The impact of paralog genes: detection of copy number variation in spinal muscle atrophy patients

Sergio LAURITO1, 2; Juan A. CUETO1, 3; Jimena PEREZ1; María ROQUÉ1, 2

1 IHEM, Universidad Nacional de Cuyo, CONICET
2 Departamento de Biología, Universidad Nacional de Cuyo, Facultad de Ciencias Exactas y Naturales
3 Instituto de Fisiología, Universidad Nacional de Cuyo, Facultad de Ciencias Médicas

Key words: SMA, CNV, MLPA, Paralog gene

Abstract: Spinal muscular atrophy (SMA) is caused by dysfunction of the alpha motor neurons of the spinal cord. It is an autosomal recessive disease associated to the SMN1 gene, located in the subtelomeric region of 5q13. A paralog SMN2 gene is located at the centromeric region of the same chromosome, which apparently originated by an ancestral inverted duplication occurring only in humans. The exon sequence differs in two nucleotides in exon 7 and exon 8, which leads to an SMN2 transcript that lacks exon 7 and results in a truncated protein. Part (10%) of the SMN2 transcripts avoids the splicing of exon 7 but most of the copies are dysfunctional. In a disease scenario, the more SMN2 copies the higher possibility to restore at least partly the effects of SMN1 deficiency. Some therapeutic approaches are being developed to increase the expression of SMN2. To determine the number of SMN1 and SMN2 copies, the methodology must distinguish accurately between both genes. In this work, we present the results obtained using multiplex ligation-dependent probe amplification (MLPA) in 60 SMA suspected patients/carriers derived from different regions of Argentina. In 32 of these DNA samples we found alterations in SMN1. Among these, 16 presented a heterozygous deletion (carrier status) and 14 an homozygous deletion (patient status) in exon 7 and 8 of SMN1. In one case, exon 7 was found homozygously deleted but exon 8 presented a single copy, and in another case, exon 7 was found heterozygously deleted while exon 8 was normal. Almost half of the patients (7/15) presented a normal diploid number of SMN2 while the other half (8/15) presented an increased number. In this work we showed how a probe-based methodology such as MLPA was able to distinguish between the paralog genes and determine the amount of copies in DNA samples from suspected patients/carriers of SMA.

Introduction

Spinal muscular atrophy (SMA) is the most common debilitating neuromuscular disease in children, together with Duchenne muscular dystrophy. It is caused by dysfunction and death of the alpha motor neurons of the spinal cord. This provokes in the patients a generalized muscle atrophy and weakness. The clinical phenotype is very heterogeneous, and ranging from severe to mild, it is divided in SMA type I, type II, type III (Prior et al., 2010) and for some authors even in type IV (Prior et al., 2010; Arnold and Fischbeck 2018). In 1997, the first description of a genetic cause was published, reporting the disease as hereditary in an autosomal recessive way (Lefebvre et al., 1997). The majority of the patients (92%) have homozygous deletions in the survival motor neuron 1 (SMN1) gene located on the telomeric region of chromosome 5q13.2 (Verhaart et al., 2017).

Only few available studies mention incidence and prevalence of the disease. The incidence of all types of SMA is cited as 1 in 10-12,000 live births (Lunn and Wang 2008; Verhaart et al., 2017). The prevalence of the disease, which is the number of living diseased individuals, is observed as 1-5 per 100,000 individuals, even though a broad variability is reported among different geographic regions and a possible sample bias has to be considered because a small error in the number of cases has a large impact on the estimated prevalence (Lunn and Wang 2008). Besides, another difficulty to assess the exact prevalence is the median life expectancy of the patients, which differs from around one year for type I patients to more than 20 years for type II patients, and figure close to the normal population life expectancy for type III patients (Zerres et al., 1997; Chung et al., 2004; Finkel et al., 2014).

Since SMA is a genetic recessive disease, there are also heterozygous carriers who are clinically unaffected but can transmit the mutation to their descendants. The carrier frequency is estimated to be higher among Caucasian and Asian populations, as compared to the African and Hispanic populations (Verhaart et al., 2017), however the exact determination has not been established, as very severe forms...
(death in utero) and very mild forms (symptom-free in adults) interfere with an accurate estimation (Verhaart et al., 2017). Carriers can be divided based on the location of normal and mutated copies of SMN1: the most frequent form is to have a functional SMN1 allele in one chromosome 5 and a deleted allele in the homolog chromosome. Less common is to present two functional genes on the same chromosome 5 and no corresponding alleles on the homolog chromosome. Some cases, very rarely present three, four or more copies of SMN1, unequally distributed among the two homolog chromosomes.

SMN1 is located in an unstable subtelomeric region of 5q13 and it has an almost identical copy gene called SMN2, at the centromeric region of the same chromosome. This paralog gene has been originated ancestrally by an inverted duplication, which apparently occurred in humans, but not in chimpanzees (Rochette et al., 2001). Interestingly, chimpanzees have many copies of the SMN gene, though they are all identical in their sequence. Instead, the human SMN2 gene has a slightly different sequence and has been detected in populations from Europe, the Central African Republic, and the Congo. There is evidence that SMN gene duplication occurred more than 5 million years ago, before the separation of human and chimpanzee lineages, but that SMN2 appears for the first time in Homo sapiens (Rochette et al., 2001). The exon sequence of both genes differs only in two nucleotides in the coding region (one in exon 7 and one in exon 8), which alters the splicing pattern, leading to an SMN2 transcript that lacks exon 7 and results in a truncated, nonfunctional protein. However, part of the SMN2 transcripts manages to avoid the splicing of exon 7, but since this occurs in a very low rate (10%), most copies are dysfunctional. Therefore, in a scenario of deleted SMN1 gene, the absence cannot be entirely compensated by SMN2 (Arnold and Fischbeck 2018). The more SMN2 copies, the higher the possibility to partly restore the effects of SMN1 deficiency. Therefore, in the last years, it has become crucial in clinic to determine the SMN2 copy number in SMA patients, because a correlation has been established between the severity of the disease and SMN2 copy number: type I patients have a normal or reduced number of SMN2 copies, but most of the type II and III patients show an increased number of SMN2 copies (3-4 copies; Calucho et al., 2018).

In the last two decades, substantial advances have been made in the understanding of the genetic basis of SMA (Finkel et al., 2016; Gomez et al., 2009; Marzese et al., 2008; Laurito et al., 2013; 2015). In brief, 50-100 ng of DNA was denatured at 95°C for 5 min and then incubated for 16 hours at 60°C with the P021/P060 probe mix (P021 for the first 4 analyses, and P060 for the remaining 56 samples). The difference between P021/P060 resides in the probe targets and probe design for SMN2, since P021 has probes flanking SMN1 and SMN2. Following probe hybridization, DNA ligase was added and incubated for 15 min at 54°C. The ligation products were then amplified by the PCR according to the manufacturer's protocol using one primer labeled with 6-FAM. PCR products were subsequently resolved using a CEQ 8000 Capillary Sequencer (Beckman Coulter). The data were analyzed by the Gene Marker version 1.75 software.

To determine the copy number of SMN1 and SMN2, the applied methodology must be able to distinguish accurately between both genes. Multiplex Ligation-dependent Probe Amplification (MLPA) assay is based on different probes which hybridize specifically and simultaneously on distinct genomic regions, and after a fluorescence-based PCR amplification the number of targeted probes can be separately quantified. In this work, we present the MLPA results obtained in our laboratory on SMA patients and carriers derived from different regions of Argentina.

Materials and Methods

Patients and samples

Blood samples of 60 SMA suspected patients/carriers were derived to our service of genetic analyses. All samples were accompanied by a medical order for SMA testing by MLPA. Clinical criteria were not always informed in the medical order. DNA was extracted from whole blood as mentioned before (Laurito and Roqué 2018). In brief, 3 ml of each EDTA-anticoagulated blood sample was washed with Tris-HCl 10 mM EDTA 10 mM. Pellets were resuspended in 3 ml of CTAB solution (2% CTAB; 700 mM NaCl; 50 mM Tris-HCl, pH 8; 10 mM EDTA; and 2% β-mercapto-ethanol) and incubated at 60°C. The solutions were further subjected to chloroform/isoamyl alcohol extraction. Finally, the aqueous phase was transferred to a clean tube and DNA was precipitated with chilled absolute ethanol.

MLPA assay

This assay for the analysis of copy number variation (CNV) was performed as communicated before for other diseases (Laurito and Roqué 2018; Mayorga et al., 2016; Gomez et al., 2009; Marzese et al., 2008; Laurito et al., 2013; 2015). In brief, 50-100 ng of DNA was denatured at 95°C for 5 min and then incubated for 16 hours at 60°C with the P021/P060 probe mix (P021 for the first 4 analyses, and P060 for the remaining 56 samples). The data were then amplified by the PCR according to the manufacturer's protocol using one primer labeled with 6-FAM. PCR products were subsequently resolved using a CEQ 8000 Capillary Sequencer (Beckman Coulter). The data were analyzed by the Gene Marker version 1.75 software.

Results

Mutation detection

Among the 60 derived samples, we found alterations in SMN1 by MLPA in 32 (54%). Among these, 16 (50%) presented a heterozygous deletion in exons 7 and 8 of SMN1 (carrier status) and 14 (43%) an homozygous deletion (patient status). In one case, exon 7 was found homozygously deleted, while exon 8 presented a single copy (3%), and in another case exon 7 was found heterozygously deleted while exon 8 was found normal (3%) (Tab. 1).
TABLE 1
Types of alterations detected by MLPA in the SMA derived samples

<table>
<thead>
<tr>
<th>CNV Detected</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMN1 0 copies</td>
<td>14</td>
<td>43.75</td>
</tr>
<tr>
<td>SMN1 0 copies</td>
<td>1*</td>
<td>3.125</td>
</tr>
<tr>
<td>SMN1 1 copy</td>
<td>16</td>
<td>50</td>
</tr>
</tbody>
</table>

TABLE 2
Double entrance table showing the copy numbers of both paralogous SMN genes in SMA patients and carriers

<table>
<thead>
<tr>
<th>SMN2</th>
<th>0 copies</th>
<th>2 copies</th>
<th>&gt;2 copies</th>
<th>ND</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMN1</td>
<td>0 copies</td>
<td>7</td>
<td>8</td>
<td>ND</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1 copy</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

FIGURE 1. A. MLPA electropherogram comparing a DNA of a patient sample (blue peaks) and DNA control sample (red peaks). The orange arrows indicate a total absence of fluorescence for the probes corresponding to exons 7 and 8 in SMN1 gene in the patient sample, and an increase of fluorescence for the probes for exon 7 and 8 of SMN2 gene (green arrows). B-F. Ratio Analysis between the study/control sample, B shows a homozygous deletion for exon 7 and 8 of SMN1 gene (patient status), and an increment of copy numbers for exon 7 and 8 of SMN2 gene, C a homozygous deletion of SMN1 gene (patient status), with normal copy number of SMN2 gene, D a heterozygous deletion of SMN1 gene (carrier status) and normal copy number of SMN2 gene. E a heterozygous deletion of SMN1 gene (carrier status) and an increment of copy numbers of SMN2. F shows a normal copy number for SMN1 and SMN2.
Only exon 7 should be considered to determine the number of copies of SMN1 gene, because gene conversion between SMN1 and SMN2 exon 8 can occur in 5% of the samples, resulting in a chimeric gene containing the SMN1 exon 7 and the SMN2 exon 8 sequences. Such a hybrid gene results in a protein functionally identical to SMN1 protein (MLPA® MRC-Holland Product Description Probemix P060-B2 SMA). We suspect that both cases marked with an asterisk present an increased number of copies of exon 8, probably due to gene conversion.

Given the clinical importance of the amount of copies of SMN2, in Tab. 2 we crossed the SMN1 status in patients and carriers (0 copy and 1 copy) with the amount of copies of SMN2. Only exon 7 of both genes was considered, to avoid the possible interference of conversion between exons 8. As it can be observed, almost half of the patients (7, which have 0 copies of SMN1) have normal diploid number of SMN2, while the remaining 8 present an increased number. In these latter cases, a less severe phenotype and a possibly better response to treatments based on SMN2 RNA modulation is expected.

The electropherograms and ratio analyses obtained by the software GeneMarker v1.75 are shown in Fig. 1 to represent the different SMN1/2 statuses. Only 4 probes (2 for exon 7 and 8 of SMN1, and 2 for exon 7 and 8 of SMN2) were considered for the analysis, because the remaining ones are reference probes which hybridize in genomic regions not involved in the disease and serve as normal copy number controls.

**Discussion**

Gene duplication is an evolutionary source of new genes and variability, which has been maintained in nature over time (Zhang, 2013). Duplication of a genomic region can occur by error during DNA duplication, and once duplicated the copies can evolve differently, generating paralog genes and new proteins. Unlike ortholog genes, a paralog gene is a new gene that holds a new function. After duplication, two sub-functional proteins may be produced in some cases, which need from each other to cooperate (Baker et al., 2013), while in other cases, a competitive interference may arise between the duplicates (Baker et al., 2013).

In the case of SMN, the SMN2 gene produces 90% of a nonfunctional protein, but a small portion has a functionality similar to that of SMN1. Depending on the amount of SMN2 copies in the genome, the absence of SMN1 in SMA patients may be partially compensated. It has been reported that several variants in SMN2 exon 7 could act as a positive modifier in several patients who display milder clinical phenotypes than other patients with the same SMN2 copy number (Alias et al., 2018). SMN protein is a 38-kDa protein that is ubiquitously expressed in the cytoplasm and nuclei of all tissues, involved in the metabolism and processing of RNA (Coovert et al., 1997).

In this work we showed how a probe-based methodology such as MLPA was able to distinguish between the paralog genes and to determine the amount of copies in 60 studied DNA samples from suspected patients or carriers of SMA. The interpretation of the results is easy, and the costs of the study are reasonable for developing countries such as Argentina. The limit of the technique is the detection over more than 3-4 copies, since the PCR amplification there reaches a plateau. However, this seems not to be a limitation for its clinical use, since more than 4 copies are rare events.

Therefore, MLPA can be considered a useful tool for diagnosis and copy number determination of the paralogous genes SMN1/2 in DNA obtained from circulating blood samples.

**Authors Contributions**

LAURITO (first author) and CUETO carried out the experiments. LAURITO and ROQUE carried out the data analysis. ROQUE wrote the manuscript with support of LAURITO, CUETO and PEREZ. All authors discussed the results and contributed to the final manuscript.

**References**


Finkel RS (2018). Diagnosis and management of spinal muscular atrophy: Part 2: Pulmonary and acute care; medications, supplements and immunizations; other organ systems; and ethics. *Neuromuscular Disorders* 28: 197-207.


Laurito S, Di Pierro J, Roqué M (2015). Neurofibromatosis tipo I. Mutación de splicing detectada por MLPA y secuenciación en la...
Argentina [Neurofibromatosis type 1. splicing mutation detected by MLPA and DNA sequencing in Argentina] Medicina (Buenos Aires) 75: 91-94.

Laurito S and Roqué M (2018). Análisis de variación del número de copias y de patrones de metilación en la región 15q11-q13 [Variation analysis of the number of copies and methylene patterns in region 15q11-q13]. Medicina (Buenos Aires) 78: 1-5.


