

Mutation entries in DMD 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 (1) the techniques used, and (2) the nomenclature.

Few things to keep in mind

- ✦ Only accurately defined mutations are suitable for inclusion in a database,
- ✦ One of the most commonly used diagnostic techniques for *DMD* mutations (multiplex PCR) is not exhaustive, as it tests only a subset of exons.
- ✦ Mutations should be accurately reported:

Since January 1, 2003, the reference coding DNA reference sequence has been defined as GenBank sequence **NM_004006.1**. Nucleotide numbering now starts with position 1 at the A of the ATG translation initiation site. This leads to a shift of 208 bp in the numeration compared to the old cDNA reference sequence, which was based on GenBank file M18533. The curator must be aware that all reports prior generated prior to January 1, 2003 are likely to be misnumbered by current standards.

- ✦ Mutations should be reported using the international mutation nomenclature for standardization. See the recommendations for the description of DNA changes established by the Human Genome Variation Society (HGVS) (<http://www.hgvs.org>).
- ✦ 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).

Report of large rearrangements

A/ Large deletions

The deletions of one or more exons are the most frequent mutations in the DMD gene: they are found in at least 60% of the DMD patients and 80% of the BMD patients (whose diagnosis has been ascertained by muscle dystrophin analysis).

▪ **How to report them?**

The easiest way to collect data for deletions is to indicate the first and the last missing exons.

Example: a deletion from exon 45 to exon 49 could be written "del 45-49".

▪ **How to validate the data?**

1 - Are the deletion borders accurately defined? Yes No

Techniques used: Southern blot multiplex PCR semi-quantitative PCR
MLPA

Exons screened by multiplex PCR:

Exons screened by semi-quantitative PCR:

- ▶ Patients for whom the extents of the deletion have not been unambiguously defined should be further analyzed with an additional set of exons (multiplex PCR, semi-quantitative PCR) or MLPA prior to inclusion in the database. Many patients have been analyzed by using the standard multiplex PCR method, which is intended to detect deletions but not necessarily to define the borders of deletions.
- ▶ If only the Southern blot technique has been performed it is recommended to re-analyze the DNA using a PCR-based technique.

2 - Single exon deletion confirmed?

by a second technique by a second set of primers

The hypothesis of a polymorphism (SNP) or a mutation in one of the primer binding sites leading to a false positive result should be ruled out.

▪ **Description of deletions according to the international mutation nomenclature**

To be entered in a database, the deletions should be described according to the international nomenclature. Using the cDNA reference sequence **NM_004006.1**, deletions are designated by "del" after an indication of the first and last nucleotides deleted.

Remember that:

- ▶ This result derives from analysis of genomic DNA
- ▶ Confirmation of the extent of the deletion is usually not performed at the RNA level
- ▶ Precise intronic breakpoints are not known
- ▶ Expected effect of the deletion on the reading frame is only a prediction

Ex: Report of a deletion of exons 45 to 49

This is indicated in the international nomenclature by c.6439-?_7200+?del

=> denotes the deletion of exons 45 to 49, starting at an unknown position in intron 44, upstream of coding DNA nucleotide 6439 (first nucleotide of exon 45), and ending at an unknown position in intron 49, downstream of coding DNA nucleotide 7200 (last nucleotide of exon 49).

▪ **Some useful tools**

Some software can assist the curator in the description of the mutations according to the international nomenclature:

- + Mutalyzer sequence variant nomenclature check: <http://www.humgen.nl/mutalyzer/1.0.1/>
- + UMD software: the UMD software used to build databases calculates automatically the international nomenclature for all mutations.

Software can help in the prediction of the impact of deletions on the reading frame:

- + DMD exonic deletions/duplications reading-frame checker 1.6 (<http://www.dmd.nl>)
- + UMD software (<http://www.umd.be>): calculates automatically the predicted impact of deletions on the reading frame.

▪ **Exceptions to the reading frame rule**

When the predicted impact of a deletion on the reading frame does not correlate with the reported phenotype in the patient *i.e.* in frame deletion/DMD phenotype or out-of-frame deletion/BMD phenotype, it is recommended to go back:

- to the genetic results: is the extent of the deletion correctly defined?
- to the clinical data: is the reported phenotype correct?

B/ Large duplications

The duplications of one or more exons represent about 5%-8% of the mutations of the DMD and BMD patients. They have been underreported for a long time due to technical difficulties. The recently described MLPA technique allows their routine detection. Like for deletions, the MLPA technique allows the exact determination of proximal and distal borders of duplications. The rules presented for deletions hold for duplications.

- **How to report them?** Indicate the first and last duplicated exons: dup 2; dup 45-48; ...

- **How to validate the data?**

- *Are the duplication borders accurately defined?* Yes No

- Techniques used: Southern blot semi-quantitative PCR MLPA

- Exons screened by semi-quantitative PCR:

- ▶ As with deletions, patients have to be re-analyzed with an exhaustive method such as MLPA if the borders of their duplication remain undefined, or if only the Southern blot technique has been used .
- ▶ If a duplication is identified using a partial gene screen (e.g., semi-quantitative PCR of a select number of exons) **it is important to screen all remaining exons for further duplications, since a non-continuous duplication may be present** (see below). This is more likely for duplications than for deletion mutations.

- **Description of duplications according to the international mutation nomenclature:**

Using the cDNA reference sequence **NM_004006.1**, duplications are designated by "dup" after an indication of the first and last nucleotides flanking the duplicated region.

Ex: dup2 => c. 32-?_93+?dup

dup 45-48 => c.650-?_961+?dup

This nomenclature relies on the assumption that the duplicated exons are the same at the transcript level as at the genomic level.

- **Some useful tools**

- *For nomenclature:*

+ Mutalyzer sequence variant nomenclature check: <http://www.humgen.nl/mutalyzer/1.0.1/>

+ UMD software: <http://www.umd.be>

- Impact on the reading frame:

The prediction of the impact of duplications on the reading frame is questionable because there is no evidence of the location and the orientation of the identified duplications in the *DMD* gene.

C/ Complex rearrangements

Over the last years, the use of the MLPA method that screens all the 79 exons of the *DMD* gene has allowed to identify more complex rearrangements such as noncontiguous deletions or duplications, triplications, or noncontiguous deletions associated with duplications or triplications.

Report of triplications: indicate the first and last triplicated exons: trip 45-48

International nomenclature: trip 45-48 => c.650-?_96+?1tri

Two mutations on the same allele:

dup 5-13 and dup 45-52 => c.[265_2380dup;5326_5922dup]

Three mutations on the same allele:

dup 52-55,dup 63-67 and trip 68-79 => c.[7543_8217+dup; 9225_9807dup; 9808_2691tri]

Rules for deletions/duplications

1. Identify the technique used.
2. Identify the exact exons screened.
3. 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. Were the deletion/duplication borders completely delineated?
 - b. For single exon deletions, was the deletion validated by a second technique/second primer set?
 - c. For duplications: if **any** exon was duplicated, were **all** exons tested?
4. Determine if the mutation and phenotype conform to the reading frame rule. If not: re-evaluate the genetic and clinical data.

Report of point mutations

Point mutations represent from ~30% (DMD patients) to ~15% (BMD patients) of the mutations in the *DMD* gene. 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: NM_004006.1 (GenBank)
- 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.

A/ Use of the standard reference sequence

Important: All point mutations identified and reported before 2003 are likely to have been described by using the outdated reference sequence GenBank file M18533; their description may have not been updated. **It is essential that the curator number nucleotides according to GenBank sequence NM_004006.1. Nucleotide numbering now starts with position 1 at the A of the ATG translation initiation site.**

B/ Use of the current international mutation nomenclature (<http://www.hgvs.org>):

1- 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

2- Use the correct symbols:

- ">" for substitutions.

Ex: c.8914C>T, denotes a change of a cytosine to a thymidine at nucleotide position 8914.

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

Ex: c.2230_2231del or c.2230_2231delAG, denotes a deletion of AG deletion from nucleotide 2230 to 2231.

Ex: c.5434_5437dup or c.5434_5437dupTTCA, denotes a duplication of the four nucleotides 5434 to 5437 (TTCA).

Ex: c.1542_1543insA, denotes an insertion of one nucleotide (A) between nucleotide 1542 and 1543.

▪ "+" and "-" for intronic mutations.

The current recommendations for the description of intronic mutations suggest identifying them relative to the coding DNA reference sequence (c.10294+.... 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. Exon 69 ends with nucleotide c.10086; a mutation at the first nucleotide in intron 69, affecting the splice site, is recorded as c.10086+1G>T.

"-": End of the intron: the number of the first nucleotide of the following exon, a minus sign and the position in the intron. Exon 14 begins with nucleotide c.1603; a mutation at the second intronic nucleotide before the exon is recorded as c.1603-2A>C (within intron 13).

For deep intronic mutations located far from the exonic borders, the recommendation is to use the shortest description as much as possible.

Ex: c.9563+1215A>G (intron 65, 2831 bp long)

Ex: c.961-5925A>C (intron 9, 52718 bp long)

Some examples

old ref sequence	old nomenclature	new ref sequence	new nomenclature	Comments
M18533		NM_004006.1		
9122C>T (exon 59)	9122C>T Q2972X	8914C>T (exon 59)	c. 8914C>T p.Gln2972X	NM_004006.1: position 9122 is located in exon 61. Position of the aa change (2972) allows the curator to check for the correct numeration at the DNA level.
	1040C>T <i>Does this mean?</i> DNA: 1040C>T? (-> p.Ala347Val) <i>or</i> Protein: C1040T? (->c.3119G>A)		c.1040C>T p.Ala347Val	The use of the "c.", "p." symbols and the three-letter amino-acid code is unequivocal.
10294+1G>T	IVS69+1G>T	c.10086+1G>T	c.10086+1G>T	Numeration of intronic mutations is based on cDNA sequence. The curator should still collect the intron number to validate the nomenclature.

C/ Report at the RNA level:

- "r." is used to indicate that a change is described at RNA level,
- Use a, c, g, u nucleotides.

Mutations can be reported at the RNA level only when experimental data are available. Dystrophin transcripts isolated from lymphocytes or from a muscle biopsy may have been analyzed in the patient, either:

- In the course of the diagnostic strategy to identify the mutation,
- To establish the effect of intronic mutations on RNA splicing or further explain exceptions to the reading frame rule (alternative exon skipping, etc.).

For more complex descriptions at the RNA level, refer to the RNA-specific rules available at the HGVS web site (<http://www.hgvs.org>).

Some examples:

- The mutation initially found by studying the transcripts and confirmed at the genomic level:
Ex: The c.8914C>T (p.Gln2972X) mutation is described as **r.8914c>u** at the RNA level. No modification in splicing of the dystrophin transcripts is observed.

- Description of the impact of a splice site mutation on RNA processing:
Ex: The effect of the c.1482+1G>T (DNA level) mutation on splicing is described as **r.1332_1482del** (i.e., skipping the entire DMD exon 12).

- When one change affects RNA processing, yielding two or more transcripts, these are described between square brackets, and separated by a "," character.

Ex. 1: An exonic mutation c.4250T>A at the DNA level (p.Leu1417X at the protein level) was found to alter splicing:

r.[4250u>a, 4234_4344del] denotes the appearance of two RNA molecules. One carries the variation r.4250u>a, and one contains a deletion of nucleotides 4234 to 4344 (corresponding to skipping of *DMD* exon 31).

Ex. 2: Description of the impact of the donor splice site mutation c.9563+1G>T (DNA level) on RNA processing:

r.[9563+1g>a, 9563_9564ins9563+1_9563+4] denotes the nucleotide change c.9563+1G>A causing an insertion of the intronic nucleotides 9563+1 to 9563+4 (shift of the splice donor site to an intronic position) with the nucleotide change 9563+1g>a.

Ex. 3: Description of the impact of a cryptic splice site mutation c.3432+2036G>A (DNA level) on RNA processing:

r.[=, 3432_3433ins3432+2037_3432+3131] denotes the nucleotide change c.3432+2036G>A causing the appearance of two RNA molecules.

One is the normal transcript (r.=). The other contains an insertion of the intronic nucleotides 3432+ 2037 to 3432+3131 (representing the inclusion of a cryptic exon in the mature transcript).

D/ Missense mutations

Missense mutations are rare in the dystrophin gene. Some validation steps should be done before considering a change in amino acid as the disease-causing mutation:

- Has the entire coding sequence 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)?
- Were large rearrangements excluded? If so, were they excluded by MLPA (considered a conclusive technique), or by Southern blot (which is less sensitive)?
- What is the impact on the structure/function of the protein? Is it located in functional domains such as the actin-binding domain or the β dystroglycan-binding domain?
- Does the putative missense mutation have an effect on exon splice enhancer or exon splice suppressor motifs (which can be checked at www.umd.be/SSF).
- Epidemiological studies: what is the frequency of the sequence variation in at least 200 chromosomes?
- Is it a *de novo* mutation?

See the rules for the description of point mutations below.

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; frameshift insertion, deletion, or insertion/deletion; missense; splice site)
 - b. Exon or intron number
 - c. Nucleotide position (DNA level)
 - d. Amino acid position (protein level)
 - e. If transcript analysis was performed: effect on splicing (RNA level)
4. Confirm the correct mutation nomenclature.
 - a. Was the mutation identified before 2003?
 - b. Is it numbered based on the correct reference sequence (GenBank accession NM_4006.1)?
 - c. Confirm the description uses (or correct the description to use) the International Mutation Nomenclature.
5. 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 MLPA performed to exclude duplications?
 - iii. Was transcript analysis performed to exclude intronic point mutations resulting in pseudoexons?
 - b. For splice site mutations:
 - i. Was the mutation at a location already known to affect splicing? (Check previously reported splicing at www.dmd.nl.)
 - ii. If not, was transcript analysis performed to confirm the effect of the putative mutation on splicing?
6. Determine if the mutation and phenotype conform to the reading frame rule. **If not:** re-evaluate the genetic and clinical data.