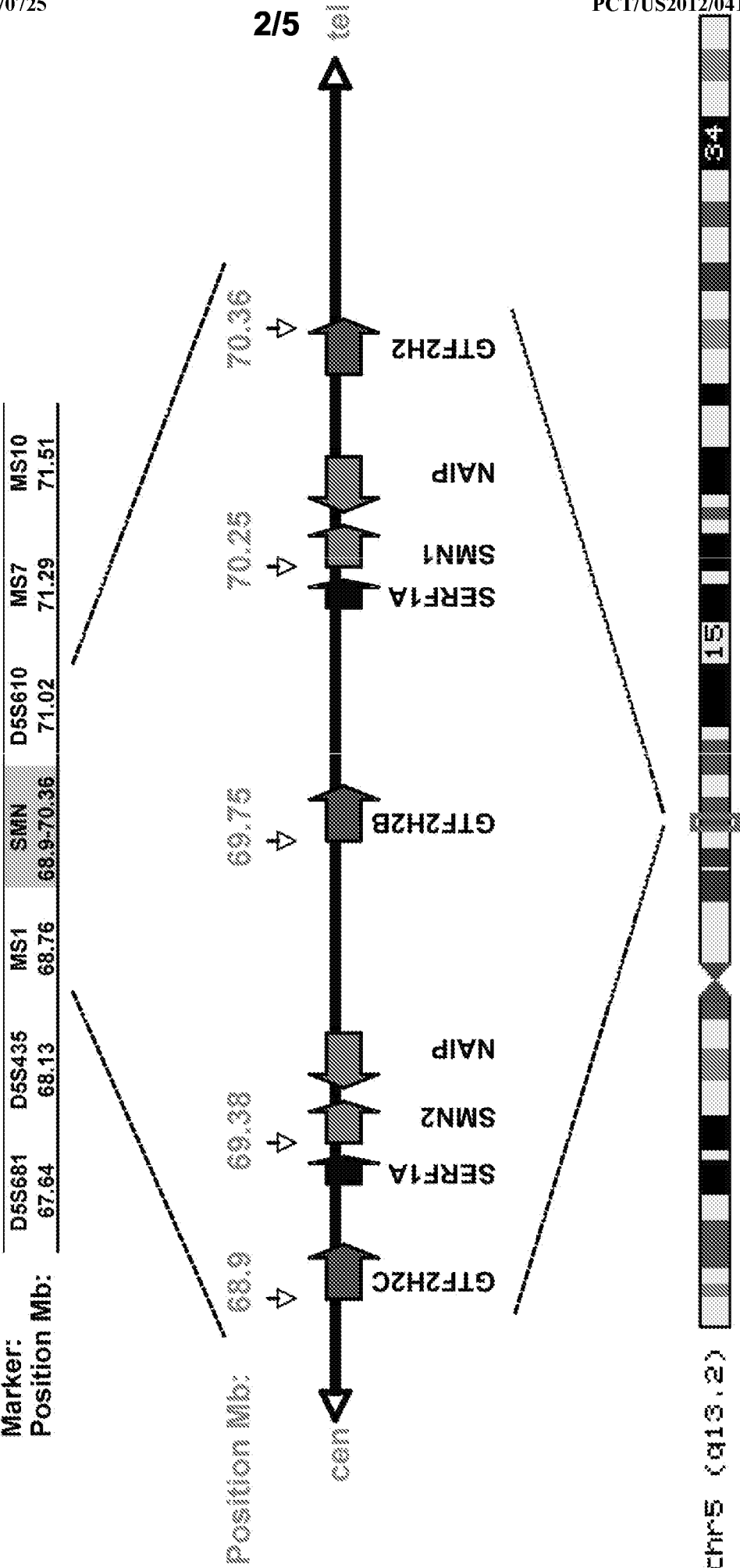


Figure 2



3/5

Figure 3A

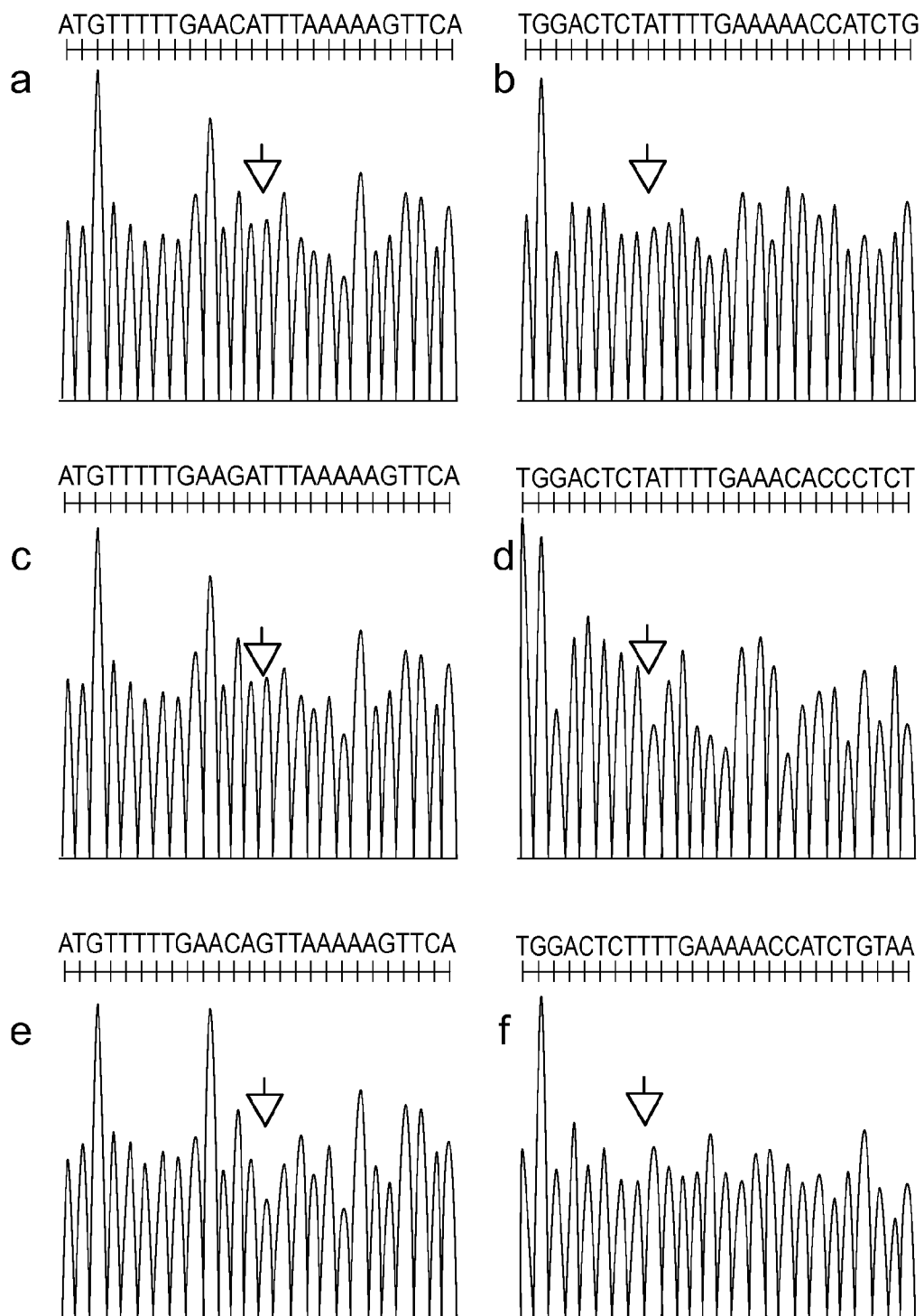
```

rs 4916, C in SMN1/T in SMN2
*
26972 TTTTAACT TCCCTTATTT TCCCTACAGG GTTTCAGACA AATCAAAA GAAGAGGT SGTCAATTC CTTAATTAA SGAGTAAGTC TGCCAGCAAT
AAAAAATTGA AGGAAATATA AGGAAATGCC CAAATCTGTG TTATGTTT CTCCTTCCA CGAGTGAAG GAATTAAT CCTCAATCAG ACGGTGATA
*****
Exon 7
SMN1-E7F
g.27134T>G
*
27072 ATGAAGTGA ATCTTACTTT TGTAAACTT TATGTTTGT GGAACAACA TGTATTGAA CATTAAAA GTTCAGATGT TAAAAAGTTG AAAGTTTAA
TACTTTCACA TAGAATGAAA ACATTITGAA ATACCAACA CTTTGTGTT ACAAAACCT GTAAATTTT CAAGTCTACA ATTTTCAAC TTTCAAATTA
***
HpyCH4III
27172 GTAAACAAT CAATATTAA GAATTTGAT GCCAAAACA TTAGATAAA GSTTAATCTA CATCCCTACT AGAATCTCA TACTTAAGT GTTGTTATG
CAATTTGTTA GTTATAATTT CTTAAACTA CGGTTTTCAT AATCTATTTT CCAATTAGAT GTAGGGATGA TCTTAAGACT ATCAATTGAC CAACCAATAC
27272 TGAAGAAGC ATACTTTCAC AATAAGAGC TTTAGGATAT GATGCCATTT TATATCACTA GTAGGCAAGC CAGCAGACTT TTTTATTG TGATATGGGA
ACCTTCTTG TATGAAGTG TTATTTCTG AATCCTATA CTACGGTAA ATATAGTAT CATCCGCTG GTCTCTTAA AAAAAATAAC ACTATACCT
*****
SMN1-I7R1
27372 TAACCTAGG ATACGCACT GTACACTCTG ACATATGAG TGCTCTAGT AGTTTAACT GGTCTCCA CAGGACATGG TTTAATGGA ATTGCTCAG
ATTGCAATCG TATGACGTA CATGAGAC TGTATATCTC ACAGATCAG TTCAAAATGA CCACAGTGT CTCCTGTACC AAATGACCT TAACAGTTC
27472 CCTGTGCTC TAATTTCTCA TTTCAGACA TTTTCTGAT TTTTCTGAT TTTTCTGAT TTTTCTGAT TTTTCTGAT TTTTCTGAT TTTTCTGAT
GGAGACCAAG ATTAAGAGT AAACGTCCTG TTTTCTGAT TTTTCTGAT TTTTCTGAT TTTTCTGAT TTTTCTGAT TTTTCTGAT TTTTCTGAT
*****
Exon 8
SMN1-E8R
g.27706_27707delAT
**
27672 ACCAATGCA ATGTGAATA TTTTACGGA CTCTATTTG AAAACATC TGTAAAGC TGGGTGGG GTGGAGGC AGCAGGCG TGAGCAGCT
TGTATGCT TACATTTAT AATGACCT GAGAAAC TTTTGTGAT ACATTTCTG ACCCACCC CACCTCCG TGTGTCCAC ATCTCTGTA
*****
27772 GAGAAATTT GAATGTGGT TAGATTTTG ATGATATGG ATAAATATG GTATTTTAT GAGGTGTG AGGTGTGT TACTTTATAG AGACTGTCT
CTCTTTTAA CTACACCTA AACTAAACT TACTATACC TATTATAC GATTAAATA CTGCACATC TTCCACAC ATCAAAAT TTCTGACAGA
*****
27872 TAATTTGAT ACTTAAGCAT TTAGGAATGA AGTTTAGAG TGCTTTAAT TGTTCATG TGTTCATG AGCTTAAGT AGCTTAAGT GGCATAGT
ATTAACCTA TGAATCTGA AATCTTACT TCACATCTC ACAGATTTT ACAGATTTT TACATAC TCCCATACA CGGTTTACA
*****
27972 TACAGATCT AACTGCTGA CAGCTCTCT CATCTCTCT TTTTCTCTA TCTCTCTCT GTTAAACT ATATATATA AATTTTAT TTTTCTCTA
ATGCTTGA TTGACACT GTACACCA GTACATGAC AAAAAGAT AGAGATTA CAATTTCA TATATATT TTATAATTA AAAAAAT
*****
SMN1-E8R

```

4/5

Figure 3B



5/5

Figure 3