

## Mutation entries in SMA databases

### Guidelines for national curators

## GENERAL CONSIDERATIONS

### Role of the curator(s) of a national database

Molecular data can be collected by many different ways. There are different options which have proven successful in existing registries: a self report filled out by the patient/parents or a form filled out by a geneticist or a clinician. The main role of the curator is to check for the accuracy of these data in collaboration with the different partners.

- ▶ If the data are collected either from the patient or from the treating physician, it will always be necessary for the curator to confirm the data directly from the **laboratory report** itself.
- ▶ Collection of data **directly from the geneticist** may be considered as the best way to collect genetic data of high quality. However, laboratories may have different diagnostic procedures based on many different techniques and levels of expertise. Even when geneticists directly report data to the database, the curator still has to check for the techniques used, and the nomenclature.

### Short compendium of SMA genetics

SMA is clinically classified into three subgroups: acute (type 1), intermediate (type 2) and mild (type 3). The estimated birth prevalence is about 1/10.000 and the carrier frequency is 1/35.

Two homologue copies of the SMN gene are playing a central role in SMA: **SMN1** (or telomeric copy, telSMN, SMNt) and **SMN2** (or centromeric copy, cenSMN, SMNc). Within the SMN region (5q12.2-q13.3) these genes are arranged in tandem on each chromosome.

Both *SMN1* and *SMN2* contain nine exons and are very similar. Their sequences differ only in five nucleotides (three are intronic and two are exonic, located within exons 6, 7, and 8). (See Biros and Forrest J Med Genetics 1999; 36:1-8.) However, only the nucleotide difference in exon 7 is of functional relevance. This C-to-T transition in *SMN2* exon 7 disrupts an exon-splicing enhancer sequence. Consequently, *SMN2* primarily produces transcripts lacking exon 7, whereas *SMN1* produces full-length transcripts. Therefore, somebody who lacks both functioning copies of *SMN1* is always a patient, whereas SMA carriers with a single copy of the *SMN1* gene are symptom-free. The copy gene *SMN2* cannot fully compensate for the lack of expression of mutated *SMN1*. However, if the *SMN2* copy number is increased, the small amount of full-length transcript generated by *SMN2* is often able to produce a milder type II or III

phenotype. In conclusion, the more *SMN2* copies a patient has, the less severe the disease is expected to be.

### A few things to keep in mind

- ▶ Approximately 95-98% of patients with the clinical diagnosis of SMA lack **both copies (homozygous deletion)** of *SMN1* **exon 7** (and **exon 8** in the majority of cases).
- ▶ As described above, the most common genetic result leading to the diagnosis of SMA is a homozygous deletion of *SMN1* exon 7 (and exon 8). This is different in only a minority of SMA patients (2-5%): these patients are compound heterozygous, e.g. carry an ***SMN1* deletion of exon 7 (and exon 8) on one of their alleles and an intragenic point mutation of the *SMN1* gene on the other allele.** Point mutations may be dispersed all over the *SMN1* gene.
- ▶ The most up-to-date diagnostic techniques for SMA deletion analysis are quantitative techniques like Real-Time PCR (TaqMan, SybrGreen I, FRET) or, most commonly used, MLPA. For an accurate genetic diagnosis and inclusion into the registry, results of non-quantitative methods (PCR-RFLP analysis) should be confirmed by a quantitative analysis.
- ▶ In order to detect point mutations, the *SMN1*-gene has to be sequenced. Sequence analysis of all *SMN1* exons and intron/exon borders may be used to identify the intragenic *SMN1* point mutations. Exonic regions must be individually amplified; therefore, sequence analysis does not detect exonic deletions or duplications.
- ▶ **In rare cases, instead of the point mutation there may be a conversion mutation from *SMN1* to *SMN2*** of one copy of the gene (exons 7 and 8), if it affects nucleotide position 873 in exon 7 (c.873C>T). Together with a heterozygous deletion of *SMN1* exon 7, these patients appear to be homozygously deleted for *SMN1* exon 7, if the test method is based on the single nucleotide difference in exon 7 only.
- ▶ The ***SMN2* copy number** is an additional item which can be entered into the database. However, this genetic item cannot be detected by all methods that are used routinely. Since this item might be of interest regarding the prediction of phenotype and possibilities for therapeutic intervention, quantitative analysis including the *SMN2* copy number is strongly encouraged.
- ▶ Each genotyped patient will correspond to one record in the database. So even within a family, each affected individual should have been genotyped to be entered in the database. (Even if it is rare, the occurrence of independent mutations within a family has already been described).
- ▶ Only accurately defined mutations are suitable for inclusion in a database.

## Report of SMA genetic results

### First step: identify the method used for genetic testing

#### A: Quantitative methods (MLPA, Real-Time PCR)

MLPA (Multiplex Ligand-dependent Probe Amplification) and Taq man/Real-time PCR are quantitative methods used to detect exon 7 deletions in the *SMN1* gene and to analyze the *SMN2* copy number.

*Possible results are:*

- ▶ **homozygous deletion of exon 7 (and exon 8) of *SMN1*:** diagnostic criteria fulfilled, patient can be entered into database
- ▶ **homozygous deletion of exon 7 and heterozygous deletion of exon 8 of *SMN1*:** diagnostic criteria fulfilled, patient can be entered into database (this can mean a conversion mutation from *SMN1* to *SMN2* of one copy of the whole gene!)
- ▶ **heterozygous deletion of exon 7 (and exon 8) of *SMN1*:**
  - patient is **carrier** or
  - sequence analysis identified **compound heterozygous point mutation in *SMN1***
- ▶ in consanguineous families, the occurrence of **the same point mutation on both alleles of the *SMN1* gene** in the patient might be considered
- ▶ **compound heterozygous for two different point mutations in *SMN1*** (not reported)
- ▶ evaluation of *SMN2* copy number is possible, but not always given in the genetic report

#### B: RFLP

RFLP (Restriction-fragment length polymorphism) is a qualitative assay to detect the homozygous absence of *SMN1* by taking advantage of the nucleotide difference between *SMN1* and *SMN2* in exon 7 (and exon 8).

- ▶ **homozygous deletion of exon 7 (and exon 8) of *SMN1*:** diagnostic criteria fulfilled, patient can be entered into database
- ▶ Qualitative RFLP analysis cannot distinguish carriers of one *SMN1* copy from individuals with two copies. The *SMN2* copy number cannot be detected by this method.

#### C: Linkage analysis

Linkage analysis is available for families if direct DNA testing is not sufficiently informative. It may be used for confirmation of carrier testing and prenatal testing results.

The detection of the *SMN2* copy number is not possible with this method.

If only a linkage analysis is used, patients should be contacted and encouraged to provide/obtain a direct SMA testing result.

## D: Sequence analysis

Sequence analysis is not offered routinely and can only be performed in some laboratories. This method is used if a patient has been diagnosed with a heterozygous deletion of *SMN1* on one allele in order to detect a compound heterozygous point mutation on the second allele.

Point mutations are found in 2-5% of patients with SMA. These include substitutions (mainly nonsense mutations and missense mutations); deletions or insertions of a small number of nucleotides (frameshifting mutations); and splice site mutations.

To avoid erroneous reports and to be unequivocal, it is essential:

- ▶ to use the standard reference sequence:  
the standard reference sequence for *SMN1* (MIM No. 600354) is NM\_000344.3 (gi:196115055).
- ▶ to follow the international mutation nomenclature.
- ▶ to collect exhaustive data: the name of the mutation at the DNA level **and** at the protein level; the correct exon/intron number, etc.

Sequence analysis cannot determine whether a point mutation is in the *SMN1* gene or in the *SMN2* gene, unless one of these genes is absent. Therefore, finding a mutation does not automatically mean that the mutation is in *SMN1*. However, several point mutations have meanwhile been described. Identification of a previously reported point mutation provides evidence that the mutation is in *SMN1*.

For confirmation that the *SMN1* gene harbours the pathogenic mutation, the two genes can be distinguished by cloning experiments prior to sequencing. Alternatively, *SMN1* can specifically be amplified by long-range PCR prior to nested internal PCR and sequencing.

A list of *SMN1* mutations can, for example, be found in "Molecular and Functional Analysis of Intragenic *SMN1* Mutations in Patients with Spinal Muscular Atrophy" by Sun et al, Human Mutation 2005;25:64-71.

Possible reasons if a sequence alteration is not detected:

- ▶ Patient does not have a mutation in the tested gene (e.g., a sequence alteration exists in another gene at another locus).
- ▶ Patient has a sequence alteration that cannot be detected by sequence analysis (e.g., a large deletion, a splice site deletion). Large deletions or possible duplications should be detected by long-range PCR.
- ▶ Patient has a sequence alteration in a region of the gene (e.g., an intron or regulatory region) not covered by the laboratory's test.

## General rules for SMA genetic reports

- 1) Identify the technique used. Is it a quantitative or qualitative technique? Although quantitative techniques are considered state-of-the-art, also results of PCR-RFLP should be accepted, since not all labs perform quantitative testing. However, each mutation should be validated by two methods (e.g., PCR-RFLP, Real-Time PCR, or MLPA)
- 2) Determine if the diagnostic criteria for SMA have been fulfilled. One of the following options has to be the case:
  - a. homozygous deletion of exon 7 (and exon 8) in *SMN1*
  - b. heterozygous deletion of exon 7 (and exon 8) in *SMN1* compound heterozygous to a point mutation in *SMN1*
  - c. homozygous deletion of exon 7 in *SMN1* and heterozygous deletion of exon 8 in *SMN1*
  - d. homozygous for one point mutation in *SMN1* (very rare, in consanguineous families)
  - e. compound heterozygous for two point mutations in *SMN1* (not yet been reported)
- 3) Report the exon deletion result of the genetic testing on the cDNA level according to the reference sequence:
  - a. deletion of exon 7: c.724\_834del
  - b. deletion of exons 7 and 8: c.724\_885del

## Second step: describe the point mutation (see [www.hgvs.org](http://www.hgvs.org))

Indicate the level of description of the mutation:

- c. coding sequence
- g. genomic DNA
- r. RNA
- p. protein: use the three-letter amino-acid code

Use the correct symbols:

- ">" for substitutions.

Ex: c.815C>T, denotes a change of a cytosine to a thymidine at nucleotide position 815 in exon 6 and corresponds to p.Y272C on the protein level

- "-" for small deletions, duplications or insertions.

Ex: c.558\_559del or c.558\_559delA, denotes a deletion of A from nucleotide 558 to 559.

Ex: c.558\_559dup or c.558\_559dupA, denotes a duplication of the nucleotide A 558 to 559.

Ex: c.208\_209insA, denotes an insertion of one nucleotide (A) between nucleotide 208 and 209.

- "+" and "-" for intronic mutations.

The current recommendations for the description of intronic mutations suggest identifying them relative to the coding DNA reference sequence (c.889+6T>G or c.1603-...) rather than relative to the intron number (IVS3+... or IVS14-...). **The coding DNA reference sequence position used is either the first or last nucleotide of a given exon, as follows:**

"+": Beginning of the intron: the number of the last nucleotide of the preceding exon, a plus sign and the position in the intron.

"-": End of the intron: the number of the first nucleotide of the following exon, a minus sign and the position in the intron.

Validate the mutation:

Some validation steps should be done before considering a change in amino acid as the disease-causing mutation:

- Has the entire coding region been sequenced for mutations?
- What was the technique used? Specifically, was direct sequencing performed, or were scanning methods used (such as DHPLC, SSCP, or other methods)?
- What is the impact on the structure/function of the protein?
- Epidemiological studies: what is the frequency of the sequence variation in at least 200 chromosomes?
- Is the sequence alteration reported as a polymorphism in international databases?
- Is it a *de novo* mutation?

## Rules for point mutations

1. Identify the technique used.
2. Identify the exact exons tested.
  - a. Was the entire gene directly sequenced, or screened by another technique?
  - b. Was the entire gene sequenced, or just one exon?
3. Collect all available data:
  - a. Mutation class (nonsense; missense; frameshift insertion, deletion or insertion/deletion; splice site)
  - b. Exon or intron number
  - c. Nucleotide position (DNA level)
  - d. Amino acid position (protein level)
4. Confirm the correct mutation nomenclature using International Mutation Nomenclature
5. Determine if the mutation has already been described in the literature
6. Determine if the testing was conclusive. Check each of the following. Testing was not conclusive if the answer to any of the following is "No":
  - a. For missense mutations:
    - i. Was the entire coding region sequenced?
    - ii. Was a quantitative test performed to exclude a homozygous deletion?
  - b. For splice site mutations:
    - i. Was the mutation at a location already known to affect splicing?
    - ii. If not, was transcript analysis performed to confirm the effect of the putative mutation on splicing?