UNIVERSITÀ DEGLI STUDI FIRENZE

DOTTORATO DI RICERCA IN NEUROSCIENZE

CICLO XXX

COORDINATORE Prof. Renato Corradetti

The functional consequence of TGS1 mutations and the interaction with Survival of Motor Neuron protein

Molecular diagnosis of the most common forms of Limb Girdle Muscular Dystrophy with autosomal recessive inheritance

Settore Scientifico Disciplinare MED/26

Dottorando

Dott.ssa Sara Aguti

Tutore

Prof. Alessandro Malandrini

Coordinatore

Prof. Renato Corradetti

Anni 2014/2017

TABLE OF CONTENTS

Part II (Molecular diagnosis of the most common forms of Limb Girdle Muscular Dystrophy with autosomal recessive inheritance)

Abstarct

Part I

The functional consequence of TGS1 mutations and the interaction with Survival of Motor Neuron protein

Abbreviations

AdoHcy	*S*-adenosyl-L-homocysteine
AdoMet	*S*-adenosyl-L-methionine
ALS	Amyotrophic lateral sclerosis
ASF/SF2	Serine/arginine-rich splicing factor 1
ASO	Antisense oligonucleotide
ASO-10-27	2'-*O*-2-methoxyethyl-modified ASO
AAV	Adeno-associated virus
C	Cytosine
CBC	Nuclear cap-binding complex
CBs	Cajal bodies
CNS	Central nervous system
cpg15	Candidate plasticity-related gene 15
CRM1	Chromosomal maintenance 1 or Exportin 1
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FDA	Food and drug administration
GAP43	Cytoskeleton-associated growth-associate protein 43
GTF2H2	General transcription factor IIH subunit 2
HDACIs	Histone deacetylases inhibitors
HEK 293T	Human embryonic kidney cells 293
hnRNP A1/A2	Heterogeneous nuclear ribonucleoparticule A1/A2
hnRNP R	heterogeneous nuclear ribonucleoproteins R
hRT	Human telomerase RNA
HuD	Neuronal mRNA-binding protein
ICM	Inner cell mass
IMP1	Insulin-like growth factor mRNA-binding protein 1

ISS-N1	Intronic splicing silencer N1
miRNPs	microRNPs
mRBPs	mRNA-binding proteins
mRNA	Messenger ribonucleic acid
mRNPs	Messenger ribonucleoproteins
NAIP	Neuronal apoptosis inhibitor protein
NBs	Nuclear bodies
NES1	Nuclear export signal 1
NES2	Nuclear export signal 2
NoLS	Nucleolar localisation signal
PHAX	Phosphorylated adapter RNA export protein
pICIn	Chloride conductance regulatory protein
PIMT	Human trimethylguanosine synthase 1
PMO	Phosphorodiamidate morpholino oligonucleotides
Pol II	RNA polymerase II
Pol III	RNA polymerase III
PRMT5	Protein arginine methyltransferase 5
PRMT7	Protein arginine methyltransferase 7
RNP	Ribonucleoprotein
rRNA	Ribosomal ribonucleic acid
scAAV9	Self-complementary adeno-associated virus serotype 9
ScaRNP	Small cajal body RNP
SERF1	Small EDRK-rich factor 1A
SH-SY5Y	Neuroblastoma cell line
SMAs	Spinal muscular atrophies
smn	Survival motor neuron (mouse)
SMN1	Survival motor neuron 1 gene (Human)
SMN2	Survival motor neuron 2 gene (Human)
snoRNPs	Small nucleolar ribonucleoproteins
snRNP	Small nuclear ribonucleoprotein

T	Thymine
TGS1	Human trimethylguanosine synthase 1
TGS1 LF	Trimethylguanosine synthase 1 full-length
Tgs1	*Yeast* trimethylguanosine synthase 1
TGS1 SF	Trimethylguanosine synthase 1 short isoform
TLC1	Template RNA component 1 (*Saccharomyces cerevisiae*)
TMG cap	Trimethylguanosine cap
UNRIP	Unr-interacting protein
wmd	*Drosophila* Unrip orthologue p
WRAP53	Human telomerase cajal body protein 1
YG box	Tyrosine/glycine-rich motif

ABSTRACT

Spinal Muscular Atrophies (SMAs) are genetic neurodegenerative disorders characterized by progressive muscle wasting and loss of muscle function due to severe motor neuron dysfunction. SMAs are phenotypically heterogeneous groups of diseases distinguished by the pattern of features, severity of muscle weakness, age of onset and mode of inheritance. The majority of patients present autosomal recessive inheritance with proximal manifestation, secondary to mutations in *survival motor neuron 1* (*SMN1*) gene. SMA represents one of the most common genetic causes of infant death with an incidence of 1 in 6000.

This study was carried out to investigate a possible involvement of a new gene in the pathogenesis of SMA disease. *Trimethylguanosine Synthase 1* (*TGS1*) gene encodes an enzyme responsible for the nuclear importation and maturation of snRNPs. TGS1 function is closely related to SMN function since both proteins are essential for snRNPs biogenesis. Even if the molecular pathway underlying SMA is still not fully understood, disruption of snRNP biogenesis followed by pre-mRNA misplacing is the most likely explanation for the pathogenesis of SMA.

The discovery of two *TGS1* gene mutations in two patients with type IV SMA clinical features has been the starting point of this project. Functional studies have been performed to understand the functional consequences of these variants and their involvement in SMA disease. Our preliminary studies, mainly performed on one mutation (p.Gly259Ser), seem to confirm our initial theory highlighting a possible relationship between *TGS1* mutations and SMA-like phenotypes.

1. INTRODUCTION

1.1 SPINAL MUSCULAR ATROPHY

The term Spinal Muscular Atrophies (SMAs) is applied to a wide group of genetic disorders that affect the anterior horn cells and selective motor cranial nerve nuclei[1]. SMAs are phenotypically heterogeneous groups of diseases with autosomal recessive, autosomal dominant, or X-linked recessive mode of inheritance. The majority of patients present autosomal recessive inheritance with proximal manifestation.

Autosomal recessive SMA (or 5q-SMA) is a neurodegenerative disorder, which is the most common genetic cause of death in infants, with an incidence of about 1.2 per 10000 live births[2]. More recent studies estimate that the incidence of SMA to be higher about 1 in 6000 with a carrier frequency of 1:38[3]. Spinal muscular atrophy is characterised by progressive symmetrical muscular atrophy and weakness as a result of degeneration of α-motor neurones in spinal cord[4].

The first clinical descriptions were published in 1891 by Guido Werding from University of Vienna [5], which described the disease in two infant brothers. The following year, Johann Hoffmann of Heidelberg University used the term "spinal muscular atrophy" in his paper ("Familiar Progressive Spinal Muscular Atrophy"). Hoffmann credited the two cases described by Guido Werding and placed himself in full agreement with Werding conclusions regarding the familiar and progressive nature of the disease. In the following years Hoffmann described seven cases in four families[6]. Werding's and Hoffmann's papers presented the clinical and pathologic aspects of infantile SMA, one of the different types of spinal muscular atrophy. But only in 1961 the wide spectrum in SMA was observed, thanks to Byers and Banker[1]. They found that life expectations were not always less than 4 years old, as originally described. Some years after, Dubowitz and Munsat studies helped to figure out the age of onset, prognosis and clinical features of SMA, which can be classified into different clinical grades based on the age of onset and severity of the disease[7,8].

In the beginning of 1990, the gene responsible of SMA was mapped to chromosome 5q13.2 and in 1995, the gene was identified as the *Survival Motor Neuron* (*SMN*) gene[9]. This gene is localized in a particular chromosome region, characterized by an inverted duplication region with 4 genes (*SERF1*, *GTF2H2*, *NAIP* and *SMN*) presents in telomeric and centromeric copies[10]. For this reason, *SMN* exists in two paralogue genes: *SMN1* the telomeric copy and *SMN2* the centromeric copy (Fig. 1).

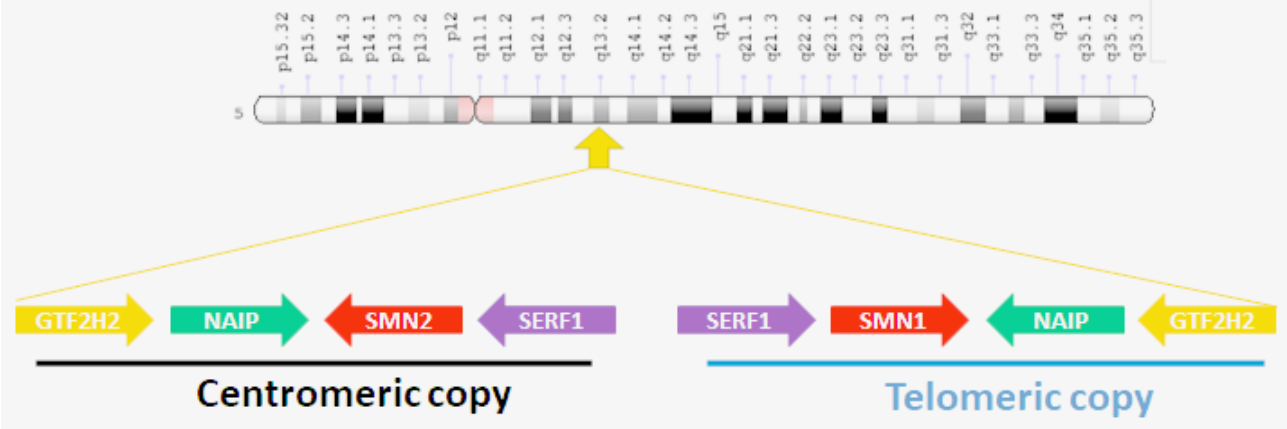


Fig 1: Gene locus of the centromeric and telomeric copy of *Survival Motor Neuron* gene in human chromosome 5.

Both genes encode for SMN protein, which is important during the assembly of small nuclear ribonucleoproteins (snRNPs) involved in pre-mRNA splicing. However, the gene responsible for the selective degeneration of lower motor neurons followed by atrophy of proximal skeletal muscle is only *SMN1*. Mutations in *SMN2* do not lead to disease. Spinal muscular atrophy is due in 95% of cases by biallelic loss of *SMN1* followed by the loss of the encoded protein. While the other 5% of cases by mutations as missense, small deletions and splice site mutations that lead to non-functional protein generation. In both cases, the centromeric copy *SMN2* is still present. This gene differs from *SMN1* gene by several exonic and intronic SNPs. Among them, the transition C>T at position +6 of exon 7 leads to disrupt an exonic splice enhancer[11,12]. As a result, ~90% of the generated mRNAs are with exon 7 skipping (SMNΔ7) and produce an unstable protein. While only ~10% of the transcripts are fully functional. However, the low abundance of the resulted protein from *SMN2* is not sufficient to fully compensate for the loss of the *SMN1* gene. Even if *SMN2* is not enough to sustain the function and the survival of spinal motor neurons, it acts as a positive modifier. This is because there are variable copy numbers of *SMN2* gene in the population and exists an inverse correlation between the numbers of copies and the severity of the disease[13,14].

1.2 CLINICAL CLASSIFICATION

Since the severity of SMA is highly variable among different cases, the International SMA Consortium defined four clinical groups, based on the age of onset and motor function achieved[15] (Tab.1).

SMA type I (MIM #253300) or even called Werding-Hoffmann disease is the most common and severe form of SMA. The incidence is estimated to be between four and ten per 100 000 live births[16].

The onset is from birth to 6 months, affected infants are hypotonic with a symmetric pattern of weakness ("floppy infant"). It is more proximal than distal, with lower limbs usually weaker than upper limbs. Patients with SMA type I are characterised by absence of deep tendon reflexes, breathing, sucking and swallowing difficulties and inability to sit unsupported. Without mechanical ventilation, death occurs before 2 years old. Patients with SMA type I carry only two copies of *SMN2* gene.

More precisely, the most severe form is SMA type 0 or Prenatal SMA, where the intrauterine movements are decreased, severe weakness and joint contractures are common at birth too. SMA type 0 has prenatal onset and patients need respiratory support at birth[17]. About 95% of the affected babies are characterized by the presence of only one copy of *SMN2* and deletion of exon 7 and 8 in both *SMN* copies.

SMA type II (MIM #253550) or Dubowitz disease (chronic spinal muscular atrophy) is the intermediate form of childhood SMA, the age of onset is between 6 and 18 months and individuals survive until 5 years. The most important clinical feature, that is not present in Werding-Hoffmann form, is postural hand tremor. Patients with SMA type II are able to sit unsupported but they cannot walk independently. In this form at least three copies of *SMN2* gene are present.

SMA type III (MIM #253400) or Kugelberg-Welander disease is the juvenile form with age of onset after 18 months and it is characterized by normal lifespan[18]. SMA type III could be divided into IIIa and IIIb, classification based on the age of onset. In type IIIa symptoms appear before 3 years old while in type IIIb after. Patients with SMA type III are able to sit and walk unsupported, but some of them lose the ability to walk during the life. Patients carry from four to eight copies of *SMN2* gene.

SMA type IV (MIM #271150) is the adult form with adult onset and is associated with normal life expectancy[19]. Patients are able to walk and they have not respiratory and nutritional problems. Like in SMA type III, four/eight copies of *SMN2* gene are present.

Type	Name	Age of onset	Clinical features	SMN2 copies	Prognosis
Type 0	**Prenatal SMA**	Prenatal	▪ Decreased fetal movements in utero ▪ Severe weakness	1 copy	Between 2 and 6 months
Type I	**Werding-Hoffmann disease**	0 to 6 months	▪ Floppy infant ▪ Breathing, sucking and swallowing difficulties ▪ Tongue atrophy and fasciculations ▪ Limb and trunk hypotonia ▪ Never sit unsupported	2 copies	Usually die by 2 years
Type II	**Dubowitz disease**	6 to 18 months	▪ Scoliosis, joint contractures and intercostal muscle weakness ▪ Able to sit unsupported ▪ Never walk independently	3 copies	Expected to live into their twenties and beyond
Type IIIa	**Kugelberg-Welander disease**	Before 3 years	▪ Scoliosis, hip abductor weakness ▪ Able to sit and walk unsupported ▪ Someone lose the ability to walk during the life	4/8 copies	Normal lifespan
Type IIIb		After 3 years		4/8 copies	
Type IV		2nd or 3rd decade	▪ Mild motor impairment	4/8 copies	As type III

Tab 1: Clinical classification of SMA disease.

1.2 THERAPEUTIC APPROACHES IN SMA DISEASE

Even if SMA disease is currently incurable, significant progresses in the development of therapeutic strategies have been achieved.

To date, most of the therapeutic strategies are focusing on increasing the expression of SMN protein, since the disease is due to a deficiency of this protein. It has been demonstrated that the increase of SMN protein can be achieved by increasing the transcriptional activity of *SMN2* gene or introducing exogenous *SMN1* gene using viral vector[20]. The most promising vector in SMA disease is adeno-associated viral (AAV) vector based on double strand genome. For the central nervous system (CNS), the vector of choice is a self-complementary adeno-associated virus serotype 9 (scAAV9). This vector has a remarkable capacity to transduce both neurons and astrocytes in adult mice following intravenous delivery[21]. Thanks to gene therapy, mice showed striking life expectation increase. SMA mice normally survive about 13 days while the ones treated with a single systemic delivery of scAAV9 the average lifespans increase to 199 days[22,23].

In addition to gene therapy, to achieve an increase in SMN protein levels it is possible to modulate *SMN2* gene transcription with antisense oligonucleotide (ASO) or drugs to stimulate promoter activity. Several ASOs are tested in preclinical models, they promote exon 7 retention and increase full-length *SMN2* transcript. This is possible for the presence of several splice silencing motifs in *SMN2* that are used as targets for antisense oligonucleotide. A strong intronic splicing silencer, ISS-N1, is located in *SMN2* intron 7 and this motif has been targeted by a number of ASOs with different chemical modifications. Due to the linkage between the specific ASO with ISS-N1, the negative splicing factor is unable to interact with the *SMN2* pre-mRNA leading the inclusion of exon 7 in the transcript followed by a complete functional protein generation.

One of these is phosphorodiamidate morpholino oligonucleotides (PMO), which promotes exon 7 retention in the mature SMN transcripts[24]. It has been shown that one of the PMOs tested increased the level of the full-length transcript of SMN protein *in vitro* and prolonged the survival *in vivo*, rescuing the phenotype in a severe mouse model SMA after intracerebroventricular delivery. Another promising ASO is 2'-*O*-2-methoxyethyl-modified ASO (ASO-10-27), which corrects effectively *SMN2* splicing and restores SMN expression in motor neurons after intracerebroventricular injection. Systemic administration of ASO-10-27 to neonates, robustly rescued severe SMA mice[25,26]. However, even if ASOs are highly effective at promoting inclusion of exon 7 splicing in *SMN2* transcripts and at increasing SMN protein levels both *in vitro* and *in vivo*, they are not capable of crossing the blood-brain barrier. Then, repeated intrathecal administration are required[27]. Another way to modulate *SMN2* transcription is to use drugs to stimulate promoter activity. For example, histone deacetylases inhibitors (HDACIs) are a group of small molecules that promote gene transcription by chromatin remodelling. Various drugs in this class have been studied in SMA. Some HADC inhibitors are able to activate SMN expression *in vitro* and they are able to improve motor performance, attenuate weight loss and increase survival in SMA model mice[28,29]. Worth to mention small molecule therapy with mechanism on modulating SMN2 exon 7 splicing. These studies are sponsored by Norvatis and Roche with promising pre-clinical efficacy[30]. There are currently 3 phase 2 trials using Roche's small molecule RG7916 that is a splicing modifier able to increase SMN2 protein production. In addition, other strategies are being used including neuroprotection with transplantation of neuron stem cells. These have resulted in improved phenotype and extended survival in SMA model mice[31,32].

In December 2016, FDA announced the approval of Spinraza (previously known as Nusinersen) to treat spinal muscular atrophy, an 2'-*O*-methoxyethyl modified ASO (2'MOE)[33]. 1st June of 2017

Spinraza was also approved in the European Union as the first treatment for spinal muscular atrophy[34]. Spinraza is now available on the market from Biogen[35]. 2'MOE is designed to bind to the *SMN2* pre-mRNA complementary to ISS-N1 in intron 7. Since ASOs cannot cross blood brain barriers, to target motor neurons Spinraza has to be administrated via intrathecal injection, which delivers therapies directly to the cerebrospinal fluid around the spinal cord[36]. Treatment with Spinraza reduces the severity of the disease, allowing children with SMA type I to reach development milestones normally associated with less severe types of SMA, like the ability to sit and stand without any support. Furthermore, children survive longer than they would previously succumb to respiratory failure without invasive ventilation. However, this drug is not a complete cure for SMA. Patients treated with Spinraza have improved, however, still defected motor function. Another issue to be solved is how to deliver the drug efficiently into spinal cord in a less invasive delivery method. All these should be considered in future for SMA therapy[37].

1.4 SURVIVAL MOTOR NEURON (SMN)

As shown previously, *SMN* gene exists in two copies, the telomeric copy *SMN1* and the centromeric copy *SMN2*, localised within a duplicated and inverted segment on the long arm of chromosome 5 (5q13), consisting of approximately 500kb. Both the centromeric and telomeric copies consist of nine exons (1, 2a, 2b, 3, 4, 5, 6, 7, 8) and five nucleotides distinguish *SMN2* from *SMN1*, but only one is localised within the coding region. This base pair exchange is the transition C to T at position +6 of exon 7 (c.840C>T). c.840C>T is a silent mutation but results in alternative splicing of exon 7 (SMNΔ7). SMNΔ7 transcript encodes for a shorter protein in which the last 16 residues encoded by exon 7 are replaced by 4 residues encoded by exon 8. This truncated and unstable protein shows a reduced oligomerization capacity, which is essential for proper SMN function[11]. The alteration of the normal splicing with the presence of the transition c.840C>T is due to the disruption of an exonic splicing enhancer (ESE), which is recognised by serine/arginine-rich splicing factor 1 (SF2/ASF), and/or creation an exon splice suppressor (ESS) recognised by heterogeneous nuclear ribonucleoparticule A1/A2 (hnRNP A1/A2). Normally, exon 7 inclusion is due to the binding of SF2/ASF in a sequence encompassing nucleotides + 6 to +11 of exon 7. The transition c.840C>T is followed by a binding decrease of SF2/ASF due to the generation of ESS recognized by hnRNP A1/A2, which antagonize SF2/ASF-dependent ESE activity and promote exon 7 skipping. The skipping of exon 7 occurs in ~90% of mRNA transcripts. Nevertheless, *SMN2* still preserves the ability in ~10% of mRNA transcripts to generate the full-length isoform of SMN (Fig 2).

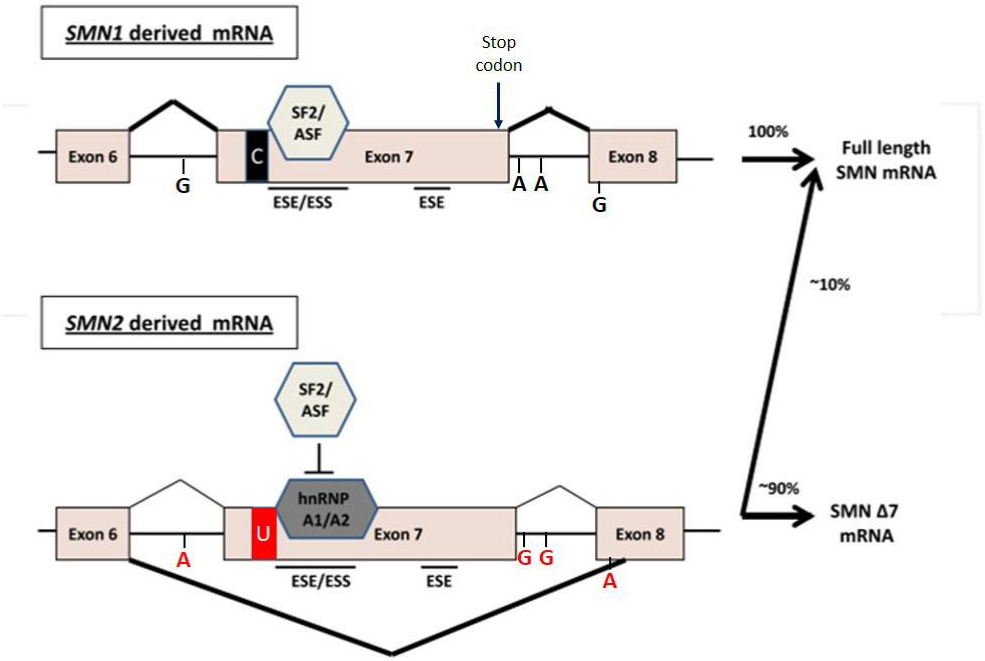


Fig 2: Difference in *SMN1* and *SMN2* splicing. (Image modified from Farooq *et al.*, 2013[38])

Compared to the full-length isoform, SMNΔ7 is characterised not only by the skipping of exon 7, which encodes for 16 amino acid, but also by the addition of four amino acid (EMLA) encoded by exon 8. The presence of these four amino acids can explain the decreased stability of this truncated protein since EMLA represents a degradation signal[39].

1.4.1 SMN protein structure

SMN protein is a ubiquitously expressed protein, consists of 294 amino acids with a molecular weight of 38 kDa. The N-terminus region coded by exons 2A and 2B represents the nucleic acid binding domain, it is conserved and overlaps with the binding site of SMN-Interacting Protein 1 (also known as Gemin 2). The middle of the protein, comprising amino acids 90-160, is characterised by the presence of a Tudor domain that is encoded by exon 3, which binds symmetrically dimethylated arginines, found within many SMN binding proteins. Tudor domain exhibits a conserved negatively charged surface that is shown to interact with C-terminus Arg and Gly-rich (RG and RGG domains) tails of Sm proteins[40]. While downstream of the Tudor domain, SMN contains a proline-rich sequence and a tyrosine/glycine-rich motif (YG box). YG box (encoded by exon 6, from amino acid 258-279) together with the last sixteen amino acid encoded by exon 7 facilitate the self-association of SMN[41,42]. That appears to be crucial for stability and subcellular localisation of SMN. Moreover, QNQKE motif inside the sequence encoded by exon 7 is a nuclear export signal, important for cytoplasmic localisation of SMN and axon function as well[43] (Fig 3).

Fig 3: Domains of SMN protein. (Image modified from Singh *et al.*, 2017)

The overall structure of the protein is critical for the function since small mutations, such as small deletions, splice mutations and missense mutations in these domains have been linked to SMA. This occurs in ~5% of SMA cases. Small mutations are found throughout the length of the protein, but the 3D structure of SMN reveals the mutations to be clustered. Three clustered groups of mutations are presented, one in the N-terminal domain, the second in the Tudor domain and the last within and near the YG box of SMN[41].

1.4.2 SMN protein

SMN protein is localised in the cytoplasm as well as in the nucleus and is particularly abundant in motor neurones of spinal cord[44]. Also, high levels of SMN protein are presented in brain, kidney and liver, while moderate levels in skeletal and cardiac muscle. Both in the nucleus and the cytoplasm SMN form a large multiprotein complex of 1MDa, consist of Gemins 2-8, Unr-interacting protein (UNRIP) and SMN.

Within the nucleus, SMN protein is concentrated in structures called Gems, dot-like structures associated with Cajal bodies (CBs)[45–47]. Gems and Cajal bodies are distinct nuclear structures since they do not colocalize in the fetal tissues but only in some cell lines in adult tissue[48]. The interaction between the two structures may be mediated by the binding capacity of SMN and Coilin (marker of Cajal bodies)[49] (Fig 4).

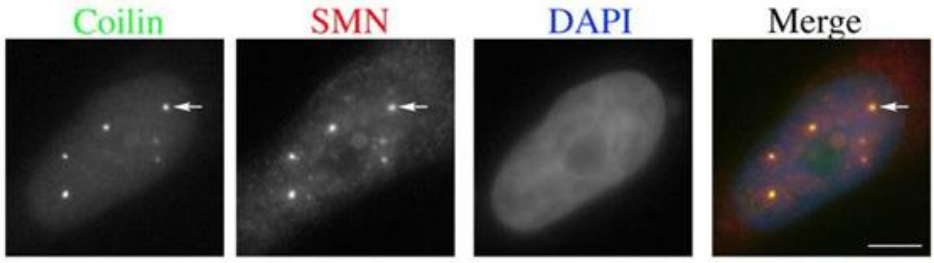


Fig 4: SMN and Coilin staining in Hela cells shows that SMN proteins, concentrated in Gems structure, colocalize with Coilin of Cajal bodies (Hearst *et al.*,2009[50])

Cajal bodies structures were discovered in both plants and animal cells in 1903[51]. CBs are one of the nuclear bodies (NBs) existing, together with splicing speckles, paraspeckles, nuclear stress bodies and histone locus bodies[52–56]. Nuclear bodies spatially compartmentalise the nuclear environment, they are steady-state structures that dynamically respond to basic physiological process[52]. In particular, Cajal bodies are the sites where many proteins and RNA molecules are modified and assembled. CBs represent a meeting place and factory for ribonucleoprotein (RNP)[54]. They contain high levels of factors involved in the transcription and processing of many types of nuclear RNAs, including the three eukaryotic RNA polymerases, small nucleolar ribonucleoproteins (snoRNPs) and small nuclear ribonucleoproteins (snRNPs)[57].

SMN protein is involved in large numbers of different pathways, but its major housekeeping function is the biogenesis of snRNP[58]. In addition to this, SMN has also an essential role in snoRNPs [59], miRNPs[60] and telomerase RNPs biogenesis [61]. Over the last decades, an additional non-canonical function has been identified, that is a role in the assembly of mRNA-binding proteins (mRBPs) and their target mRNAs into a messenger ribonucleoprotein complexes (mRNPs)[62,63]. SMN protein seems to be involved in regulation of the axonal localisation of mRNPs[64–67].

1.4.3 Role of SMN in axonal mRNA metabolism

SMN protein is abundant in axons, dendrites and axonal growth cones of motoneurons. SMN is localised in granules associated with cytoskeletal filament systems, SMN granules are actively transported both retro- and antero-gradely in a microtubule-dependent manner[68–70]. The involvement of SMN protein in axonal transport has been shown in motoneurons isolated from SMA mouse model, which exhibit normal survival but reduced axons growth in which SMNΔ7 is confined to the nucleus[71]. In fact, the distribution of SMN-Gemin complex through the cytoplasm is due to the QNQKE motif (nuclear export signal) that is encoded by exon 7. That motif leads to the exportation from the nucleus to the cytoplasm where SMN is localised in granules within neurites[43]. It has been shown that SMN interacts with some mRBPs and in SMA disease the loss of this interactions can affect targets mRNAs stability and their transport through the axon[72] (Fig 5).

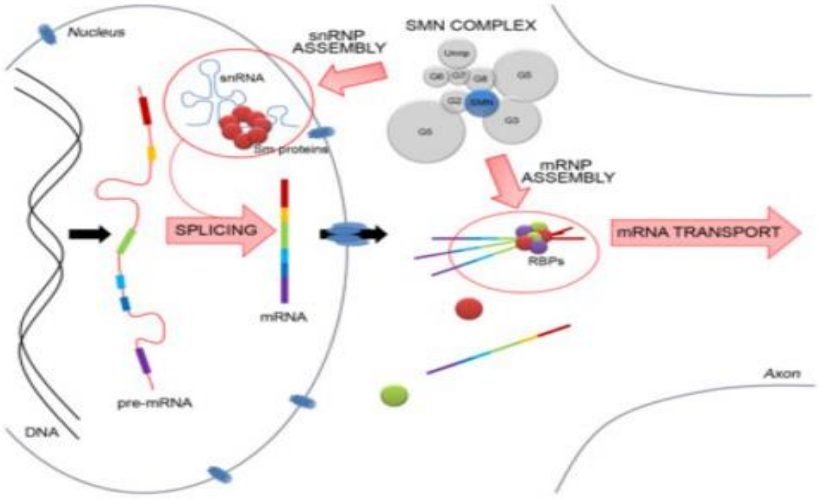


Fig 5: Role of SMN in mRNP assembly (Fallini *et al.*, 2012[63]).

The first clue about the function of SMN in axons came from the significant reduction in β-actin mRNA levels at the axon tip of motoneurons derived from severe SMA mice[71,73]. Translocation of β-actin mRNA at the axon tip results necessary for the neurite outgrowth. Reduction in the level of β-actin mRNA is due to the absence in SMA mice of the linkage between Smn and RNA-binding heterogeneous nuclear ribonucleoproteins R (hnRNP R)[74]. This interaction allows the following association between hnRNP R and β-actin mRNA, which leads to translocation of this mRNA into axons and axon terminals. Therefore, Smn-hnRNP R linkage is crucial for the transport of β-actin mRNA to the distal part of the axon and probably it is important for the control of the stability of β-actin mRNA in motor growth cones. Mutations in hnRNP R, which affect the binding with Smn and β-actin mRNA, lead to the reduction in neurite outgrowth in differentiating PC12 cells line[71]. Smn and hnRNP R interaction has been studied also *in vivo* in zebrafish and *in vitro* and *in vivo* in mice. It has been shown that these two proteins contribute to the differentiation and maintenance of neuromuscular endplate and depletion of Smn or hnRNP R leads to comparable phenotypes[67,75,76]. The interaction described is not the only one to be crucial for β-actin mRNA transport. The insulin-like growth factor mRNA-binding protein 1 (IMP1) is actively transported in motor neuron axon due to the association with the Tudor domain of SMN. IMP1 directly controls the transport and local translation of β-actin mRNA in fibroblasts and neurons and shows a dramatic reduction in motor neurons but not in the brain lysate from SMA mice[72,77]. This suggests that the role of IMP1 in β-actin mRNA transport could be a specific phenomenon occurring to motor neurons and not a general phenomenon occurring in all cell types.

Translation of β-actin mRNA is also under the control of neuronal mRNA binding protein HuD, which interacts with the Tudor domain of SMN and the survival motor neuron protein is co-transported with HuD in axonal granules[65,78] (Fig 6).

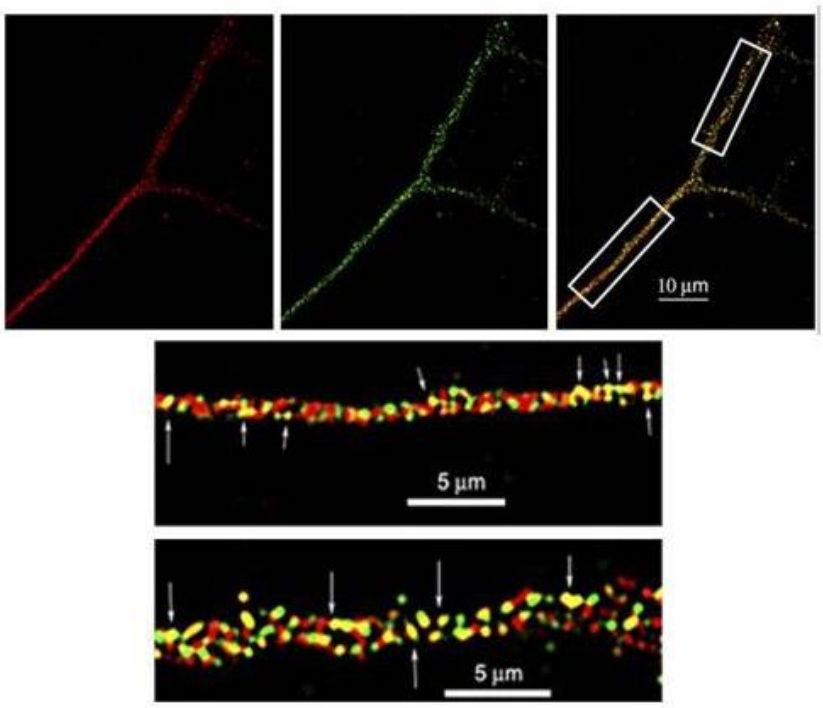


Fig 6: Colocalization of HuD protein (red) with smn granules (green) along the mouse motor neuron axon. (Image taken from Akten *et al*,. 2011[78]).

HuD in neurons is involved in the neuronal development and plasticity. It binds to AU-rich elements in the 3'UTR of target mRNAs and controls their stability and translation[79]. A missense mutation in SMN Tudor domain, which is known to cause SMA, impaired the interaction with HuD and even a deficiency of SMN in primary motor neuron leads to a reduction of HuD levels followed by important impairment of poly(A)-mRNA localisation in axon[80]. One of the HuD targets is β-actin mRNA but is not the only one, other HuD targets are the candidate plasticity-related gene 15 (cpg15) mRNA and growth-associated protein 43 (GAP43) mRNA. It has been shown that cpg15 mRNA interacts with HuD and SMN as well since colocalizes with both proteins in motor neurons. HuD and SMN are involved in the transport of cpg15 mRNA through the axons followed by its translation inside the growth cones. cpg15 is highly expressed in the developing ventral spinal cord and it has a role in the development of motor axons and neuromuscular junctions[78]. Interestingly, neurons with SMN deficiency are characterised by decreased levels of cpg15 mRNA and its overexpression in zebrafish partially rescues SMA phenotype. That shows a possible role of cpg15 mRNA as a modifier of SMA deficiency[78]. Moreover, even the growth-associated protein 43 (GAP43) mRNA is a target of HuD

and it plays a key role in regulating growth cone stability and axon growth and its levels are reduced in motor neurones from SMA mouse model[64,81,82].

These observations show that SMN, in association with the binding of mRNA-binding proteins such as HuD, hnRNP R and IMP1, plays a crucial role in the axonal transport of some mRNAs involved in axon outgrowth.

1.4.4 SMN and snRNPs assembly

Spliceosomal snRNPs assembly is a well-characterized function of SMN, where it is involved in Sm core assembly and in snRNPs nuclear importation.

snRNPs are best known for their role in RNA splicing complexes, involving in the maturation of pre-mRNAs. Spliceosomal snRNP assembly is a multistep process and requires a nucleus and cytoplasmic phase. In most of the eukaryotes, two spliceosomes coexist: the U2-dependent spliceosome, composed of U1, U2, U4/U6 and U5 snRNPs and the less abundant U12-dependent spliceosome, composed of U11, U12, U4atac/U6atac and U5 snRNPs. The first one catalyses the U2-type introns removal, while the other catalyses the splicing of the rare U12-type class of introns[83]. Each spliceosomal snRNP consists of one (U1, U2, U11, U12 and U5) or two (U4/6 and U4atac/U6atac) small nuclear RNAs (snRNAs), eight Sm proteins (B, D1, D2, D3, E, F, and G) and specific polypeptides that are associated with only one individual U snRNP[84].

The first step for spliceosomal snRNP assembly is the transcription of pre-snRNAs that occurs in the nucleus by RNA polymerase II (Pol II), with the exception of U6 snRNA and U6atac generated by RNA polymerase III (Pol III). After transcription, pre-snRNAs are exported to the cytoplasm by their m^7G cap. This structure is recognised by the nuclear cap-binding complex (CBC) composed by CBP20 and CBP80. CBC complex and pre-snRNAs are recognised by the adaptor protein PHAX (phosphorylated adaptor for RNA export) that mediates with the export receptor CRM1 the exportation of newly transcribed snRNAs[48]. In the meantime, in cytoplasm, Sm proteins (B, D1, D2, D3, E, F, and G) interact with chloride conductance regulatory protein (pICIn). This interaction leads to the prevention of premature RNA associations with Sm proteins[85]. pICIn exists in a complex known as the PRMT5 complex, responsible together with PRMT7 for the methylation of specific arginine residues in D1 and D3 Sm proteins allowing the transferring of Sm proteins to SMN complex[86,87], in which SMN directly binds B, D1 and D3 Sm proteins with the Tudor domain. SMN complex consists of UNRIP and seven Gemin proteins (Gemin 2-8). After the interaction of SMN with Gemin proteins, Gemin7 recruits Gemin 6 and UNRIP. In addition to these interactions, self-associations were observed for SMN and Gemin 2. The self-association of SMN proteins occurs with

the N-terminal domain but for the stabilisation of SMN oligomerization is essential Gemin 2-Gemin 2 interaction[88,89]. SMN oligomerization allows the formation of higher-order complexes ranging from 20S to 80S, which is impaired in SMA patients (Fig 6).

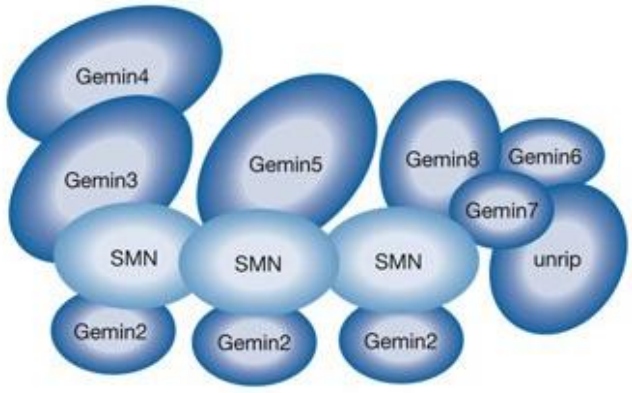


Fig 6: SMN complex. Image taken from Pellizzoni *et al.*, 2007[90].

After the SMN complex formation and the exportation of snRNAs into the cytoplasm, PHAX is dephosphorylated bringing to the release of snRNAs. Now SMN complex plays a key role in assembling Sm proteins onto the just exported snRNAs to produce an active snRNP[84,87,91]. Sm proteins by the linkage with the Sm sites of snRNAs lead to the formation of a heptameric ring around Sm site (Sm core)[92]. A properly assembled Sm core is required for cap hypermethylation and 3'-end maturation.

snRNP assembly is followed by nuclear importation, that occurs after the hypermethylation of m^7G cap to a 2,2,7-trimethylguanosine (TMG) cap and trimming of the 3'-extension. The enzyme responsible for m^7G cap conversion is Trimethylguanosine Synthase 1 (TGS1), also known as PIMT in higher eukaryotes[93,94]. It has been shown that there is an interaction between SMN and TGS1, which could indicate that SMN recruits TGS1 to the snRNAs[95]. This interaction appears to be essential for TMG cap formation[96], that represents together with Sm core a bipartite nuclear localisation signal required to promote the nuclear uptake of snRNPs. The importation inside the nucleus is due to the binding of Snurpotin-1 to TMG cap and the association of SMN with importin β, that is facilitated by WRAP53[97–99] (Fig 7).

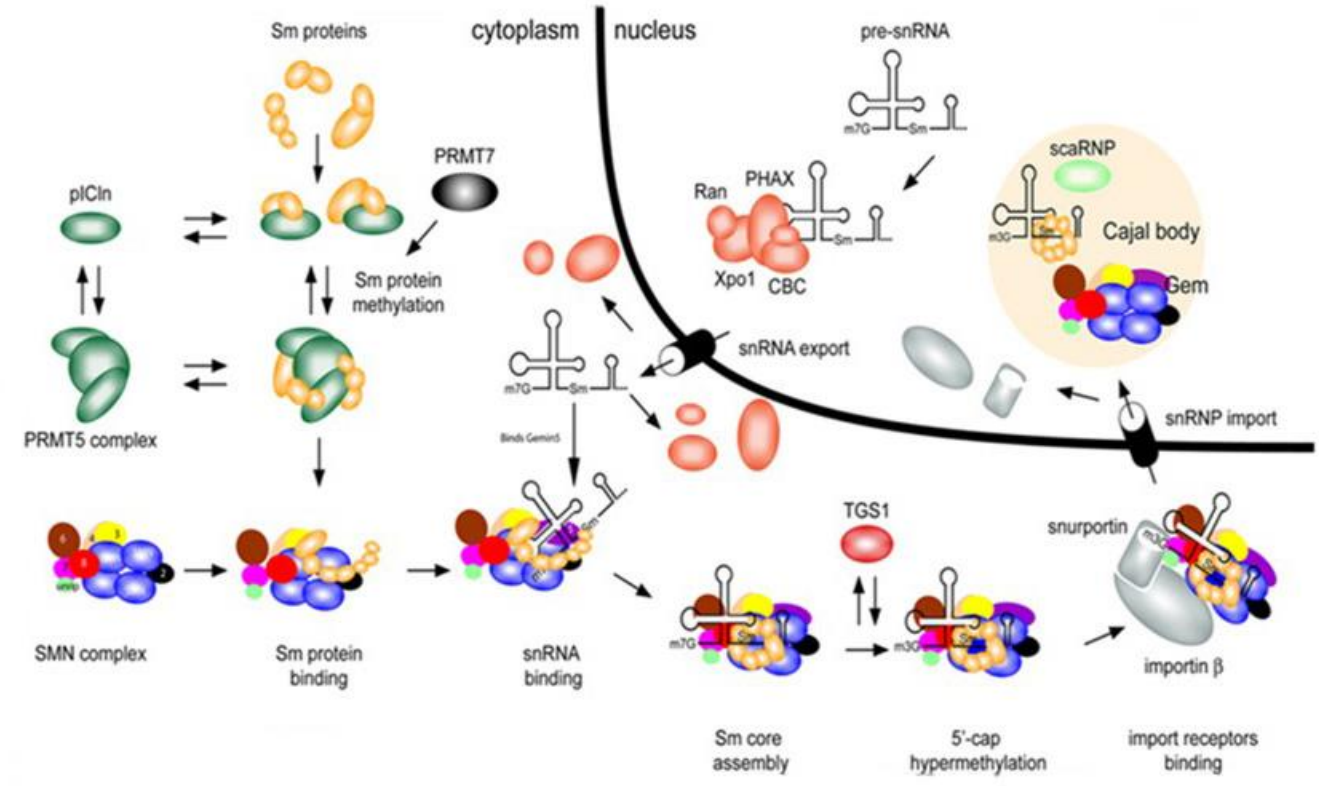


Fig 7: Maturation of snRNPs requires nuclear and cytoplasmatic regulatory steps. (Image taken form Burghes and Beattie 2009[87]).

Once inside the nucleus, the imported complex undergoes to the dissociation and snRNPs and SMN transiently localised to Cajal bodies [93,96,100]. Inside these structures, CB specific ribonucleoproteins (scaRNPs) complete the maturation of snRNPs with 2'-O-methylation and pseudouridylation followed by the association with individual snRNP-specific proteins. After the completed maturation process, that occurs in the Cajal bodies, mature snRNPs are localized to speckles or nucleoplasm, where they are stored until they are required for pre-mRNA splicing[101].

SMN protein hence becomes extremely important for the assembly of snRNPs and the further maturation, for this reason, SMN deficiency causes widespread defects in splicing[42,102]. SMN mutations lead to disruption of the binding between Sm core proteins and snRNAs, followed by a defect in Cajal bodies accumulation of snRNPs and SMN[103,104]. Defects in Cajal bodies accumulation and their disruption are not only due to depletion of SMN protein but also the lack of other proteins involved in snRNPs maturation, like TGS1 and PHAX[105]. Knockdown of any one of these proteins leads to the dispersion of coilin, the marker protein for CBs, into numerous small foci in the nucleoplasm and nucleolus and the colocalization of snRNPs with coilin is lost. This indicates that CBs are lost when the biogenesis of snRNPs is blocked.

1.4.5 SMN and snoRNPs

SMN complex is also involved in the biogenesis of snoRNPs. The small nucleolar ribonucleic protein consists of snoRNA and a number of common core proteins. These complexes are localised in the cell nucleolus and they are required for the maturation of pre-rRNA by chemical modifications of rRNA during ribosome biogenesis (methylation and pseudouridylation) and post-transcriptional processing. snoRNPs are categorised into two main classes based on their function and structure: the box C/D and box H/ACA snoRNAs[106]. The C/D class of snoRNPs carries out 2'-*O-methylation*, whereas the H/ACA class is responsible for pseudouridylation of their RNA targets. C/D snoRNAs are associated with a set of small proteins: 15.5K, Nop56, Nop58 and fibrillarin, while H/ACA snoRNAs are associated to Nhp2, Nop10, Gar1 and dyskerin[107,108]. snoRNAs can be transcribed by Pol II, like snRNAs, but even as polycistronic units. In this case, when snoRNAs are transcribed as polycistronic units they do not require hypermethylation process for completing the maturation. Instead, snoRNAs transcribed by Pol II, m^7G cap needs the hypermethylated by TGS1 that take place into CBs[109,110].

SMN is involved in the snoRNP assembly and their biogenesis inside the CBs[111]. SMN complex interacts with both box C/D and H/ACA snoRNPs, due to the binding with fibrillarin and Gar1 that are specific markers of the two classes of snoRNPs[59]. In contrast with snRNAs, the biogenesis of snoRNAs does not require a cytoplasmic phase, but the maturation happens through Cajal bodies and nucleolus. The assembly of snoRNAs starts in the CBs, but for the localisation of snoRNA precursors inside these nuclear bodies the association of the adaptor PHAX with m^7G cap is required. Inside the Cajal bodies, they acquire TMG cap after hypermethylation process by TGS1[113]. TMG cap is then linked to chromosome region maintenance 1 (CRM1) that is required for the nucleolar transport of the mature snoRNPs[112].

The role of SMN in snoRNPs biogenesis is still not well characterised, SMN might be involved in a late assembly step. It has been shown that SMN is required for accumulation of snoRNPs chaperone Nopp140 in CBs. Nopp140 interacts with both classes of snoRNPs, acting as a chaperone of snoRNPs in yeast and vertebrate cells[114]. Moreover, the presence of Nopp140 into the CBs is inversely correlated with SMA disease severity[115]. However, the clear function of Nopp140 and SMN in CBs is unknown but SMN seems to be important for some aspects of snoRNP biogenesis in CBs that involves Nopp140. Decreased level of Nopp140 in CBs leads to a specific reduction of H/ACA snoRNPs. The reduction of H/ACA versus C/D snoRNPs in SMA cells is due to a tighter association of Nopp140 with box H/ACA snoRNPs compared to the association with box C/D class[114].

In conclusion, SMN protein plays a key role in the assembly and the maturation of both snRNPs and snoRNPs and it is one of the essential proteins for these processes. Another important protein during the biogenesis of both complexes is TGS1, an enzyme responsible for the hypermethylation m^7G cap. With the hypermethylation snRNPs acquire the nuclear localisation signal, which is required to promote their nuclear uptake, while in snoRNPs leads to nucleolar transport of the mature snoRNPs. Therefore, TGS1 plays a critical role in targeting sn- and snoRNA precursors to their final maturation and function site.

1.5 TRIMETHYLGUANOSINE SYNTHASE 1 (TGS1)

TGS1 protein was first discovered in yeast in 2002 as the enzyme responsible for hypermethylation of snRNPs and snoRNPs m^7G cap structure synthesised by RNA polymerase II[93], as well as the telomerase RNA[116] and selenoprotein mRNA[117]. TGS1 is an S-adenosyl-L-methionine (S-AdoMet)-dependent methyltransferase. The gene coding TGS1 was discovered in 2001 and it was called *PIMT* since its ability to interact with the receptor PRIP[118]. This gene maps to chromosome 8q12.1, it covers 52.40kb and it is composed of 13 exons encoding for 853 amino acids.

TGS1 is highly conserved in metazoan, plants, lower eukaryotes, bacteria and archaebacteria. The highest conserved region is the C-terminal part of the protein, which contains the catalytic methyltransferase domain between the amino acid 576-853[119]. While the N-terminal domain varies in size between different species and is required for self-association of the protein. It is not clear if TGS1 association leads a dimer formation or higher order oligomers. This event greatly enhances the efficiency of snRNPs and snoRNPs trimethylation[120]. Furthermore, the N-terminal domain contains two nuclear export signals: NES1 and NES2[109]. NES2 is required for the shuttle of TGS1 between the nuclear and the cytoplasm compartments in a CRM1-dependent manner. In yeast mutant Tgs1 lacking N-terminus domain is localised into the nucleus but fails to be enriched in the nucleolus and also leads to undermethylation of snRNAs and snoRNAs[120].

1.5.1 TGS1 function

TGS1 is involved in methylation steps for the conversion of the m^7G caps to TMG cap structure. Functions conserved in yeast, *Drosophila* and human somatic cells[93,105,121]. Eukaryotic mRNAs, synthetized by Pol II, are characterised by the presence of m^7G cap structure at their 5'-end. The TMG cap is a signature feature of eukarya mRNA that protects against exonucleolytic decay and promotes translation initiation. TGS1 catalyses two successive methy-trasfer reactions from *S*-adenosyl-L-methionine (AdoMet) to the exocyclic NH_2 group of the N7-methylguanine, generating two *S*-

adenosyl-L-methionine molecules (AdoHcy) and the modified $m^{2,2,7}G$ (m_3G)-cap. TGS1 is a guanine specific methyltransferase and compared to the other RNA guanine-N2 methyltransferases the methylation of guanine N2 is strictly dependent on prior methylation of guanine N7[122]. For its activity, TGS1 requires AdoMet as the methyl donor and the binding with snRNP or snoRNP, which represent the methyl acceptor. In snRNP it seems that the SmB/B' protein facilitates hypermethylation of snRNAs by providing a docking site for the MTase[123].

In addition to the hypermethylation of snRNPs and snoRNPs, TGS1 is involved in m_3G cap structure formation in telomerase RNA. The first discovery of this additional function was done in *Saccharomyces cerevisiae* in 2008[116]. Telomerase component 1 (TLC1) encodes the template RNA of telomerase and it shares with snRNA and snoRNA the presence of a m_3G cap structure at the 5' end[124]. It has been shown that the lack of Tgs1 activity leads to a premature ageing due to a modification in telomere length, structure and function. Even in human telomerase RNP biogenesis, TGS1 is involved to TMG cap formation. After the transcription of the human TER, designed as hRT in the original identification[125], and the assembly with H/ACA core proteins, the hRT precursor is then processed at its 3'-end followed by the transport to the Cajal bodies by the transporter factor PHAX and Nopp140[114]. Where, one of the two isoforms of TGS1 (TGS1 short isoform,TGS1 SF) gives hTR its TMG cap[126,127].

1.5.2 TGS1 isoforms

TGS1 exists as two isoforms: full-length isoform (TGS1 LF) with a molecular weight of 110kDa and the shorter isoform (TGS1 SF) with a molecular weight of 65-70kDa. The full-length isoform of TGS1 is located predominantly in the cytoplasm while the shorter is located in the nuclear compartment[95,126]. The nuclear isoform (TGS1 SF) is not generated by alternative splicing but it is the result from a post-translational mechanism, and lacks between 320 and 400 residues in the N-terminal region. With immunoprecipitation experiments, it has been shown that TGS1 SF binds preferentially to common proteins of C/D and H/ACA snoRNPs, while TGS1 LF to snRNPs[126]. TGS1 LF is responsible for the acquisition of the trimethylguanosine cap structure in snRNPs into the cytoplasm. While TGS1 SF is involved in snoRNPs hypermethylation inside the CBs, since snoRNPs do not require a cytoplasmic transit for the acquisition of the trimethylguanosine cap structure[123,128] (Fig 8).

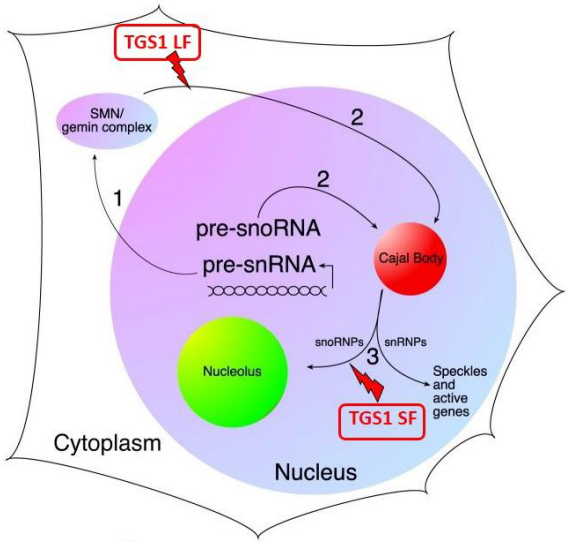


Fig 8: TGS1 LF is responsible for TMG cap formation in snRNPs followed by their uptake into the nucleus while TGS1 SF is involved in the maturation of snoRNPs. With the acquisition of TMG cap structure, snoRNPs complete their maturation into the nucleolus (Image modified from Ogg *et al.*, 2002[129]).

TMG cap formation required first the interaction between TGS1 and snRNPs or snoRNPs. This occurs with N-terminal region of TGS1 full-length for snRNPs and with the C-terminal region of TGS1 short isoform for snoRNPs. With snRNPs, TGS1 LF is able to interact with only the C-terminal extension of SmB and SmD1 proteins[93]. The common feature of SmB and SmD1 compared to the other Sm proteins is a basic rich C-terminal tail. This means that TGS1 interacts with highly charged regions. At the same time, TGS1 is able to bind with C-terminal domain SMN protein. SMN interacts with TGS1 LF in the region encoded by exons 4-7, interaction lost in the spinal muscular atrophy for the inability of SMNΔ7 to bind TGS1 LF[95]. This direct binding between TGS1 LF and SMN highlights a role of the latter in m_3G cap formation. SMN could allow the engagement of TGS1 LF with the m^7G-capped snRNP particle.

For the hypermethylation of snoRNPs the interaction occurs in the C-terminal extensions of C/D and H/ACA box that, like C-terminal tail of snRNPs, is rich in basic amino acid. These interactions with C-terminal extensions of box C/D and H/ACA occurs with C-terminal domain of the short isoform of TGS1[126]. TGS1 interacts with snoRNPs inside the Cajal bodies followed by hypermethylation of m^7G structure. In CBs there is not only the short TGS1 isoform but also the full-length, which binds the C-terminal domain of Nop58 and Nop56 in snoRNPs. This interaction is regulated by CRM1 and this highlights a new role of TGS1 LF in the snoRNPs biogenesis unrelated to cap hypermethylation[109]. TGS1 LF in CBs is not required for the hypermethylation but the binding with C-terminal domain of snoRNPs hides the nucleolar localisation signal (NoLS). With this linkage, snoRNPs are held in the Cajal bodies to complete their assembly. While for hypermethylation is

required the short isoform of TGS1. After cap hypermethylation, NoLS has uncovered thanks to CRM1 which removes TGS1 LF. With the exposure of NoLS, mature snoRNPs are targeted to the nucleolus[109].

1.5.3 TGS1 studies in model organisms

The first mutagenesis studies were performed in yeast in 2002[93]. Yeast snRNPs and snoRNPs acquire their TMG cap structure inside the nucleolus. For a correct localisation of Tgs1 within the nucleolus and efficient sn(o)RNPs trimethylation, the N-terminus region of the protein is required [120]. In N-terminous region a bipartite NLS signal and the domain essential for self-association are present and both are needed for the correct nucleolar localisation and efficient sn(o)RNPs trimethylation activity. Deletion of yeast Tgs1 produces only a cold-sensitive splicing defect but does not interfere with the viability of the yeast[93]. In the absence of Tgs1, the conversion of the m^7G cap to the m_3G structure is missing, resulting in a splicing defect at the restrictive temperature that correlates with the retention of U1 snRNA in the nucleolus. In contrast to yeast, in mouse and *Drosophila Tgs1* is an essential gene. The lack of this protein in mouse leads to early embryonic lethality as well as in *Drosophila* to early pupal lethality[121,130]. In *Drosophila* Tgs1 is required in both muscle and motor neurons to sustain the development of an adult viable organism. Not only disruption but even an overexpression of Tgs1 is toxic and leads to lethality[96].

In vitro experiments in HeLa cells showed that lack of TGS1 is not essential for cell viability but leads to loss of canonical CBs. With the disruption of TGS1, coilin (CBs marker) is distributed throughout the entire nucleus in numerous smaller foci and partially is accumulated in nucleoli, where it forms more predominant structure[105]. Loss of CBs is due to inhibition of cytoplasmatic maturation of snRNPs since they are not able to complete their biogenesis without TMG cap structure.

1.6 PATHOGENESIS

Although the genetic defect in SMA is well defined, its underlying molecular pathway is still not fully understood. For example, how mutations in the ubiquitously expressed SMN genes cause a neurodegenerative condition where only motoneurons become dysfunctional and degenerated while other types of neurons are spared or much less affected. Whether SMA results from a disruption of snRNP biogenesis followed by pre-mRNA misplacing of genes that are critical for the function of the neuromuscular system is still unclear. It is not well understood why the loss of snRNP biogenesis in all cells leads to a selective damage in neuromuscular system, only some suggestions have been proposed[87,97,131]. It is possible that, relative to other cell types, neuromuscular tissues have a greater

requirement or sensitivity for SMN function. An alternative hypothesis that could explain the extreme tissue-specificity observed in SMA patients is that the disruption of snRNPs biogenesis leads to inappropriate or inefficient splicing of some mRNA transcripts crucial for the survival of neuromuscular tissues. Which kind of mRNAs are affected are still to be determined. Interestingly, SMN deficiency changes the snRNP profile of tissues in a non-uniform manner and appears to preferentially reduce accumulation of U12-dependent spliceosomal[28,132]. It has been discovered that SMN is required for U4atac/U6atac/U5 formation since in type I SMA patients is dramatically decreased, even splicing events of some U12-type class of introns is inhibited[133]. Then SMN deficiency perturbs splicing and decreases the expression of a subset of U12 intron-containing genes both in mammalian cells and *Drosophila* larvae[134]. So far, the only SMN-dependent U12 splicing event that could be linked to the SMA phenotype, is the splicing of *stasimon* gene. This gene is a target of SMN and encodes for a protein required for the regulation of synaptic transmission of motor neurons. This event could be linked with SMA phenotypes since the restoration of Stasimon expression corrects defects in motor neurons in both invertebrate and vertebrate models of SMA. But it remains unclear the precise mechanism by which reduced Stasimon perturbs motor circuit activity[134].

In the last decade, another hypothesis is becoming increasingly widespread that is the explanation of SMA phenotypes with the disruption of the non-canonical function of SMN[63,87]. As previously described SMN interacts with some mRBPs as HuD, hnRNP R and IMP1 and in SMA disease loss of these interactions can affect targets mRNAs stability and their transport through the axon[72]. So far, defect in β-actin, cpg 15 and GAP-43 mRNA transport through axon have been described[64,69,78]. All these mRNA are involved in the development and axon outgrowth, then defect in translocation and transduction of these mRNA could explain the defective motor axon growth typically present on depletion of SMN[76].

Moreover, similar phenotype observed with SMN loss is shown with knockout or overexpression of other factors involved in snRNP biogenesis. In addition to SMN, many other proteins are implicated in snRNPs biogenesis. Lack of proteins that are part of SMN complex most likely leads to the motor dysfunction. For example, loss of the orthologue protein of Unrip in *Drosophila* (wmd) gives rise to viability and/or motor defects[96]. Even the disruption in all tissue of the other component of SMN-complex, like Gemin 2, Gemin 3 and Gemin 5 leads lethality in a variety of organisms as yeast, worm, fly and mouse[89]. In *Drosophila* loss of these proteins in all tissues results in death at larval and/or prepupal stages, while if the loss of function is restricted to CNS or muscle confers loss of mobility

and a flightless phenotype. In mouse, loss of function of Gemin 2 in all tissues results in early embryonic lethality while the loss of function in CNS and/or in muscle leads to motor neuron degeneration[135–137]. Not only knockout but also overexpression of SMN complex protein leads to motor defects. Overexpression of Gemin 2 in *Drosophila* is sufficient to depress normal motor function and its enhanced upregulation in all tissues interferes in fly viability[138]. Even in yeast *S.pombe* the toxicity is conserved and Gemin 2 overexpression leads Sm proteins increased in the cytoplasm. Sm proteins accumulation inside the cytoplasm is due to a block in snRNPs biogenesis, followed by a decreased levels of snRNPs in the nucleus[138].

In addition to the proteins that compound SMN-complex, there are other proteins, which are not strictly involved in SMN complex formation but still important for snRNPs biogenesis. Altered expression of these proteins leads to motor dysfunction as well. For example, the disruption of either Tgs1 and plCln mirror the motor phenotypes that is observed with the loss of the constituent members of the SMN-complex[96]. It has been shown in HeLa cells that the absence of TGS1 leads to the same phenotypes observed with the lacking of SMN or PHAX[105]. Lack of function of these proteins results in coilin dispersion followed by the loss of CBs, which could be explained with a block in snRNPs biogenesis[105]. Moreover, overexpression of both Tgs1 and plCln factors has deleterious effects on adult viability and in neuromuscular activity in *Drosophila*, situation analogous to that reports for overexpression of Gemin 2[96,138]. In particular, plCln overexpression results in an accumulation of Sm proteins in the cytoplasm, similar to the overexpression of Gemin 2 in *S.pombe*. Overexpression of Tgs1 in *Drosophila* leads to a failure to contract adequately during puparation as reflected by significant difference in the puparial axial ratio. This phenotype mimics that secondary to disruption of SMN, Gemin 2 or Gemin 3 and overexpression of plCln[96,135,136,138–140]. The similar phenotypic outcomes following disruption of Tgs1, plCln or any member of SMN-complex points out the participation of these factors in a shared pathway.

Therefore, all these studies highlight the importance of snRNP biogenesis, which is required for normal motor behaviour, since disruption and/or overexpression of proteins involved in this pathway results in a motor dysfunction similar to what occurs in SMA disease. Defects in snRNPs biogenesis are the most likely explanation for SMA pathogenesis.

2. AIM OF THE PROJECT

Two genetic variations in *TGS1* gene (NC_000008.11) were discovered in patients by the next generation sequencing. These two patients, who are negative for SMN genetic defects, share SMA-like clinical features. They showed symptoms similar to SMA type IV phenotypes. One of the variations is a missense nucleotide substitution in heterozygous state (p.Gly259Ser), localised in N-terminal region of the protein. While another, is also a missense variation in heterozygous state (p.Asn704Ser) but is localised in the C-terminal region of TGS1.

Trimethylguanosine Synthase 1 (*TGS1*) gene encodes an enzyme responsible for the nuclear importation and maturation of snRNPs. As reported in literature, the disruption of other proteins, correlates with snRNPs biogenesis, can lead to the SMA-like phenotype at least in the model organisms. Thus, the present study was carried out to investigate the possible involvement of this new gene in the pathogenesis of SMA disease. So far there are no studies about that, it is not yet described a possible direct correlation between spinal muscular atrophy and loss and/or gain of function in TGS1 protein. Then the aim of this project was to clarify the possible involvement of *TGS1* in spinal muscular atrophy and to investigate the functional consequences of two different *TGS1* genetic variations in patients with SMA-like phenotypes.

2. MATERIAL AND METHODS

3.1 ETHICS

Human cells were obtained with written informed consent of patients. The Declaration of Helsinki protocols were followed.

3.2 CELL CULTURE MAINTENANCE

Cell lines used:

HEK 293T (Human embryonic kidney cells 293, this cell line expresses a mutant version of the SV40 large T antigen)

SH-SY5Y (Human neuroblastoma cell line)

Fibroblast cell lines:

L1266: the patient's cell line with TGS1 p.Gly259Ser variant.

7384: SMA type III patient with 3 copies of SMN2 gene.

7215: SMA type III patient with 3 copies of SMN2 gene.

6141: SMA type II patient with 2 copies of SMN2 gene.

6142: SMA type II patient with 2 copies of SMN2 gene

6076: healthy control fibroblast.

8230: healthy control fibroblast.

33F: control fibroblast with neuromuscular condition excluded.

Lymphoblastoid cell lines:

BW343: patient's cell line with TGS1 p.Asn704Ser variant.

BW344: cell line with TGS1 p.Asn704Ser variant.

BW345: cell line in which the variant is not present.

CTRL-E: control cell line with neuromuscular condition excluded.

HEK 293T cells, fibroblast cell lines from patients and SH-SY5Y cells were cultured at 37°C in DMEM (D5671, Sigma) supplemented with 10% fetal bovine serum (D8537, Sigma), 1% GlutaMAX™ Supplement (35050061, Thermo Fisher Scientific) and 1% Penicillin-Streptomycin (P4458, Sigma). Cells were split when ~80% confluence was reached.

Lymphoblastoid cell lines were cultured at 37°C in RPMI Medium 1640 (61870010, Thermo Fisher Scientific) and 20% fetal bovine serum (D8537, Sigma). Cells were split every 3-4 days.

3.3 CLONING VECTORS

Wild-type and mutant plasmid carrying p.Gly259Ser mutation were obtained from an Italian research group. TGS1 was cloned into the vector pCDH-CMV-MCS-EF1-Puro under CMV promoter. It is a mammalian Flag-tagged construct carrying ampicillin as bacterial resistance and puromycin as a selectable marker. Molecular cloning was performed in One Shot® TOP10 Chemically Competent E. coli (C404006, Thermofisher). 50ng of each plasmids were used for chemically transformation and pUC19 as a positive control. Vials incubated on ice for 30 minutes and then heated for exactly 30 seconds at 42°C in a water bath followed by 2 minutes on ice. The cells were then mixed in 250µl of pre-warmed SOC medium (15544034, Thermofisher) and incubated horizontally at 37°C for 1 hour at 225rpm in a shaking incubator. After the incubation, 50µl from each transformation were spread on a pre-warmed selective LB agar plate and incubated at 37°C overnight. Each selective LB agar plate (L2897 Sigma) contained 100µl/ml of ampicillin (A9393, Sigma). The following day, single *E.Coli* colonies were picked from agar plates into 5ml LB Broth (L3522 Sigma) containing the appropriate antibiotic (100µl/ml of ampicillin) and incubated at 37°C overnight at 225rpm in a shaking incubator.

DNA was extracted using QIAprep Spin Miniprep Kit (27106, QIAGEN), according to standard protocols and sequencing was performed using 5'-GGTCAAGCACTATCTTCTGAACC-3' primer to check *TGS1* insert. After confirming the presence of *TGS1* insertion inside the colonies picked, QIAGEN Plasmid Maxi Kit (12163, QIAGEN) was used to obtain a high amount of plasmids for a lentivirus generation.

3.4 LENTIVIRAL VECTORS (PCDH-CMV-MCS-EF1-PURO)

Lentiviral vectors (pCDH-CMV-MCS-EF1-Puro) were produced by co-transfecting 1.8 x 10^7 HEK 293T cells with transgene plasmids and the packaging and enveloping constructs, pCMV delta R8.2 and pMD2.G, respectively (Addgene). For 1.8 x 10^7 HEK293T cells plated in a 15 cm dish, 40µg plasmid DNA were co-transfected with 30µg of pCMV delta R8.2 and 10µg of pMD2.G. DNA was mixed in 5ml Opti-MEM® (31985-070, Thermofisher) and combined with 5ml Opti-MEM® containing 1µM polyethyleneimine (408727, Sigma). The mixtures obtained was applied to HEK 293T cells after 20 minutes incubation at room temperature. After 4 hours of incubation, the mixture was removed and fresh medium was added to the cells. Medium containing the virus was collected

after 48 hours post-transfection. After each collection, the supernatant was filtered through a cellulose acetone membrane (0.22µm pore) and then was centrifuged (4°C) at 21900 X g for 20 hours. Following centrifugation, the supernatant was removed and the pellet was resuspended in 50µl of Opti-MEM®.

3.4.1 Virus titration

Virus titration was done in frozen samples in order to obtain an accurate result since lentivirus titer decreasing after defrosting. For virus titration, 2 x 10^5 HEK 293T cells were plated into each well of a 6 well plate and transduced with a range of volumes of the concentrated lentivirus. After seventy-two hours of transduction, genomic DNA was extracted with DNeasy Blood & Tissue Kit (69504, QIAGEN) and the proviral titre was calculated by qPCR. 20ng/µl of the sample was used in qPCR and each sample was run duplicately together with a standard curve. TaqMan Universe PCR MasterMix (4304437, Thermo Fisher Scientific) was used in qPCR, together with 5µM of each primer and the probe.

3.4.2 Transduction experiments

Transduction was performed in twelve-well matrigel coated plates. SH-SY5Y cells were seeded at 1x10^5 cells in each well. The following day, wild-type and mutant lentivirus vectors were used with a multiplicity of infection (MOI) of 1, 15 and 70. Cells were transduced for 24 hours before being replaced with the medium containing puromycin for 10 days. Survival cells were then isolated.

Transduction of fibroblast cell lines was performed in twelve-well matrigel coated plates. Fibroblast cells were seeded at a density of 5x10^4 cells per well. The day after, cells were transduced for 24 hours with a multiplicity of infection of 1, followed by puromycin selection for 10 days. Survival cells were then isolated.

3.4.3 Puromycin selection

In order to figure out the optimum concentration of puromycin (P8833, Sigma) for SH-SY5Y cells and fibroblast cells, 15 different concentrations were used for each cell lines. Cells were seeded into a 24 well plate in order to have 30-40% confluent at the following day. 15 different concentrations, from 0.1µg/ml to 2µg/ml of puromycin were used for fibroblast cells while from 1µg/ml to 6µg/ml for SH-SY5Y cells. Medium with antibiotic was replaced every 2 days. The optimum concentration to be used for each cell lines is the lowest concentration which kills 100% of untreated control cells exactly at day 5.

3.4.4 *In vitro* SH-SY5Y cells differentiation

Transduced SH-SY5Y cells were seeded at a density of $1x10^4$ cells in a 6-well matrigel coated plate with DMEM 10%. The day after media was replaced by complete Neurobasal ™ medium (21103049, Thermo Fisher Scientific) supplemented with 1% GlutaMAX™ Supplement, 2% B27 (17504044, Thermo Fisher Scientific) and 10µM of fresh retinoic acid (R2625, Sigma-Aldric). Complete Neurobasal media were replaced every 2 to 3 days.

3.5 LENTIVIRAL VECTORS WITH GREEN FLUORESCENT PROTEIN

Three other lentiviral vectors, produced by a scientist at the Institute of Child Health, were used to transduce both SH-SY5Y cells and fibroblast cell lines:

- pCCL-SFFV-TGS1.WT-IRES-EGFP (WT vector)
- pCCL-SFFV-TGS1.p.Gly259Ser-IRES-EGFP (MUT vector)
- pCCL-SFFV-EGFP (CT vector)

These lentivirus vectors differed from the previous ones in the presence of enhanced green fluorescent protein (EGFP) tag instead of FLAG tag. Moreover, the full-length *TGS1* was under the control of the spleen focus forming virus (SFFV) promoter instead of CMV promoter. The third lentivirus vector was used as a negative control since it expresses only the green fluorescent protein and not full-length *TGS1*. These three vectors were trited by flow cytometric readout of GFP-positive cells, following HEK 293T transduction. Vector tritation was completed by flow cytometry: $1x10^5$ HEK293T cells were plated into each well of a 6-well plate and transduced with a range of volumes of concentrated lentivirus. After 72 hours of transduction, cells were trypsinized and EGFP-positive cells were quantified.

3.5.1 Transduction experiments

SH-SY5Y cells were plated at a density of $1x10^5$ cells in six-well plates. The following day the three lentiviruses were then added to give a multiplicity of infection of 10. After 24 hours, the medium was changed and cells sorting were performed after 3 days of transduction. GFP-positive cells were sorted on a MoFlo sorting machine. Cells sorting allowed to obtain two different cell populations: cells with low and high GFP expression. Then, transduced cells were plated at a density of $1x10^4$ cells in a 6-well matrigel coated plate with DMEM 10%. The day after media was replaced by complete Neurobasal ™ medium supplemented with 1% GlutaMAX™ Supplement, 2% B27 and 10mM of fresh retinoic acid. Complete Neurobasal media were replaced every 2 to 3 days.

3.6 IMMUNOSTAINING

For TGS1 and SMN immunofluorescence staining in fibroblast cell lines from patients, controls and in transduced fibroblast cells, $2x10^4$ cells were seeded in 8-well chamber slides with 500µl of DMEM 10%. After 48 hours, cells were fixed in 4% paraformaldehyde (15710, Emigrd) diluted in PBS (D8537, Sigma) at room temperature (RT) for 15 minutes. Ice cold methanol were added in fixed cells for 1 hour at -20°C. Then cells were washed with PBS 1X for 5 minutes and blocked in 5% goat serum (G9023, Sigma) diluted in PBS 1X for 1 hour at RT. After the blocking, cells were incubated with mouse anti-SMN antibody (1:200) (610646, BD Transduction Laboratories™), rabbit anti-TGS1 antibody (1:100) (HPA025024, Atlas Antibodies) diluted in PBS/ 0.1% Triton X-100 (X100, Sigma) for 1 hour at RT. The cells were then washed with PBS 1X three times for 5 minutes each and incubated with Alexa 594 goat anti-mouse (1:200) (R37121, Thermo Fisher Scientific) and Alexa 488 goat anti-rabbit (1:200) (R37116, Thermo Fisher Scientific) secondary antibodies for 1 hour at RT. They were washed three times for 5 minutes each with PBS 1X and mounted with mounting medium/DAPI.

For TGS1 and SMN immunofluorescence in differentiated SH-SY5Ycells, the procedure is almost the same as described above. There was only one difference in the fixation procedure. The cells were seeded onto a matrigel-coated glass coverslips at $1x10^5$ cells in 12-well plates. The day after, they were fixed in pre-warmed 2% paraformaldehyde in PBS 1X for 10 minutes at RT followed by 30 minutes at -20°C with ice cold methanol. Fixed cells were stained as described above.

For β-tubulin III immunofluorescence staining in differentiated SH-SY5Y cells, $1x10^4$ cells were seeded onto matrigel-coated glass coverslips in 12-well plates. The following day, cells were washed with PBS 1X once and then fixed in pre-warmed 2% paraformaldehyde in PBS 1X for 10 minutes at RT. Fixed cells were blocked in 5% goat serum diluted in PBS 1X for 1 hour at RT. After the blocking, cells were incubated with mouse anti- β-tubulin III antibody (1:500) (MAB1195, R&D System) for 1 hour at RT. The cells were then washed with PBS 1X three times for 5 minutes each and incubated with Alexa 594 goat anti-mouse for 1 hour at RT. They were washed three times for 5 minutes each with PBS 1X and mounted with mounting medium/DAPI.

Images were captured with a Leica microscope using Metamorph software.

3.6.1 Gems count

SMN staining was followed by gems count. This analysis was assessed in two different ways, each performed duplicately. The counts were performed at a magnification of 63X in 200 cells for each

cell lines. The first method used was to count the number of gems in 200 cells. While the second was to count the number of positive cells, that means cells with at least one gem inside the nucleus, in 200 cells.

3.7 PROTEIN EXTRACTIONS

For whole cells extraction, cells were counted, washed with PBS and then lysed in an appropriate volume of lysis buffer. As a ratio cell/lysis buffer: approximately 10-20 x 10^6 cells per 200µl lysis buffer was used. The buffer used consist of: RIPA buffer (R0278, Sigma) + 75mMTris/HCl SDS 1% + protease inhibitor cocktail buffer (11697498001, Roche). Samples were centrifuged for 10 minutes at 12000 g at 4°C. The supernatant was transferred to a fresh tube and the protein concentration was determined by Pierce™ BCA Protein Assay Kit (23225, Thermo Fisher Scientific) according to the manufacturer's protocol. Then protein concentration was measured with BioPhotometer Eppendorf.

3.7.1 Western blot analysis

Protein extracts were fractionated by NuPAGE ™ 4-12% Gel (NP0322BOX, Thermo Fisher Scientific) using 1 X NuPAGE ™ MOPS SDS Running Buffer (NP0001, Thermo Fisher Scientific). This buffer was used for fill the external chamber, while in the inner chamber was added Running Buffer 1X with 0.25% NuPAGE ™ Antioxidant (NP0005, Thermo Fisher Scientific). 40µg of protein samples were mixed with 1X NuPAGE ™ LDS Sample Buffer (NP0007, Thermo Fisher Scientific) and 1X NuPAGE ™ Sample Reducing Agent (NP0009, Thermo Fisher Scientific). Then the samples were loaded on 4-12% gel with SeeBlue ™ Pre-staining Protein Standard (LC5625, Thermo Fischer Scientific). Gels were run at 150 V for 90 minutes before blotted onto polyvinylidene difluoride membrane (03010040001, Roche) at 350mA for 2 hours using 1X NuPAGE ™ Transfer Buffer (NP00061, Thermo Fischer Scientific), 5% methanol and 0.1% NuPAGE ™ Antioxidant. The membranes were blocked in Odyssey blocking solution (P/N 927-40100, LI-COR Biosciences) for 1 hour at RT, before incubated overnight at RT with primary antibodies:

- Rabbit anti-TGS1 antibody (1:500) (HPA025024, Atlas Antibodies)
- Mouse anti-SMN antibody (1:500) (610646, BD Transduction Laboratories ™)
- Mouse anti-β-tubulin antibody (1:3000) (236-10501, Thermo Fischer Scientific)

The following day membranes were washed three times for 10 minutes with PBS containing 1% Tween 20 (PBST) before incubation with secondary antibodies for 1 hour at RT.

- Anti-rabbit secondary antibody, HRP (1:50000) (65-6120, Thermo Fischer Scientific)

- Anti-mouse secondary antibody, HRP (1:50000) (65-6520, Thermo Fischer Scientific)

Membranes were washed three times for 10 minutes at RT with PBST and Clarity ™ Western ECL Substrate (1705060, Bio-Rad) was used according to the manufacturer's instructions and visualised with ChemiDoc Imaging System (Bio-Rad).

3.8 3D MODELING

TGS1 3D structure was predicted by utilizing Automated Mode of SWISS-MODEL software (http://swissmodel.expasy.org/). The 3D structure was generated with the alignment of methyltransferase domain of human trimethylguanosine synthase 1 (3GDH) and it was displayed by PyMol (PyMOL v.1.7.0.1; Python v.2.7.2).

4. RESULTS

4.1 PROTEIN EXPRESSION

4.1.1 SMN expression

SMN expression levels were analysed only in fibroblast cell lines and not in lymphoblastoid cell lines. Western blot of SMN from whole cell lysates showed no reduction in L1266 (TGS1 patient carrying p.Gly259Ser mutation) cell line (Fig.9).

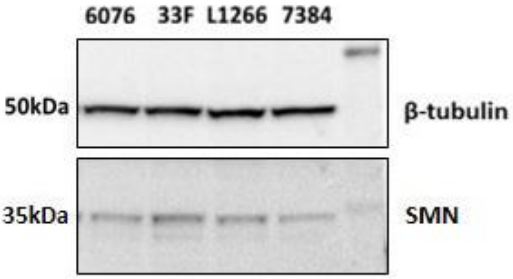


Fig.9: Western blot shows SMN protein expression levels from whole cell lysate in TGS1 patient carrying p.Gly259Ser (L1266), SMN patient (7384) and two healthy controls (6076, 33F). Results obtained with β-tubulin are shown as controls for the amount of protein loading in each lane.

4.1.2 TGS1 expression

TGS1 expression levels were detected by western blot analysis from whole cell lysate in both TGS1 patients and relative controls. By staining with the Atlas TGS1 antibody, two bands of TGS1 protein were detected, at approximately 96kDa and 65kDa. The top band is supposed to be the full-length isoform present mainly in the cytoplasm while the lower band could be the short isoform located in the nucleus. β-tubulin was used as sample loading control. No difference of TGS1 expression was observed in L1266 patient carrying p.Gly259Ser mutation and 343 patient carrying p.Asn704Ser mutation compared to the healthy controls. Interestingly, SMA patient (7384) showed a decreased expression of the short isoform of TGS1, when compared to both L1266 and the other two controls (6076 and 33F) (Fig.10).

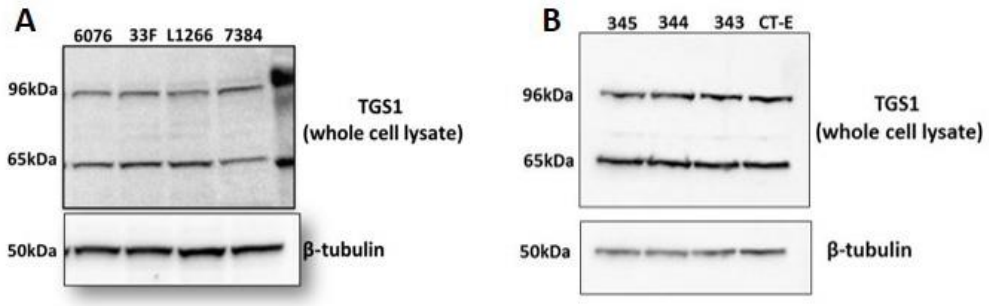


Fig.10: Panel A: Western blot shows TGS1 protein expression levels from whole cell lysate in TGS1 patient carrying p.Gly259Ser (L1266), SMN patient (7384) and two healthy controls (6076, 33F). Results obtained with β-tubulin are shown as controls for the amount of protein loading in each lane. Panel B: Western blot shows TGS1 protein expression levels from whole cell lysate in TGS1 patient carrying p.Asn704Ser (343), in cell line with p.Asn704Ser mutation (344) and two healthy controls (345, CT-E). Results obtained with β-tubulin are shown as controls for the amount of protein loading in each lane.

Western blot analysis was extended further to other SMN patients: two SMA type II patients (6141 and 6142) and one SMA type III patient (7215). As in 7384 SMA sample, the sample 6141 showed a decreased expression of the short isoform of TGS1. The ratio of the full-length isoform to the short isoform highlighted the decreased expression of the short isoform in two of four SMN patients (Fig.11).

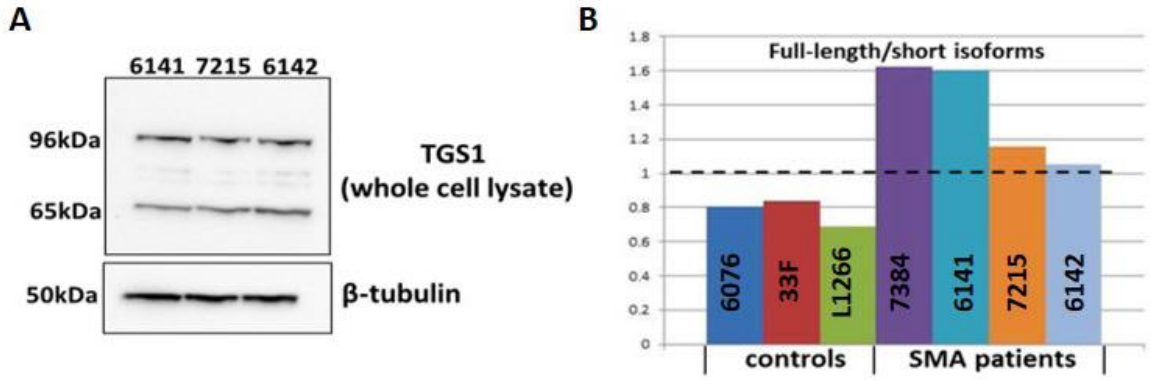


Fig.11: Panel A: Western blot shows TGS1 protein expression levels from whole cell lysate in three SMN patients (6141, 7215, and 6142). Results obtained with β-tubulin are shown as controls for the amount of protein loading in each lane. Panel B: Ratio of the full-length isoform to the short isoform of TGS1.

4.2 IMMUNOFLUORESCENCE

Immunofluorescence experiments were performed only in the fibroblast cell lines and not in lymphoblastoid cell lines.

4.2.1 TGS1 staining

With TGS1 immunostaining in fibroblast cell lines, no obvious difference was observed between L1266 and the three control cell lines. The staining showed that TGS1 is evenly distributed both in nucleus and cytoplasm. However, a decrease of the nuclear staining of TGS1 short isoform was observed in 7384 and 6141 samples (Fig.12). Immunostaining results with TGS1 antibody confirmed the western blot results, which showed a reduction of the short isoform, corresponding to the nuclear isoform, in these two SMA patients.

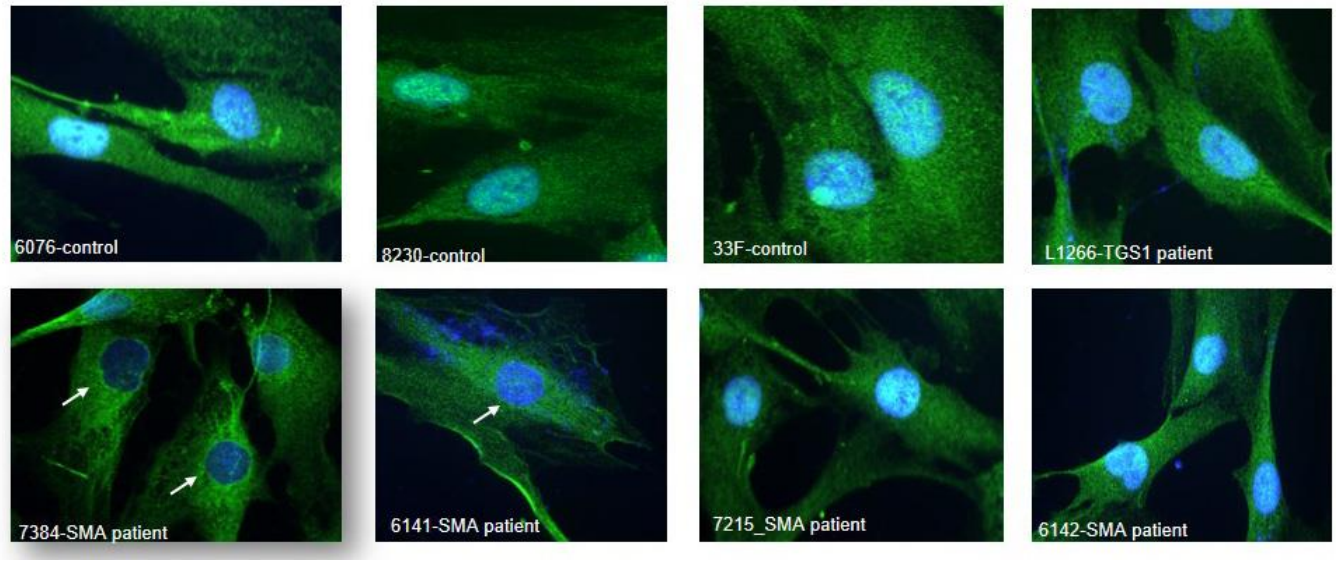


Fig.12: Staining with TGS1 antibody in fibroblast cell lines: three healthy controls (6076, 8230 and 33F), TGS1 patient carrying p.Gly259Ser mutation (L1266) and four SMA patients (7384, 6141, 7215 and 6142). The arrows highlight the fibroblast cells with lower TGS1 short isoform expression. Images took at 63x magnification.

4.2.2 SMN staining and gems count

SMN staining was performed only in fibroblast cell lines due to the small dimension of lymphoblastoid cell lines in which it was not possible to highlight any gems. Three healthy controls, one SMN patient and TGS1 patient carrying p.Gly259Ser mutation (L1266) were used for the staining. As expected, SMN patient showed no gems or very low gem numbers. Interestingly, L1266 sample showed a significant increase of gem numbers inside the nucleus, with gems bigger compared to controls (Fig.13).

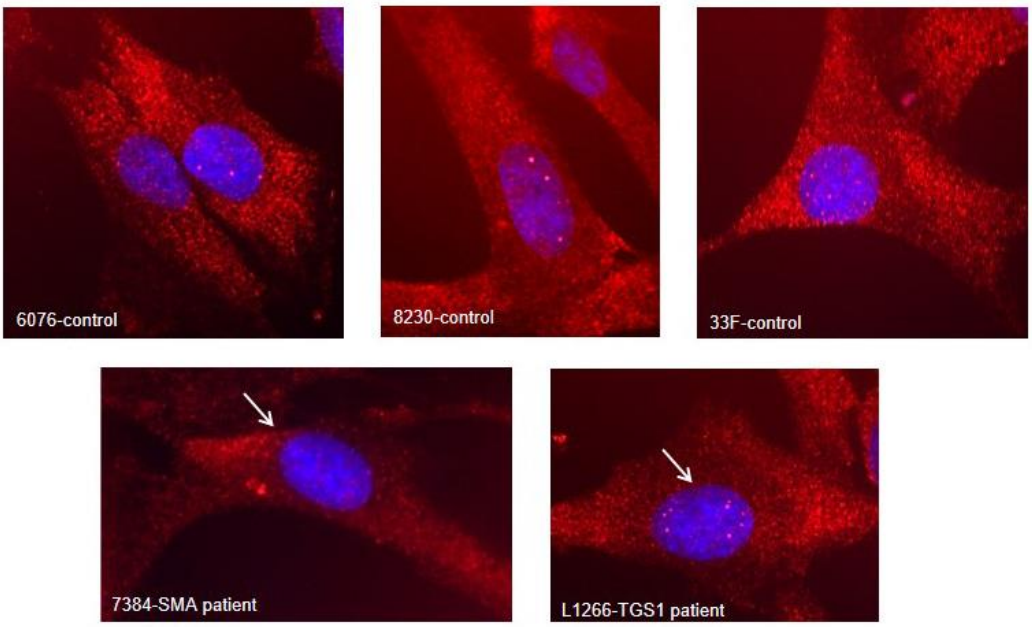
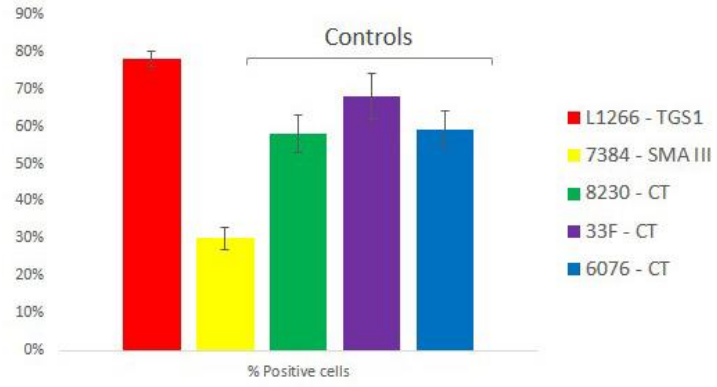


Fig.13: SMN staining in three healthy controls (6076, 8230 and 33F), one SMN patient (7384) and TGS1 patient carrying p.Gly259Ser mutation (L1266). The arrow in 7384 sample highlights the absence of gems inside the nucleus while in L1266 patient highlights the presence of numerous gems. Images took at 63x magnification.

In order to confirm the increased amount of gems number, gems count and the number of positive cells (at least one gem inside the nucleus) were performed in 200 cells. These analyses were performed twice and showed a significant increase of gems number in L1266 fibroblast cell line, which is 40% higher than controls. While SMA patient (7384) showed a reduction of gems number of 50% of the controls. Moreover, the count of positive cells highlighted that the percentage of cells with at least one gem inside the nucleus, as for gems count, is less than 50% compared to controls in 7384 sample. Whereas, the percentage of positive cells in L1266 sample is only 20% higher than controls. This means that in L1266 sample the higher number of gems is due to an increased gems number inside each nucleus differently from 7384 sample, in which the lower gems number is due to a reduction of 50% of positive cells (Fig.14).

A

	Number of cells	% Positive cells
L1266 - TGS1	200	78%
7384 - SMA III	200	30%
8230 - CT	200	58%
33F - CT	200	68%
6076 - CT	200	59%



B

	Number of cells	Nuclear gems
L1266 - TGS1	200	456
7384 - SMA III	200	140
8230 - CT	200	293
33F - CT	200	335
6076 - CT	200	309

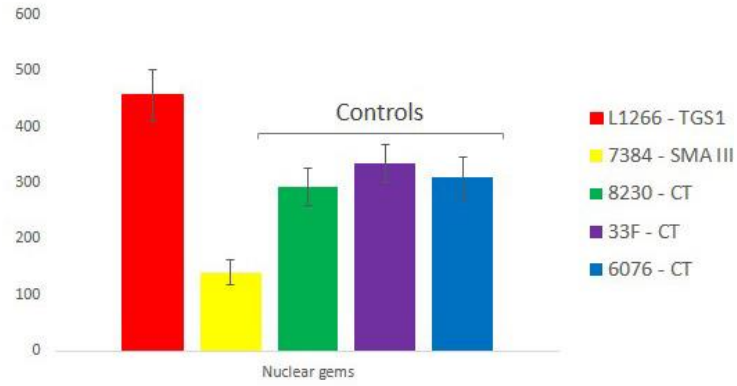


Fig.14: Panel A: Histograms show the percentage of positive cells in L1266 sample (shown in red), SMA patient (shown in yellow) and three healthy controls. Count performed in 200 cells in two different staining. Panel B: Histograms show the gems number in L1266 sample (shown in red), SMA patient (shown in yellow) and three healthy controls. Count performed in 200 cells in two different staining.

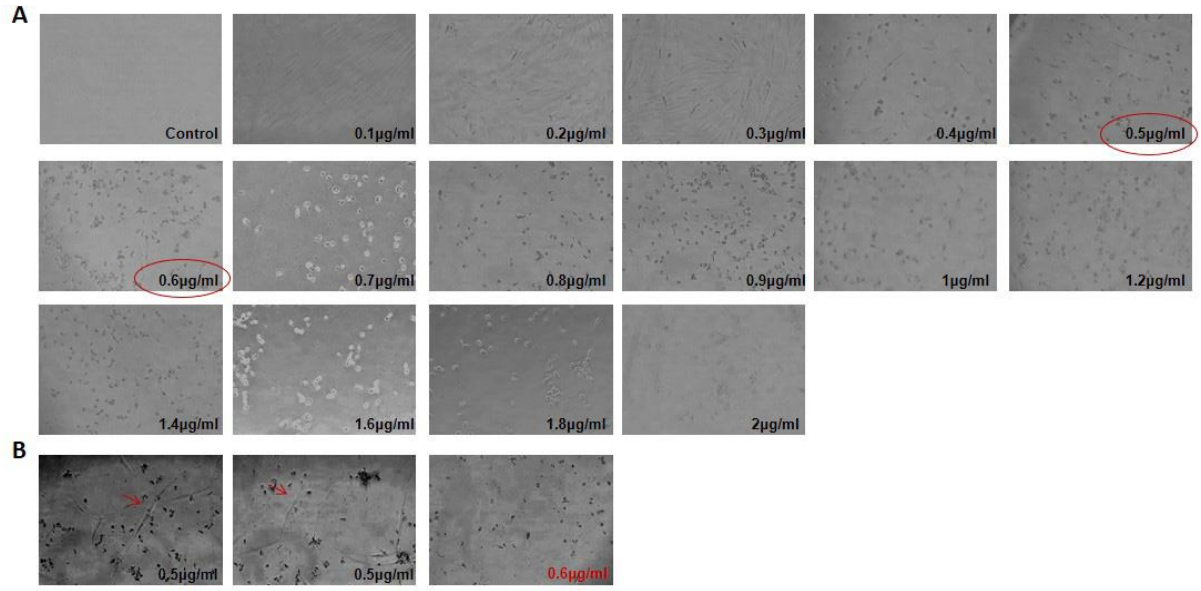
4.3 LENTIVIRUS (PCDH-CMV-MCS-EF1-PURO)

4.3.1 Lentivirus titration

Proviral titre was calculated by qPCR that allowed to measure the number of integrated DNA lentiviral copies per cell. The number of lentiviral vector copies was calculated by absolute quantification. Absolute DNA titers were achieved by comparing crossing point values derived from DNA samples to those obtained from a standard curve of known concentrations of plasmid lentivirus DNA. Lentiviral titer obtains by qPCR was of $1.71x10^7$ in lentivirus carrying WT construct while $7.38x10^6$ in lentivirus carrying construct with p.Gly259Ser mutation.

4.3.2 Puromycin selection in SH-SY5Y cells and fibroblast cells

The optimum concentration of puromycin in SH-SY5Y cells and fibroblast cells was obtained using 15 different concentrations of puromycin antibiotic, from 0.1µg/ml to 2µg/ml for fibroblast cells and from 1 µg/ml to 3.8 µg/ml for SH-SY5Y cells. The right antibiotic concentration was the lowest concentration that killed all cells exactly at day 5. After day 5, all SH-SY5Y cells died with 1.4µg/ml and 1.6µg/ml of puromycin, while fibroblast cells died with 0.5 µg/ml and 0.6µg/ml. In order to be sure that no more cells were still alive in these wells, DMEM 10% was added for 6 days. After 6 days, SH-SY5Y cells with 1.4µg/ml and fibroblast cells with 0.5µg/ml started to grow. While no cells were detected in wells with 1.6µg/ml and 0.6µg/ml of puromycin. Then, the optimum concentration of puromycin was 1.6µg/ml for SH-SY5Y cells and 0.6µg/ml for fibroblast cells (Fig.15)



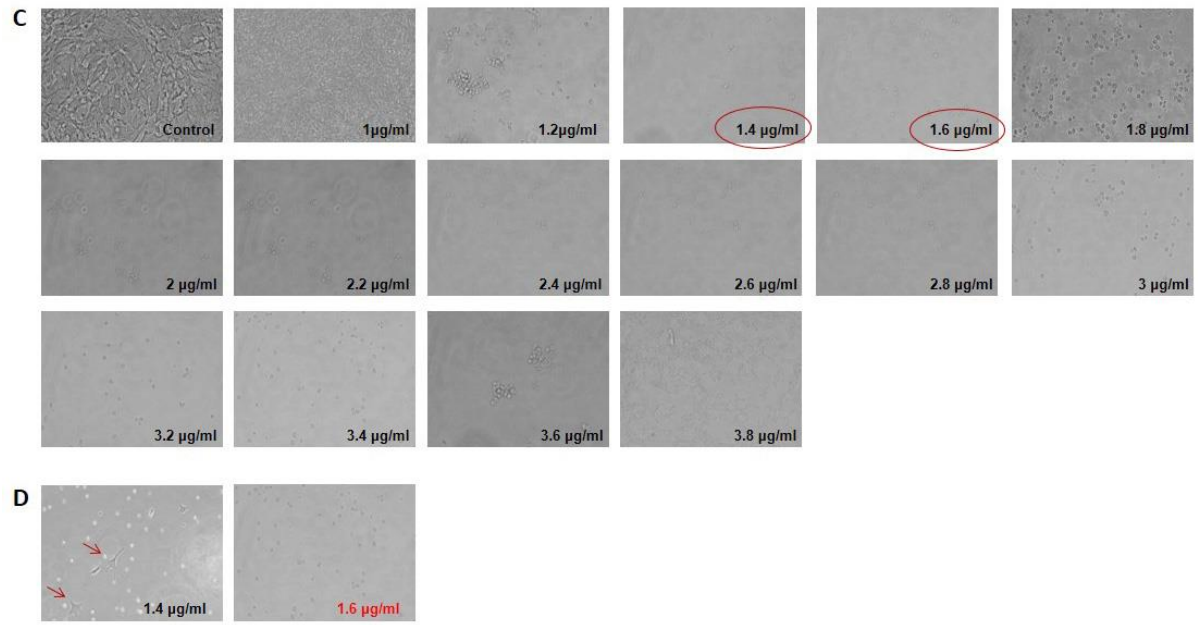


Fig.15: Panel A: Puromycin selection in fibroblast cell line at day 5. Panel B: Fibroblast cells selected with 0.5 µg/ml and 0.6µg/ml of puromycin at day 5 followed by 6 days in DMEM 10%. In red is shown the optimum concentration of puromycin antibiotic for the selection of fibroblast cell line. Panel C: Puromycin selection in SH-SY5Y cells at day 5. Panel D: SH-SY5Y cells selected with 1.4 µg/ml and 1.6µg/ml of puromycin at day 5 followed by 6 days in DMEM 10%. In red is shown the optimum concentration of puromycin antibiotic for the selection of SH-SY5Y cells.

4.3.3 Transduction experiments in SH-SY5Y cells

SH-SY5Y cells were transduced with the two lentiviruses carrying wild-type and mutant constructs using MO1 1, MOI 15 and MOI 70. After 24 hours of transduction, cells were selected with medium containing 1.6µg/ml of puromycin for 10 days. The presence of the constructs inside the cells was detected by DNA and cDNA samples sequencing. In both DNA and cDNA samples, a couple of primers were designed to amplify specifically the coding region between exon 3 and exon 5, where the p.Gly259Ser variant is localised. Amplification of DNA samples with primers specifically designed for cDNA has allowed confirming the integration of the two constructs within the DNA of SH-SY5Y cells (Fig.16).

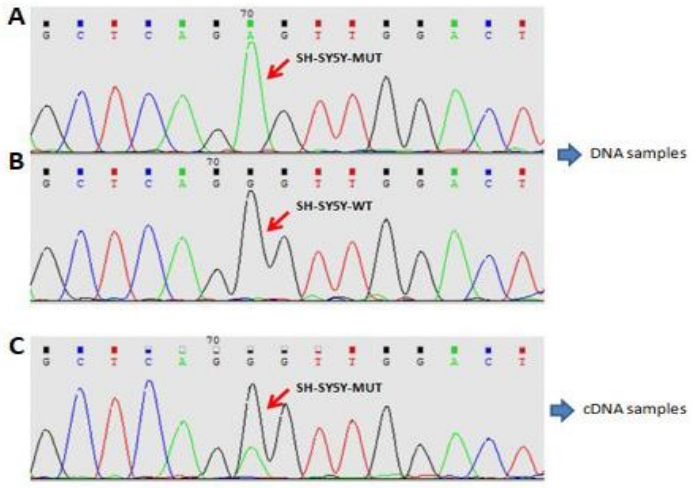


Fig.16: Panel A: DNA sequence of SH-SY5Y cells transduced with lentivirus carrying the mutant construct. Electropherogram shows the variant c.775G>A in homozygous state. Panel B: DNA sequence of SH-SY5Y cells transduced with lentivirus carrying the wild-type construct. Electropherogram shows the wild-type sequence. Panel C: cDNA sequence of SH-SY5Y cells transduced with lentivirus carrying the mutant construct. As expected electropherogram shows the variant c.775G>A in heterozygous state.

The selection of cells with puromycin was followed by differentiation with the complete neurobasal medium. SH-SY5Y cells were cultured for 10 days prior to being utilized for further experiments. In the first experiments performed, retinoic acid was not used for the differentiation but after βIII-tubulin staining, not all cells were differentiated. Then, retinoic acid was added in the complete neurobasal medium in order to promote the differentiation.

4.3.3.1 Transduction experiments in SH-SY5Y cells with MOI 1

At the beginning, transduced experiments were performed with MOI 1 but no significant differences were observed in SH-SY5Y cells transduced with wild-type and mutant lentiviral vectors compared to untreated cells. Lentiviral transduction was followed by differentiation for 10 days, however, no variations in neurite outgrowth in both wild-type and mutant SH-SY5Y cells compared to untreated cells were observed (Fig.17).

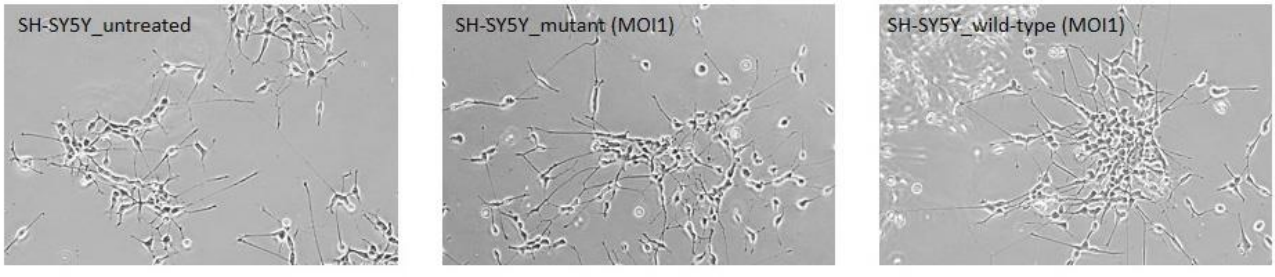


Fig.17: Untreated SH-SY5Y cells and SH-SY5Y cells transduced with MOI 1 at day 10 with neurobasal complete medium.

After cells differentiation, TGS1 and SMN staining were performed. These staining did not show differences in localization and amount of TGS1 and SMN proteins in transduced cells compared to untreated cells (Fig.18).

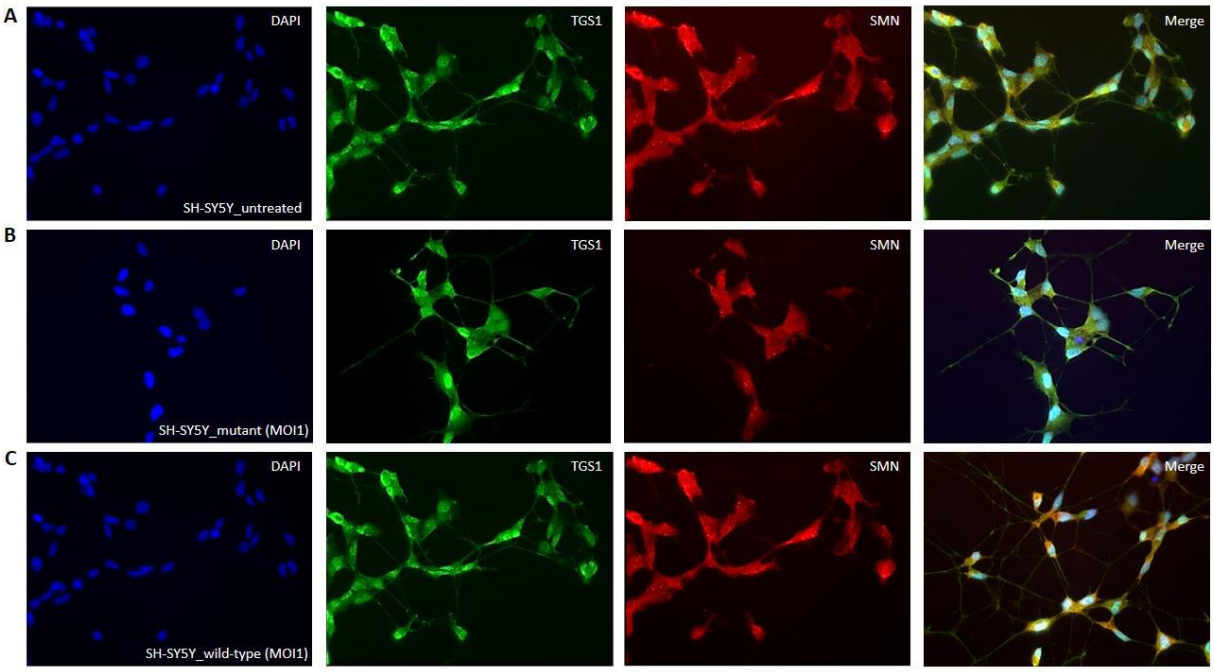
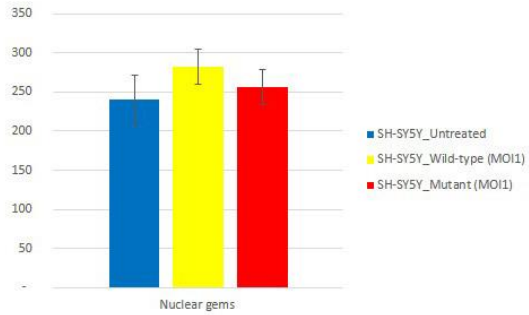


Fig.18: Panel A: Single and merged staining for TGS1 (green), SMN (red) and DAPI (blue) in SH-SY5Y cells not transduced. Panel B: Single and merged staining for TGS1 (green), SMN (red) and DAPI (blue) in SH-SY5Y cells transduced with lentivirus carrying mutant construct with MOI 1. Panel C: Single and merged staining for TGS1 (green), SMN (red) and DAPI (blue) in SH-SY5Y cells transduced with lentivirus carrying wild-type construct with MOI 1. Images took at 20x magnification

After TGS1 and SMN staining, gems count and the count of positive cells were performed for each cell lines in 200 cells, twice. Both these analyses showed no differences in wild-type and mutant SH-SY5Y cells compared to untreated cells (Fig.19).

A

	Number of cells	Nuclear gems
SH-SY5Y_Untreated	200	240
SH-SY5Y_Wild-type (MOI1)	200	282
SH-SY5Y_Mutant (MOI1)	200	256



B

	Number of cells	% Positive cells
SH-SY5Y_Untreated	200	72,5%
SH-SY5Y_Wild-type (MOI1)	200	70%
SH-SY5Y_Mutant (MOI1)	200	72%

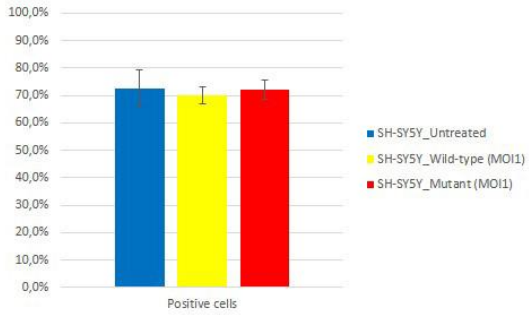


Fig.19: Panel A: Histograms show the gems number in SH-SY5Y_untreated (shown in blue), SH-SY5Y_wild-type (shown in yellow) and SH-SY5Y_mutant (shown in red). Count performed in 200 cells in two different staining. Panel B: Histograms show the percentage of positive cells in SH-SY5Y_untreated (shown in blue), SH-SY5Y_wild-type (shown in yellow) and SH-SY5Y_mutant (shown in red). Count performed in 200 cells in two different staining.

4.3.3.2 Transduction experiments in SH-SY5Y cells with MOI 15

Transduction experiments performed with MOI 15 showed differences in cells transduced with wild-type and mutant lentiviral vectors compared to the untreated cells. Transduction and the selection of the transduced cells were followed by differentiation. After 10 days in complete neurobasal medium, alterations in neurite outgrowth were observed. SH-SY5Y cells transduced with lentivirus carrying wild-type and mutant vectors showed shorter neurites with a high number of branch points compared to the untreated cells (Fig.20).

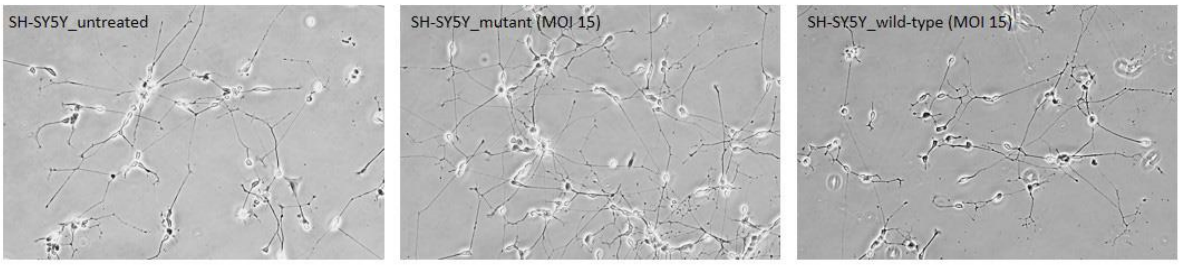


Fig.20: Untreated SH-SY5Y cells and SH-SY5Y transduced with MOI 15 at day 10 with complete neurobasal medium.

These cells were stained with βIII-Tubulin to highlight better the differences in neurite outgrowth. As shown in figure 21, SH-SY5Y transduced with wild-type and mutant vectors shown in the

majority of cells defects in neuron development with short neuritis, characterized by numerous ramification and branch points.

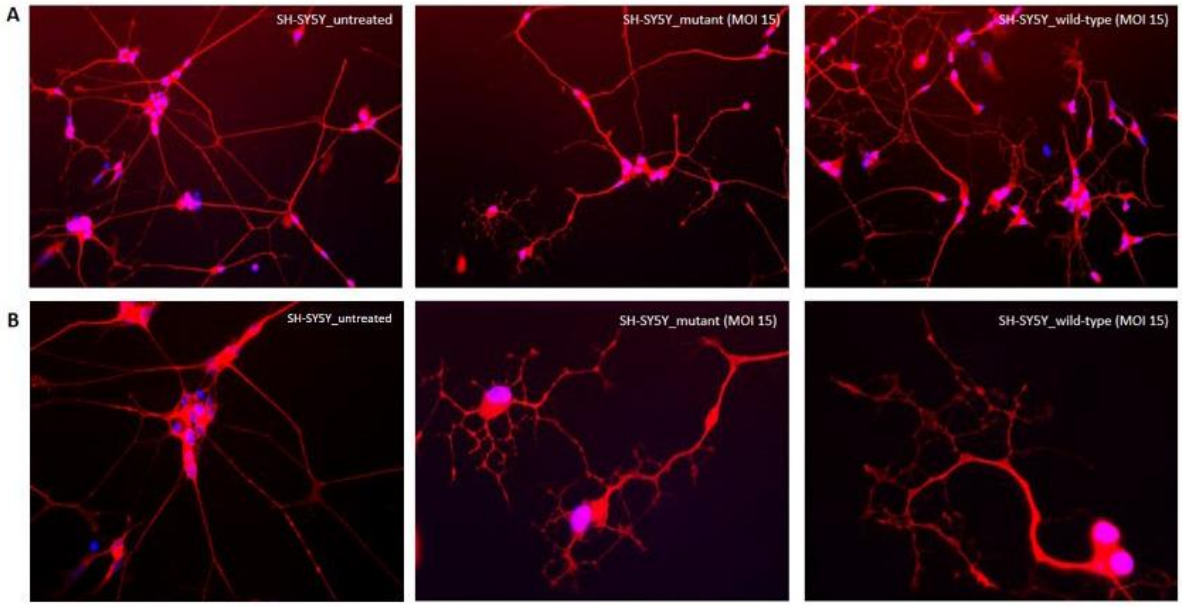
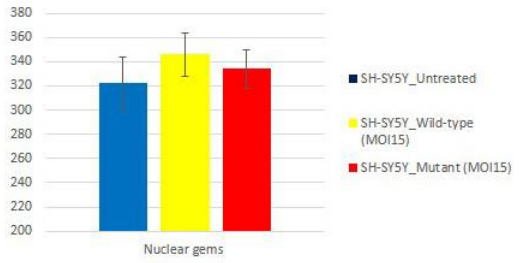


Fig.21: Panel A: βIII-Tubulin staining in SH-SY5Y cells untreated and SH-SY5Y cells transduced with wild-type and mutant constructs. Images were taken at 20X magnification. Panel B: βIII-Tubulin staining in SH-SY5Y cells untreated and SH-SY5Y cells transduced with wild-type and mutant constructs. Images were taken at 40X magnification.

After SMN staining, gems count and the number of positive cells were measured for each cell lines in 200 cells, twice. Both analyses showed no differences in wild-type and mutant SH-SY5Y cells compared to untreated cells (Fig.22).

A

	Number of cells	Nuclear gems
SH-SY5Y_Untreated	200	322
SH-SY5Y_Wild-type (MOI15)	200	346
SH-SY5Y_Mutant (MOI15)	200	334



B

	Number of cells	% Positive cells
SH-SY5Y_Untreated	200	85,00%
SH-SY5Y_Wild-type (MOI15)	200	88%
SH-SY5Y_Mutant (MOI15)	200	92%

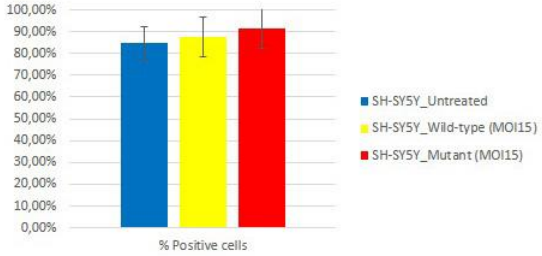


Fig.22: Panel A: Histograms show the gems number in SH-SY5Y_untreated (shown in blue), SH-SY5Y_wild-type (shown in yellow) and SH-SY5Y_mutant (shown in red). Count performed in 200 cells in two different staining. Panel B: Histograms show the percentage of positive cells in SH-SY5Y_untreated (shown in blue), SH-SY5Y_wild-type (shown in yellow) and SH-SY5Y_mutant (shown in red). Count performed in 200 cells in two different staining.

4.3.3.3 Transduction experiments in SH-SY5Y cells with MOI 70

Transduction experiments performed with MOI 70 showed the same differences observed in cells with MOI 15. SH-SY5Y cells transduced with wild-type and mutant lentiviral vectors presented phenotype features diverge from the untreated phenotype (Fig.23)

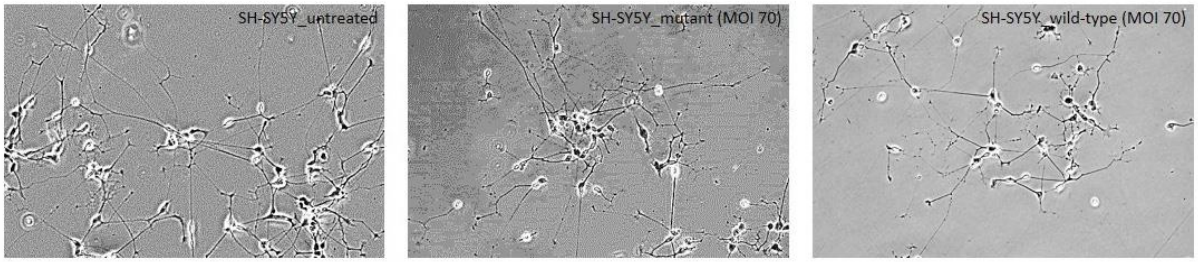


Fig.23: SH-SY5Y cells transduced with MOI 70 at day 10 with complete neurobasal medium.

βIII-Tubulin staining presented better the phenotype features in this case. These cells showed numerous short neurites with numerous branch points (Fig.24).

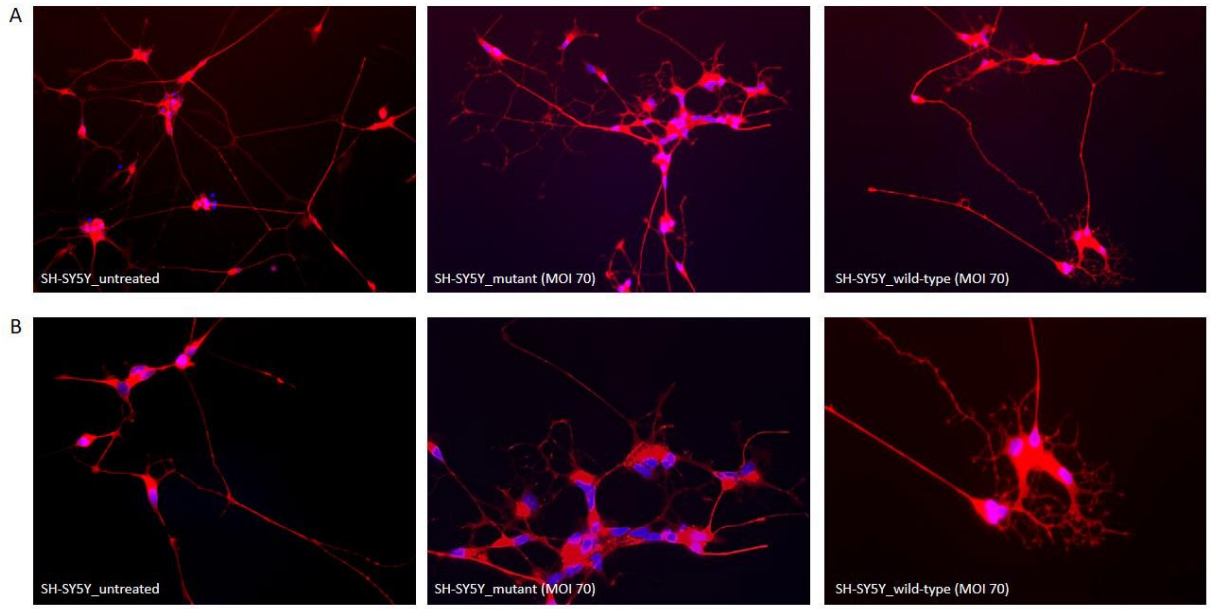


Fig.24: Panel A: βIII-Tubulin staining in SH-SY5Y cells untreated and SH-SY5Y cells transduced with wild-type and mutant constructs. Images were taken at 20X magnification. Panel B: βIII-Tubulin staining in SH-SY5Y cells untreated and SH-SY5Y cells transduced with wild-type and mutant constructs. Images were taken at 40X magnification.

4.3.4 Transduction experiments in fibroblast cell lines

Transduction experiments in fibroblast cell lines were performed with MOI 1 in:

8230: healthy control fibroblasts.

7384: SMA type III patient with 3 copies of *SMN2* gene.

7215: SMA type III patient with 3 copies of *SMN2* gene.

6141: SMA type II patient with 2 copies of *SMN2* gene.

After 24 hours of transduction followed by selection with puromycin for 10 days, 6141 fibroblast cell line died while only a few cells survived in the other two SMA fibroblast cell lines. Moreover, these cells showed reduced cell growth and not spindle shape that is characteristic of fibroblast cell lines. However, this was not observed in the fibroblast cell line of a healthy control, in which growth and cell shape were similar to the untreated cells (Fig.25).

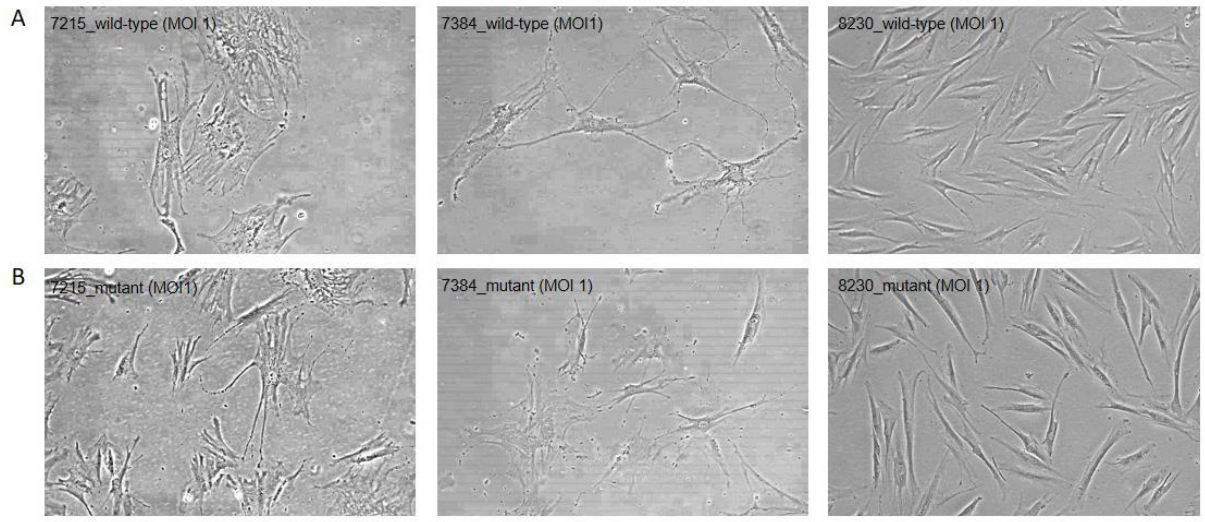


Fig.25: Panel A: Fibroblast cell lines from two SMA patients (7215 and 7384) and fibroblast cell line from a healthy control (8230) transduced with lentivirus carrying wild-type construct after puromycin selection. Panel B: Fibroblast cell lines from two SMA patients (7215 and 7384) and fibroblast cell line from a healthy control (8230) transduced with lentivirus carrying mutant construct after puromycin selection.

SMN and TGS1 staining allowed to point out differences in SMN expression in transduced fibroblast cell lines. Compared to untreated cells, the amount of SMN protein seems to be high especially within the nucleus of cells transduced with both wild-type and mutant lentivirus. SMA cell lines are characterized by few numbers of gems inside the nucleus but after transduction of both lentiviral vectors led to increasing of SMN protein. However, even though SMN level was increased in 7215 and 7384 cell lines, SMN seems not to be organized in gems structures. Similar character, although less evident than in SMA cell lines, was observed in control cell line (8230),

which showed increased amount of SMN protein inside the nucleus. Even in this cell line, SMN seems not to be organized in gems structure (Fig.26)

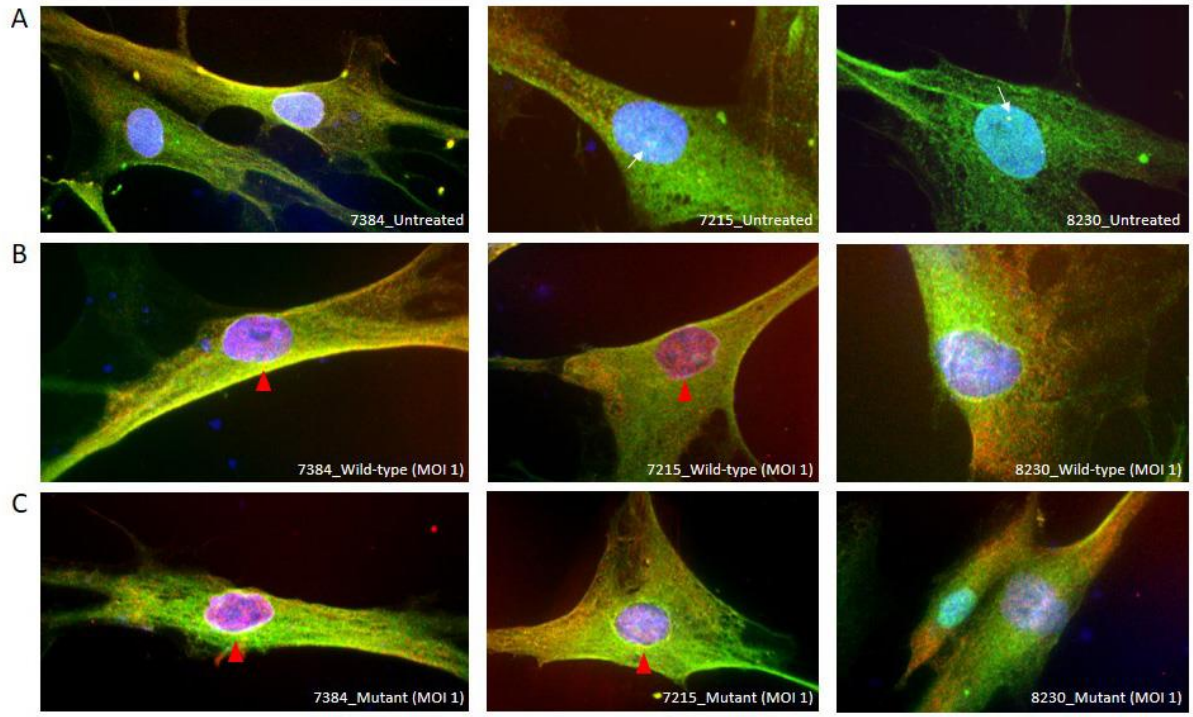


Fig.26: Panel A: TGS1 (green) and SMN (red) staining in untreated fibroblast cell lines of two SMA type III patients (7384, 7215) and healthy control (8230). The white arrow highlights the presence of gems inside the nucleus. Panel B: TGS1 (green) and SMN (red) staining in fibroblast cell lines of two SMA type III patients (7384, 7215) and healthy control (8230) after transduction with wild-type lentiviral vector. The red arrowhead highlights the increased level of SMN staining without the presence of gems. Panel C: TGS1 (green) and SMN (red) staining in fibroblast cell lines of two SMA type III patients (7384, 7215) and healthy control (8230) after transduction with mutant lentiviral vector. The red arrowhead highlights the increased level of SMN staining without the presence of gems.

4.4 LENTIVIRUS TRANSDUCTION WITH GREEN FLUORESCENT PROTEIN

4.4.1 Transduction experiments in SH-SY5Y cells

SH-SY5Y cells were transduced with MOI 10 and cell sorting has been performed after three days from transduction. Cell sorting showed the efficiency of transduction for each lentiviral vectors (Fig.27B). Moreover, during the cell sorting, two populations were sorted. One population with slightly increased fluorescence over the untransfected control and a GPF bright population with high expression of GPF (Fig.27A).

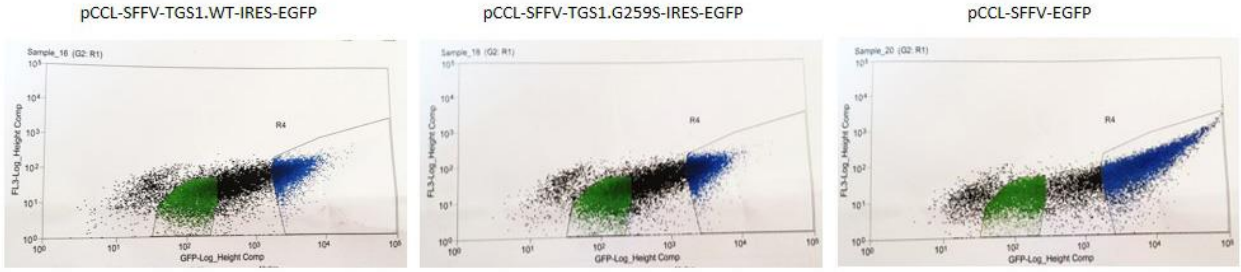
First populations sorted:

- SH-SY5Y_WT_low GFP
- SH-SY5Y_MUT_low GFP
- SH-SY5Y_CT_low GFP

Second populations sorted:

- SH-SY5Y_WT_high GFP
- SH-SY5Y_MUT_ high GFP
- SH-SY5Y_CT_ high GFP

A



B

	SH-SY5Y (MOI 10)		
	Total	Low GFP	High GFP
pCCL-SFFV-TGS1.WT-IRES-EGFP	8.46%	2.62%	0.7%
pCCL-SFFV-TGS1.G259S-IRES-EGFP	6.76%	1.84%	0.56%
pCCL-SFFV-EGFP	6.27%	0.43%	3.41%

Fig.27: Panel A: Sorting of two populations in SH-SY5Y cells. In green is shown the cell population with low expression of GFP while in blue the cell population with high expression of GFP. Panel B: The first column of the table represents the percentage of SH-SY5Y cells transduced, the second column is the percentage of cells with low GFP expression while the third column is the percentage of cells with high GFP expression.

4.4.1.1 SH-SY5Y cells differentiation

After cell sorting, each population of SH-SY5Y cells have been differentiated together with untreated cells. At day 5, untreated cells (Fig.28A) and SH-SY5Y_ CT_low GFP (Fig.28F) showed the same neurite outgrowth. While in SH-SY5Y_CT_high GFP (Fig.28G) and SH-SY5Y cells transduced with WT and MUT vectors in both populations with low and high GPF expression (Fig.28B-C-D-E), shorter neurite outgrowth was observed compared the untreated or SH-SY5Y_CT_low GFP.

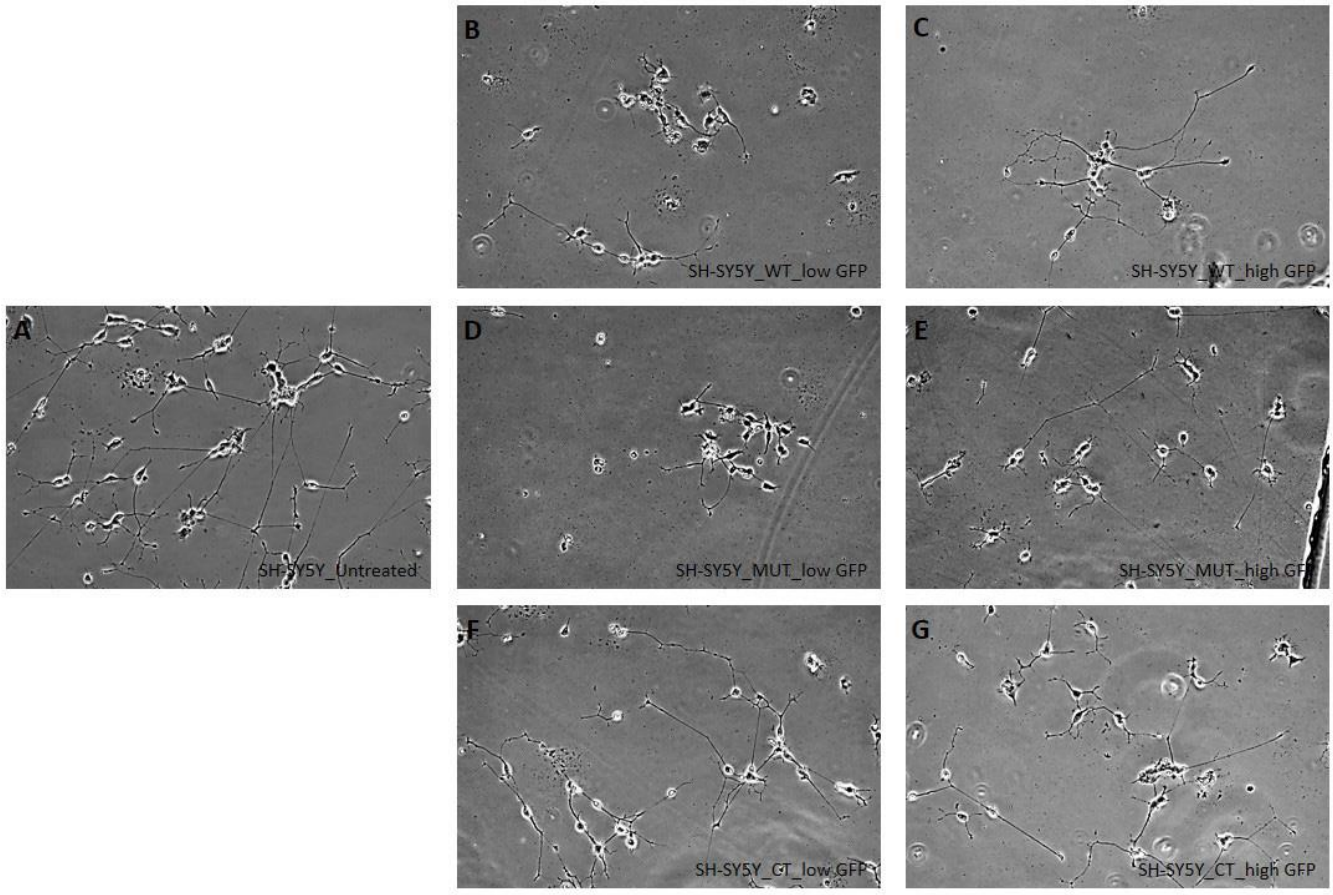


Fig.28: Images took at day 5 of differentiation in untreated SH-SY5Y cells (A), SH-SY5Y cells transduced with wild-type vector with low GFP expression (B) and high GFP expression (C), SH-SY5Y cells transduced with mutant vector with low GFP expression (D) and high GFP expression (E) and SH-SY5Y cells transduced with CT vector with low GFP expression (F) and high GFP expression (G)

4.4 3D MODELING

It has been possible only to analyze p.Asn704Ser variation within 3D structure since the N-terminal sequence of the protein where p.Gly259Ser is located was not available in PDB database. C-terminal region of TGS1 is characterized by the presence of AdoMet-binding site composed of ^{694}VVDAFCGVGGN^{704} and ^{717}IAIDI^{721} motifs and m^{7}G binding site composed of ^{763}SPPWGG^{768141} motif. p.Asn704Ser is located in one α-helix within the AdoMet-binding site and it is one of the residues directly involved with the linkage between the protein and AdoHcy. Therefore, p.N704 residue results essential for a correct enzymatic activity of the protein. For the difference in size of Asparagine and Serine, the amino acid substitution p.Asn704Ser could modify the binding between AdoHcy and TGS1 interfering with the enzymatic activity of the protein (Fig.29).

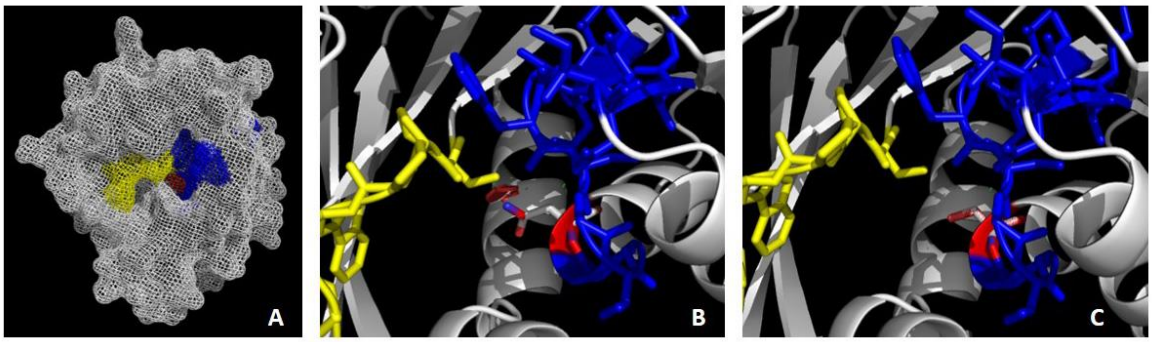


Fig.29: Panel A: Representation of TGS1 3D structure. In yellow is shown the m7G binding site, in blue the AdoMet-binding site and in red p.N704 residue. Panel B: In red is shown p.N704 residue within of TGS1 3D structure. Panel C: representation of p.S704 using PyMol mutagenesis tool.

5. DISCUSSION AND CONCLUSION

Spinal muscular atrophy is due to the loss of survival motor neuron (SMN) protein, caused, in 95% of patients, by homozygous deletions in the *survival motor neuron 1* (*SMN1*) gene. SMN protein is crucial for motor neuron survival since lack of this protein leads to the spinal motor neuron death followed by skeletal muscle atrophy and paralysis. However, the molecular pathway underlying SMA is still not fully understood. It is not clear why the loss of snRNP biogenesis in all cells leads to a selective damage in neuromuscular system, only some suggestions have been proposed. Experiments in model organisms highlighted that similar phenotype observed with SMN loss occurs even with knockout or overexpression of other factors involved in snRNP biogenesis. This was the starting point of the project. The discovery, in two unrelated patients with SMA type IV-like clinical features, of two different mutations in one gene involved in snRNP biogenesis pointed out a possible involvement of this gene in the pathogenesis of SMA disease. Two missense mutations were found in *trimethylguanosine synthase 1* (*TGS1*) gene: p.Gly259Ser (c.496G>A) and p.Asn704Ser (c.2111A>G). *TGS1* gene encodes for TGS1 enzyme that is responsible for hypermethylation of snRNPs and snoRNPs m^7G cap structure. The hypermethylation event is essential for the complete maturation of both snRNPs and snoRNPs.

Thus, the aim of this project was to perform functional studies for both mutations. We carried out the experiments with fibroblast cell line of patient carrying p.Gly259Ser mutation (L1266) and lymphoblastoid cell line of patient carrying p.Asn704Ser mutation (343). First, we evaluated TGS1 and SMN expression in whole cell lysate. Western blot in patients showed two bands of TGS1 protein, but no difference of TGS1 expression was observed in both patients (Fig.10). While SMN expression was evaluated only in patient carrying p.Gly259Ser mutation (L1266) and no differences were observed between L1266 and three control cell lines. The further experiments have been carried on just for L1266 since for lymphoblastoid cell line of patient 343 it was not possible to perform TGS1 or SMN immunofluorescence analysis. TGS1 staining in L1266 sample showed that TGS1 was expressed in both nucleus and cytoplasm and no difference was observed between L1266 and the three control cell lines (Fig.12). Immunofluorescence experiment was performed in L1266 patient, three healthy controls and one SMA type III patient (7384). Interestingly, 7384 sample showed a decreased expression of the short isoform of TGS1, when compared to L1266 and the other three

controls. Thus, the study was extended to other three SMA patients and one of these showed the same reduction of TGS1 short isoform (Fig.12). The decrease of TGS1 protein expression in fibroblast cell lines from SMA patients was further confirmed by western blotting analysis in whole cell lysates (Fig.10 and 11).

After the expression of SMN protein was evaluated. Western blot of SMN from cell lysates showed no reduction in L1266 patient cell line (Fig.9). However, SMN staining pointed out some differences compared to controls (Fig.13). L1266 showed 40% of increased numbers of gems within the nucleus than the controls. Together with the gem counts, was performed the count of positive cells (cells with at least one gem inside the nucleus) to highlight if the higher number of gems was due to a higher number of positive cells or an increased gems number inside each nucleus. This study allowed to clarify that L1266 sample is characterized by a higher number of gems inside each cell (Fig.14). The increased amount of gems number could be explained with a gain of function effect of p.Gly259Ser mutation. TGS1 is involved in the methylation steps for the conversion of the 7-monomethylguanosine (m^7G) caps to a 2,2,7-trimethyguanosine ($m^{2,2,7}G$ or TMG) cap structure. TMG cap structure represents together with Sm core a bipartite nuclear localisation signal required to promote the nuclear uptake of snRNPs. Then, an increased activity of TGS1 may lead to increased snRNPs importation followed by higher gems formation. Moreover, the functional effect of p.Gly259Ser mutation was studied even in neuronal cell lines. Two lentiviral vectors were generated: one carrying wild-type full-length *TGS1* and the other carrying p.Gly259Ser mutation. After the generation of two lentiviruses, transduction experiments in SH-SY5Y cells were carried out in order to highlight possible alteration in neurite outgrowth. Three different multiplicity of infection were used: MOI 1, MOI 15 and MOI 70. These cell lines were then differentiated into neurons and the efficiency of differentiation was verified using the antibody against the neuron-specific class III beta-tubulin. In cell lines derived from transduction with MOI 1, no significant differences were observed compared to untreated cells. They did not show modifications in neurite outgrowth or in the localization of TGS1 and SMN protein (Fig.17 and 18). Even the gems count did not show any difference (Fig.19). While in cell lines transduced with MOI 15 and MOI 70 it has been possible to point out some specific phenotypic features (Fig.21 and 24). βIII-tubulin staining in differentiated cells showed the presence of defects in neuron development. Interestingly, the same development defects have been detected in SH-SY5Y cells transduced with wild-type or mutant lentiviruses. Cells transduced with wild-type and mutant lentivirus showed in the majority of cells short neurites with numerous branch points and tips compared to the untreated control characterized by long neurites. These morphologic features are observed even in SMA

disease[71, 76,142]. Since this particular phenotype has been observed in both cell lines with an overexpression of wild-type or mutant TGS1, it suggests that the neurons degeneration is due to an increased activity of TGS1 protein. Therefore, even transduction experiments in SH-SY5Y cells allowed to hypothesize a gain of function effect of p.Gly259Ser mutation. As it has been hypothesized in the previous experiments.

Furthermore, transduction experiments were performed even in fibroblast cell lines of one healthy control, two SMA type III patients and one SMA type II patient. The multiplicity of infection used for fibroblast was MOI 1 and it turned out to be quite toxic in particular for SMA fibroblast cell lines. SMA Type II fibroblast cell line died after transduction while in SMA type III patients few cells alive showed reduced cell growth and not spindle shape, characteristics of fibroblast cell. However, this toxicity was not observed in cell line of the healthy control (Fig.25). In these transduced cell lines, SMN and TGS1 staining were performed. Interestingly, even in this case, the phenotypic features observed have been the same in cell lines transduced with wild-type or mutant vectors. SMA fibroblast cell lines showed an increased level of SMN protein inside the nucleus. The same happens in the healthy control even if less evident than in SMA cell lines. It remains to investigate if the increased amount of SMN protein is organized in gems structure. However, even this result could be explained with a gain of function effect of p.Gly259Ser mutation since fibroblast cell lines transduced with wild-type and mutant vectors showed the same phenotype with an increase in SMN protein within the nucleus.

All these experiments lead to hypothesize a gain of function effect of p.Gly259Ser mutation with a possible involvement in spinal muscular disease. To confirm this hypothesis other three lentiviruses were generated. One carrying wild-type *TGS1*, one carrying p.Gly259Ser mutation and one without *TGS1* insertion but only with GFP tag (CT vector). The lentiviral vector without TGS1 insertion was generated to verify if the phenotype observed in the previous experiments was caused to the increased activity of TGS1 or to the toxicity of the viruses. Transduction of SH-SY5Y cells with these three lentiviral vectors was followed by cell sorting. The sorting allowed to obtain two different populations: one population with low GFP expression and one population with high GFP expression. These cells were differentiated and at day 5 some differences were observed (Fig.28). Cells transduced with the CT vector were used as controls for evaluating the toxicity of the virus. It was then possible to establish that high GFP expression was toxic to the cells since SH-SY5Y_CT_high GFP showed an alteration in neurite outgrowth. While the population with low GFP expression (SH-SY5Y_CT_low GFP) showed a similar phenotype of the untreated cells

highlighting the non-toxicity of the virus. Thus, the observation of neurite outgrowth was performed only in the populations with low GFP expression. In SH-SY5Y_WT_low GFP and SH-SY5Y_MUT_low GFP, as in the experiments performed with the first two lentiviruses, an alteration in neurite outgrowth was noticed. Even in this case, SH-SY5Y cells transduced with wild-type and mutant vectors showed the same phenotype with short neurites.

In conclusion, our results in this study indicate 1) the gain-of-function dominant negative effect of p.Gly259Ser mutation in TGS1 gene on SMN biogenesis and cell survival; 2) this mutation acts negatively on cell survival at the presence of SMN deficiency; 3) p.Gly259Ser mutation inhibits neurite outgrowth in neuronal cells *in vitro* and may indicates its pathogenic effect.

This study has strengthened our starting hypothesis by highlighting a possible involvement of *TGS1* gene in the spinal muscular atrophy-like clinical features. Defects in TGS1 activity could interfere with SMA pathogenesis since the function of this enzyme is closely interrelated with SMN function. Then, it is likely that alteration in TGS1 function could lead a similar phenotype shown by SMA patients. So far, our study was focused on p.Gly259Ser mutation, which seems a pathogenic mutation with a gain of function effect. It is possible that an increased function of TGS1 protein causes an increase importation inside the nucleus of SMN protein in fibroblast cell lines. Moreover, the increased activity of TGS1 seems to induce degeneration of neurons as occurs in spinal muscular atrophy disease. However, further analysis is required to complete the study for this mutation and to investigate the functional effect of the other missense mutation.

6. REFERENCES

1. Banker BQ, Byers RK. Infantile muscular atrophy. *arch neurol* **5,** 140–164 (1961).
2. Pearn, J. Incidence, prevalence, and gene frequency studies of chronic childhood spinal muscular atrophy. *Eur. J. Hum. Genet.* **15(6),** 409–413 (1978).
3. Wilson RB. & Ogino S. Carrier frequency of spinal muscular atrophy. *Lancet* **372,** 1542 (2008).
4. Roberts DF, Chavez J, Court SDM. The Genetic Component in Child Mortality. *Arch. Dis. Child.* **45(239),** 33–38 (1970).
5. Werdnig G. Two early infantile hereditary cases of progressive muscular atrophy simulating dystrophy, but on a neural basis. 1891. *arch neurol* **25(3),** 276–8 (1971).
6. Hoffmann J. Ueber chronische spinale Muskelatrophie im Kindesalter, auf familiarer Basis. *Dtsch. Z. Nervenheilkd.* **3,** 427–470 (1893).
7. Victor Dubowitz. Infantile muscular atrophy a prospective study with particular reference to a slowly progressive variety. *Brain* **87,** 707–718 (1964).
8. Munsat TL, Woods R, Fowler WPC. Neurogenic muscular atrophy of infancy with prolonged survival. The variable course of Werdnig-Hoffmann Disease. *Brain* **92(1),** 9–24 (1969).
9. Lefebvre S, *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80(1),** 155–165 (1995).
10. Mahadevan MS, *et al*. I. SMA genes: deleted and duplicated. *Nat. Genet.* **9(2),** 112–113 (1995).
11. Lorson CL, *et al*. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci.* **96(11),** 6307–6311 (1999).
12. Monani UR, *et al*. A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum Mol Genet.* **8(7),** 1177–1183 (1999).
13. Feldkötter M, *et al*. Quantitative Analyses of SMN1 and SMN2 Based on Real-Time LightCycler PCR: Fast and Highly Reliable Carrier Testing and Prediction of Severity of Spinal Muscular Atrophy. *Am. J. Hum. Genet.* **70(2),** 358–368 (2002).
14. Taylor JE, *et al.* Correlation of SMNt and SMNc gene copy number with age of onset and survival in spinal muscular atrophy. *Eur. J. Hum. Genet.* **6(5),** 467–474 (1998).
15. Munsat TL, Davies KE. International SMA consortium meeting. *Neuromuscul Disord* **2(5-6),** 423–428 (1992).
16. Munsat TL. Workshop report. International SMA collaboration. *Neuromuscul Disord* **1(2),** 81 (1991).
17. Dubowitz V. Very severe spinal muscular atrophy (SMA type 0): an expanding clinical phenotype. *Eur J Paediatr Neurol.* **3(2),** 49–51 (1999).
18. Zerres K, *et al*. A collaborative study on the natural history of childhood and juvenile onset proximal spinal muscular atrophy (type II and III SMA): 569 patients. *J Neurol Sci.* **146(1),**

67–72 (1997).

19. Moulard B, *et al*. Association between centromeric deletions of the SMN gene and sporadic adult-onset lower motor neuron disease. *Ann Neurol.* **43(5),** 640–644 (1998).

20. Sardone V, *et al*. Antisense Oligonucleotide-Based Therapy for Neuromuscular Disease. *Molecules* **22(4),** doi: 10.3390/molecules22040563 (2017).

21. Duque S. *et al.* Intravenous Administration of Self-complementary AAV9 Enables Transgene Delivery to Adult Motor Neurons. *Mol. Ther.* **17(7),** 1187–1196 (2009).

22. Valori CF, *et al*. Systemic delivery of scAAV9 expressing SMN prolongs survival in a model of spinal muscular atrophy. *Sci Transl Med.* **2(35),** 35–42 (2010).

23. Dominguez E, *et al.* Intravenous scAAV9 delivery of a codon-optimized SMN1 sequence rescues SMA mice. *Hum. Mol. Genet.* **20(4),** 681–693 (2011).

24. Mitrpant C, *et al.* Improved Antisense Oligonucleotide Design to Suppress Aberrant SMN2 Gene Transcript Processing : Towards a Treatment for Spinal Muscular Atrophy. *PLoS One* **8(4),** e62114 (2013).

25. Hua Y, *et al.* Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature* **478(7367),** 123–126 (2011).

26. Zhou H, *et al*. A Novel Morpholino Oligomer Targeting ISS-N1 Improves Rescue of Severe Spinal Muscular Atrophy Transgenic Mice. *Hum. Gene Ther.* **24(3),** 331–342 (2013).

27. Scoto M, *et al*. Therapeutic approaches for spinal muscular atrophy (SMA). *Gene Ther.* **24(9),** 514-519 (2017).

28. Avila AM, *et al.* Trichostatin A increases SMN expression and survival in a mouse model of spinal muscular atrophy. *J. Clin. Invest.* **117(3),** 659–671 (2007).

29. Mohseni J, Zabidi-Hussin ZAMH, Sasongko TH. Histone deacetylase inhibitors as potential treatment for spinal muscular atrophy. *Genet. Mol. Biol.* **36(3),** 299–307 (2013).

30. Palacino J, *et al*. SMN2 splice modulators enhance U1–pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.* **11(7),** 511–517 (2015).

31. Corti S, *et al.* Embryonic stem cell-derived neural stem cells improve spinal muscular atrophy phenotype in mice. *Brain* **133(2),** 165–181 (2010).

32. Corti S. *et al.* Genetic Correction of Human Induced Pluripotent Stem Cells from Patients with Spinal Muscular Atrophy. *Sci Transl Med.* **4(165),** 165ra162 (2012).

33. FDA approves first drug for spinal muscular atrophy. (2016). Available at: https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm534611.htm.

34. SPINRAZA® (Nusinersen) Approved in the European Union as First Treatment for Spinal Muscular Atrophy. Available at: http://media.biogen.com/press-release/investor-relations/spinraza-nusinersen-approved-european-union-first-treatment-spinal-.

35. U.S. FDA Approves Biogen's SPINRAZA™ (nusinersen), The First Treatment for Spinal Muscular Atrophy. Available at: http://media.biogen.com/press-release/neurodegenerative-diseases/us-fda-approves-biogens-spinraza-nusinersen-first-treatment.

36. Evers MM, *et al*. Antisense oligonucleotides in therapy for neurodegenerative disorders. *Adv.*

Drug Deliv. Rev. **87,** 90–103 (2015).

37. Talbot K & Tizzano EF. The clinical landscape for SMA in a new therapeutic era. *Gene Ther.* **24(9),** 529-533 (2017).

38. Farooq FT, Holcik M, MacKenzie A. Spinal Muscular Atrophy: Classification, Diagnosis, Background, Molecular Mechanism and Development of Therapeutics. *Neurodegenerative Diseases* 561–579 (2013).

39. Cho S & Dreyfuss G. A degron created by SMN2 exon 7 skipping is a principal contributor to spinal muscular atrophy severity. *Genes Dev.* **24(5),** 438–442 (2010).

40. Selenko P, *et al*. SMN tudor domain structure and its interaction with the Sm proteins. *Nat Struct Biol.* **8(1),** 27–31 (2001).

41. Seng CO, *et al*. The SMN structure reveals its crucial role in snRNP assembly. *Hum. Mol. Genet.* **24(8),** 2138–2146 (2015).

42. Singh RN., *et al.* Diverse role of survival motor neuron protein. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1860(3),** 299–315 (2017).

43. Zhang H, *et al*. QNQKE Targeting Motif for the SMN-Gemin Multiprotein Complex in Neurons. *J. Neurosci. Res.* **85(12),** 2657–2667 (2007).

44. Coovert DD, *et al.* The survival motor neuron protein in spinal muscular atrophy. *Hum Mol Genet.* **6(8),** 1205–1214 (1997).

45. Liu Q & Dreyfuss G. A novel nuclear structure containing the survival of motor neurons protein. *EMBO J.* **15(14),** 3555–3565 (1996).

46. Brasch K & Ochs RL. Nuclear bodies (NBs): A newly 'rediscovered' organelle. *Exp Cell Res.* **202(2),** 211–223 (1992).

47. Gall JG, *et al*. Is the sphere organelle/coiled body a universal nuclear component? *Dev Genet.* **16(1),** 25–35 (1995).

48. Paushkin S, *et al*. The SMN complex, an assemblyosome of ribonucleoproteins. *Curr Opin Cell Biol.* **14(3),** 305–312 (2002).

49. Hebert MD, *et al*. Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein. *Genes Dev.* **15(20),** 2720–2729 (2001).

50. Hearst SM, *et al.* Cajal-body formation correlates with differential coilin phosphorylation in primary and transformed cell lines. *J Cell Sci.* **122(11),** 1872–1881 (2009).

51. Ramon y Cajal, S. R. Un sencillo metodo de coloracion seletiva del reticulo protoplasmico y sus efectos en los diversos organos nerviosos de vertebrados y invertebrados. *Trab. Lab. Invest. Biol.* **2,** 129–221 (1903).

52. Mao YS, Zhang B, Spector DL. Biogenesis and function of nuclear bodies. *Trends Genet.* **27(8),** 295–306 (2011).

53. Sleeman JE & Trinkle-Mulcahy L. Nuclear bodies: New insights into assembly/dynamics and disease relevance. *Curr Opin Cell Biol.* **28,** 76–83 (2014).

54. Machyna M, Heyn P, Neugebauer KM. Cajal bodies: where form meets function. *Wiley Interdiscip Rev RNA.* **4(1),** 17–34 (2013).

55. Spector DL & Lamond AI. Nuclear Speckles. *Cold Spring Harb Perspect Biol.* **3(2),** doi:10.1101/cshperspect.a000646 (2011).

56. Wang Q, *et al.* Cajal bodies are linked to genome conformation. *Nat. Commun.* **7,** 10966 (2016).

57. Gubitz AK, Feng W, Dreyfuss G. The SMN complex. *Exp. Cell Res.* **296(1),** 51–56 (2004).

58. Li, W. How do SMA-linked mutations of SMN1 lead to structural / functional deficiency of the SMA protein ? *PLos One* **12(6),** e0178519 (2017).

59. Pellizzoni L, *et al*. The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Curr Biol.* **11(14),** 1079–1088 (2001).

60. Mourelatos Z, *et al.* miRNPs: A novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16(6),** 720–728 (2002).

61. Bachand F, *et al*. The Product of the Survival of Motor Neuron (SMN) Gene is a Human Telomerase-associated Protein. *Mol Biol Cell.* **13(9),** 3192–3202 (2002).

62. Rossoll W & Bassell GJ. Spinal Muscular Atrophy and a Model for Survival of Motor Neuron Protein Function in Axonal Ribonucleoprotein Complexes. *Results Probl Cell Differ* **48,** 289–326 (2009).

63. Fallini C, Bassell GJ, Rossoll W. Spinal muscular atrophy: The role of SMN in axonal mRNA regulation. *Brain Res.* **1462,** 81–92 (2012).

64. Fallini C, *et al*. Deficiency of the Survival of Motor Neuron Protein Impairs mRNA Localization and Local Translation in the Growth Cone of Motor Neurons. *J. Neurosci.* **36(13),** 3811–3820 (2016).

65. Fallini C, *et al.* The Survival of Motor Neuron (SMN) Protein Interacts with the mRNA-Binding Protein HuD and Regulates Localization of Poly(A) mRNA in Primary Motor Neuron Axons. *J Neurol Sci.* **31(10),** 3914–3925 (2011).

66. Hubers L, *et al.* HuD interacts with survival motor neuron protein and can rescue spinal muscular atrophy-like neuronal defects. *Hum Mol Genet.* **20(3),** 553–579 (2011).

67. Dombert B, *et al.* Localization of Smn and hnRNP R in Axon Terminals of Embryonic and Postnatal Mouse Motoneurons. *PLoS One* **9(10),** e110846. (2014).

68. Béchade C, *et al*. Subcellular distribution of survival motor neuron (SMN) protein: Possible involvement in nucleocytoplasmic and dendritic transport. *Eur J Neurosci.* **11(1),** 293–304 (1999).

69. Fallini C, Bassell GJ & Rossoll W. High-efficiency transfection of cultured primary motor neurons to study protein localization , trafficking , and function. *Mol Neurodegener.* **5,** 17 (2010).

70. Zhang H, *et al.* Multiprotein Complexes of the Survival of Motor Neuron Protein SMN with Gemins Traffic to Neuronal Processes and Growth Cones of Motor Neurons. *J Neurosci* **26(33),** 8622–8632 (2006).

71. Rossoll W, *et al.* Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of β-actin mRNA in growth cones of motoneurons. *J. Cell Biol.* **163(4),** 801–812 (2003).

72. Fallini C, *et al*. Dynamics of survival of motor neuron (SMN) protein interaction with the mRNA-binding protein IMP1 facilitates its trafficking into motor neuron axons. *Dev Neurobiol.* **74(3),** 319–332 (2014).

73. Chia JX, Efimova N, Svitkina TM. Neurite outgrowth is driven by actin polymerization even in the presence of actin polymerization inhibitors. **27(23),** 3695–3704 (2016).

74. Mourelatos Z, *et al*. SMN interacts with a novel family of hnRNP and spliceosomal proteins. *EMBO J.* **20(19),** 5443–5452 (2001).

75. Glinka M, *et al*. The heterogeneous nuclear ribonucleoprotein-R is necessary for axonal β-actin mRNA translocation in spinal motor neurons. *Hum Mol Genet.* **19(10),** 1951–1966 (2010).

76. Mcwhorter ML, *et al*. Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. *J Cell Biol.* **162(5),** 919–931 (2003).

77. Ross AF, *et al*. Characterization of a beta-actin mRNA zipcode-binding protein. *Mol Cell Biol.* **17(4),** 2158–2165 (1997).

78. Akten B, *et al.* Interaction of survival of motor neuron (SMN) and HuD proteins with mRNA cpg15 rescues motor neuron axonal deficits. *PNAS* **108(25),** 10337–10342 (2011).

79. Tiruchinapalli DM, Ehlers MD, Keene JD. Activity-dependent expression of RNA binding protein HuD and its association with mRNAs in neurons. *RNA Biol.* **5(3),** 157–168 (2008).

80. Fallini C, *et al*. The Survival of Motor Neuron (SMN) Protein Interacts with the mRNA-Binding Protein HuD and Regulates Localization of Poly(A) mRNA in Primary Motor Neuron Axons. *J Neurosci* **31(10),** 3914–3925 (2011).

81. He Q, Dent EW & Meiri KF. Modulation of Actin Filament Behavior by GAP-43 (Neuromodulin) Is Dependent on the Phosphorylation Status of Serine 41, the Protein Kinase C Site. *J. Neurosci.* **17(10),** 3515–3524 (1997).

82. Denny JB. Molecular mechanisms, biological actions, and neuropharmacology of the growth-associated protein GAP-43. *Curr Neuropharmacol.* **4(4),** 293–304 (2006).

83. Patel AA & Steitz JA. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol.* **4(12),** 960–970 (2003).

84. Lührmann R, Kastner B & Bach M. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim Biophys Acta.* **1087(3),** 265–292 (1990).

85. Grimm C, *et al.* Structural Basis of Assembly Chaperone-Mediated snRNP Formation. *Mol. Cell* **49(4),** 692–703 (2013).

86. Meister G, *et al.* Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr. Biol.* **11(24),** 1990–1994 (2001).

87. Burghes AH & Beattie CE. Spinal Muscular Atrophy: Why do low levels of SMN make motor neurons sick? *Nat Rev Neurosci* **10(8),** 597–609 (2009).

88. Ogawa C, *et al.* Gemin2 plays an important role in stabilizing the survival of motor neuron complex. *J Biol Chem.* **282(15),** 11122–11134 (2007).

89. Cauchi RJ. SMN and Gemins: 'We are family' ... or are we?: Insights into the partnership

between Gemins and the spinal muscular atrophy disease protein SMN. *Bioessays* **32(12),** 1077–1089 (2010).

90. Pellizzoni L. Chaperoning ribonucleoprotein biogenesis in health and disease. *EMBO Rep.* **8(4),** 340-345 (2007).

91. Fischer U, Liu Q, Dreyfuss G. The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell* **90(6),** 1023–1029 (1997).

92. Will CL, Luhrmann R. Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol.* **13(3),** 290–301 (2001).

93. Mouaikel J, *et al*. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol. Cell* **9(4),** 891–901 (2002).

94. Plessel G, Fischer U, Lührmann R. m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. *Mol Cell Biol.* **14(6),** 4160–4172 (1994).

95. Mouaikel J, *et al.* Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. *EMBO Rep.* **4(6),** 616–622 (2003).

96. Borg RM, *et al*. Disruption of snRNP biogenesis factors Tgs1 and pICln induces phenotypes that mirror aspects of SMN-Gemins complex perturbation in Drosophila, providing new insights into spinal muscular atrophy. *Neurobiol Dis.* **94,** 245–258 (2016).

97. Li DK, *et al*. SMN control of RNP assembly: from post-transcriptional gene regulation to motor neuron disease. *Semin Cell Dev Biol.* **32,** 22–29 (2014).

98. Massenet S, *et al*. The SMN Complex Is Associated with snRNPs throughout Their Cytoplasmic Assembly Pathway. *Mol Cell Biol.* **22(18),** 6533–6541 (2002).

99. Mahmoudi S. *et al.* Wrap53 Is Essential for Cajal Body Formation and for Targeting the Survival of Motor Neuron Complex To Cajal Bodies. *PLoS Biol.* **8(11),** e1000521. (2010).

100. Fischer U, Englbrecht C, Chari A. Biogenesis of spliceosomal small nuclear ribonucleoproteins. *Wiley Interdiscip Rev RNA* **2(5),** 718–731 (2011).

101. Lamond AI & Spector DL. Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol.* **4(8),** 605–612 (2003).

102. Pellizzoni L, *et al*. A Novel Function for SMN, the Spinal Muscular Atrophy Disease Gene Product, in Pre-mRNA Splicing. *Cell* **95(5),** 615–624 (1998).

103. Shpargel KB & Matera AG. Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. *Proc Natl Acad Sci.* **102(48),** 17372–17377 (2005).

104. Narayanan U, *et al*. Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol Cell* **16(2),** 223–234 (2004).

105. Lemm I, *et al*. Ongoing U snRNP Biogenesis Is Required for the Integrity of Cajal Bodies. *Mol Biol Cell* **17(7),** 3221–3231 (2006).

106. Watkins NJ & Bohnsack MT. The box C/D and H/ACA snoRNPs: Key players in the

modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA* **3(3),** 397–414 (2012).

107. Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol.* **8(3),** 209–220 (2007).

108. Reichow SL, *et al*. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res.* **35(5),** 1452–1464 (2007).

109. Verheggen C & Bertrand E. CRM1 plays a nuclear role in transporting snoRNPs to nucleoli in higher eukaryotes. *Nucleus* **3(2),** 132–137 (2012).

110. Hirose T, Shu MD, Steitz JA. Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol Cell* **12(1),** 113–123 (2003).

111. Jones KW, *et al.* Direct Interaction of the Spinal Muscular Atrophy Disease Protein SMN with the Small Nucleolar RNA-associated Protein Fibrillarin. *J Biol Chem.* **276(42),** 38645–38651 (2001).

112. Boulon S, *et al.* PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Mol Cell* **16(5),** 777–787 (2004).

113. Pradet-Balade B, *et al.* CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *EMBO J.* **30(11),** 2205–2218 (2011).

114. Yang Y, *et al*. Conserved Composition of Mammalian Box H / ACA and Box C / D Small Nucleolar Ribonucleoprotein Particles and Their Interaction with the Common Factor Nopp140. *Mol Biol Cell* **11(2),** 567–577 (2000).

115. Renvoisé B, *et al.* The loss of the snoRNP chaperone Nopp140 from Cajal bodies of patient fibroblasts correlates with the severity of spinal muscular atrophy. *Hum Mol Genet.* **18(7),** 1181–1189 (2009).

116. Franke J, Gehlen J, Ehrenhofer-Murray AE. Hypermethylation of yeast telomerase RNA by the snRNA and snoRNA methyltransferase Tgs1. *J Cell Sci.* **121(21),** 3553–3560 (2008).

117. Wurth L, *et al.* Hypermethylated-capped selenoprotein mRNAs in mammals. *Nucleic Acids Res.* **42(13),** 8663–8677 (2014).

118. Zhu Y, *et al.* Cloning and characterization of PIMT, a protein with a methyltransferase domain, which interacts with and enhances nuclear receptor coactivator PRIP function. *PNAS* **98(18),** 10380–10385 (2001).

119. Hausmann S, *et al.* Genetic and biochemical analysis of yeast and human cap trimethylguanosine synthase: Functional overlap of 2,2,7-trimethylguanosine caps, small nuclear ribonucleoprotein components, PRE-mRNA splicing factors, and RNA decay pathways. *J Biol Chem.* **283(46),** 31706–31718 (2008).

120. Boon KL, Pearson MD, Koš M. Self-association of Trimethylguanosine Synthase Tgs1 is required for efficient snRNA/snoRNA trimethylation and pre-rRNA processing. *Sci Rep.* **5,** 11282 (2015).

121. Komonyi O, *et al.* DTL, the Drosophila homolog of PIMT/Tgs1 nuclear receptor coactivator-interacting protein/RNA methyltransferase, has an essential role in development. *J Biol Chem.* **280(13),** 12397–12404 (2005).

122. Hausmann S & Shuman S. Specificity and Mechanism of RNA Cap Guanine-N2 Methyltransferase (Tgs1). *J Biol Chem.* **280(6),** 4021–4024 (2005).

123. Mouaikel J, *et al*. Sequence-structure-function relationships of Tgs1, the yeast snRNA/snoRNA cap hypermethylase. *Nucleic Acids Res.* **31(16),** 4899–4909 (2003).

124. Singer MS & Gottschling DE. TLC1: template RNA component of Saccharomyces cerevisiae telomerase. *Science (80-.).* **266(5184),** 404–409 (1994).

125. Feng J, *et al*. The RNA component of human telomerase. *Science (80-.).* **269(5228),** 1236–1241 (1995).

126. Girard C, *et al.* Characterization of a short isoform of human Tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *J Biol Chem.* **283(4),** 2060–2069 (2008).

127. Egan ED & Collins K. Biogenesis of telomerase ribonucleoproteins. *Rna* **18(10),** 1747–1759 (2012).

128. Terns MP, *et al*. A common maturation pathway for small nucleolar RNAs. *EMBO J.* **14(19),** 4860–4871 (1995).

129. Ogg SC & Lamond AI. Cajal bodies and coilin - Moving towards function. *J Cell Biol.* **159(1),** 17–21 (2002).

130. Jia Y, *et al*. Early embryonic lethality of mice with disrupted transcription cofactor PIMT/NCOA6IP/Tgs1 gene. *Mech Dev.* **129(9-12),** 193–207 (2012).

131. Workmana E, Kolb SJ, Battle DJ. Spliceosomal small nuclear ribonucleoprotein biogenesis defects and motor neuron selectivity in spinal muscular atrophy. *Brain Res* . **1462,** 93–99 (2012).

132. Zhang Z. *et al.* SMN Deficiency Causes Tissue-Specific Perturbations in the Repertoire of snRNAs and Widespread Defects in Splicing. *Cell* **133(4),** 585–600 (2008).

133. Boulisfane N, *et al.* Impaired minor tri-snRNP assembly generates differential splicing defects of U12-type introns in lymphoblasts derived from a type I SMA patient. *Hum Mol Genet.* **20(4),** 641–648 (2011).

134. Lotti F. *et al.* A SMN-Dependent U12 Splicing Event Essential for Motor Circuit Function. **151(2),** 440–454 (2012).

135. Borg R & Cauchi RJ. The gemin associates of survival motor neuron are required for motor function in Drosophila. *PLoS One* **8(12),** e83878 (2013).

136. Cauchi RJ, Davies KE, Liu JL. A motor function for the DEAD-box RNA helicase, Gemin3, in Drosophila. *PLoS Genet.* **4(11),** e1000265 (2008).

137. Jablonka S, *et al.* Gene targeting of Gemin2 in mice reveals a correlation between defects in the biogenesis of U snRNPs and motoneuron cell death. *Proc Natl Acad Sci U. S. A.* **99(15),** 10126–10131 (2002).

138. Borg RM, *et al*. Genetic interactions between the members of the SMN-Gemins complex in Drosophila. *PLoS One* **10(6),** e0130974 (2015).

139. Grice SJ & Liu JL. Survival Motor Neuron Protein Regulates Stem Cell Division, Proliferation, and Differentiation in Drosophila. *PLoS Genet.* **7(4),** e1002030 (2011).

140. Timmerman C & Sanyal S. Behavioral and electrophysiological outcoms of tissue-specific Smn knockdown in Drosophila melanogaster. *Brain Res* . **13(1489),** 66–80 (2012).

141. Chang J, Schwer B & Shuman S. Mutational analyses of trimethylguanosine synthase (Tgs1) and Mud2: proteins implicated in pre-mRNA splicing. *RNA* **16(5),** 1018–31 (2010).

142. Jablonka S, *et al*. Distinct and overlapping alterations in motor and sensory neurons in a mouse model of spinal muscular atrophy. *Hum Mol Genet* **15(3),** 511–518 (2006).

Part II

Molecular diagnosis of the most common forms of Limb Girdle Muscular Dystrophy with autosomal recessive inheritance

Abbreviations

AldoA	Aldolase A
BMD	Becker muscular dystrophy
C2L	C2-domain-like
CaM	Calmodulin
CAPN1	μ-calpain or calpain 1
CAPN2	m-calpain or calpain 2
CAPN3	Calpain 3
CAPN4	Calpain 4
CBSW	Calpain-type beta-sandwich domain
CK	Creatine kinase
CysPc	Calpain-type cysteine protease domain
DMAT	Dystal myopathy with anterior tibial onset
DMD	Duchenne muscular dystrophy
DYSF	Dysferlin
EMG	Electromyogram
iPS	Induced pluripotent stem cell
LGMD1	Limb girdle muscular dystrophy whit autosomal dominant inheritance
LGMD2	Limb girdle muscular dystrophy whit autosomal recessive inheritance
LGMD2A	Limb girdle muscular dystrophy type 2A
LGMD2B	Limb girdle muscular dystrophy type 2B
LGMD2I	Limb girdle muscular dystrophy type 2I
LGMD2J	Limb girdle muscular dystrophy type 2J
LGMDs	Limb girdle muscular dystrophies
MARPs	Muscle ankyrin repeat proteins
MG53	Mitsigumin-53
MM	Miyoshi myopathy
MRI	Magnetic resonance imaging
NGS	Next-generation sequencing
PEF domain	Penta-EF-hand domain
PI(4,5)P2	Phosphatidylinositol 4,5-bisphosphate

rAVV	Recombinant adeno-associated virus
RyR	Ryanodine receptor
WES	Whole-exome sequencing

ABSTRACT

Limb Girdle Muscular Dystrophies (LGMDs) are a heterogeneous group of inherited progressive muscle disorders affecting predominantly the pelvic and shoulder girdle muscles. LGMDs are classified into type 1 or 2 depending on the different inheritance patterns. Type 1 LGMDs are inherited with an autosomal dominant pattern and are much less common than the type 2 LGMDs, which are inherited in an autosomal recessive pattern. LGMDs disorders result from defects in genes required for normal muscle function.

This project was carried out to achieve a molecular diagnosis for patients with clinical LGMD2 diagnosis. For that reason, we focused our studies on two genes responsible for the most frequent forms of LGMD2. The two most common forms are limb girdle muscular dystrophy type 2A and type 2B caused by mutations in *CAPN3* and *DYSF* gene, respectively. Another aim of this project was to amplify *CAPN3* and *DYSF* mutations spectrum, contributing to the study of genotype-phenotype relation in LGMD patients.

After the screening of 90 patients, seven pathogenic variants have been detected in seven patients belonging to five families in *CAPN3* gene. While three pathogenic variants have been detected in three patients in *DYSF* gene. Among these mutations, we found two novel variants in *CAPN3* gene and one novel variant in *DYSF* gene.

According to literature, our results highlight the genetic and phenotypic variability present in LGMD patients, especially in LGMD2A. In addition to clinical variability among patients, we detected intra-family variability in the C1 family. This leads to increase the difficulties to correlate a type of mutation with the onset and clinical progression of the disease. Moreover, even in our study, we found only one mutation in heterozygous state in three patients. As reported in literature ~22% of LGMD2A patients only one mutation is detected but the presence of another mutation is not excluded. However, in one of our patients, the presence of a second mutation seems unlikely. Thus, this result leads to hypothesize the presence of other factors involved in the pathogenesis of LGMD2A disease.

While for LGMD2B a genetic and phenotypic correlation seems less challenging. Among these patients, we found for the first time a mutation in the transmembrane domain of dysferlin protein. This case is particularly interesting since this missense mutation was found in the proband and in her asymptomatic daughter. We highlighted that this variant leads to an increased protein expression followed by a reduction of the protein at sarcolemma level. Then, it is likely that the 50% of mutant protein fails to insert to the sarcolemma, remaining within the cytoplasm. However,

the reduction of dysferlin has been seen only in the proband. It would be interesting to analyze the daughter to clarify if the reduction observed is only due to the presence of the missense variant or to the presence of another variant not detected.

1. INTRODUCTION

1.1 MUSCULAR DYSTROPHY

The term muscular dystrophies are reserved for classes of hereditary, progressive muscle diseases characterised by necrosis of muscle tissue and replacement by connective and fatty tissue, increase in fibre size variation, fibre splitting and centrally located myonuclei. Clinically, these disorders present themselves muscle weakness at any age from birth to middle years, frequently accompanied with high serum creatine kinase (CK) and result in significant morbidity and disability. These are disorders in which the primary pathological change is in the muscle itself and is not a consequence of disease of the central or peripheral nervous system (muscle atrophy)[1]. Muscular dystrophies occur as a result from mutations affecting proteins involved in muscle integrity and function. This group of disorders display clinical heterogeneity that reflects the heterogeneity of molecular mechanisms responsible for them. In fact, muscular dystrophies are among the most genetically heterogeneous human condition with up to 165 muscular dystrophies and myopathies-causative genes described[2]. Muscular atrophies differ from muscular dystrophies since muscle defects are not the primary pathological change as in muscular dystrophies but it is secondary to the diseases of the nervous system. The first serious attempt to separate muscular dystrophy from muscular atrophy was made by Leyden in 1876[3].

The first historical report of muscular dystrophy appeared in 1830 when Charles Bell wrote about an illness that caused progressive weakness in boys. After that a case was reported by Conte and Gioja in 1836, in which they described two brothers with progressive weakness starting at 10 years old. The two brothers showed symptoms that are now known to be characteristic of the milder Becker Muscular Dystrophy (BMD). At that time, however, Conte and Gioja did not describe that disorder as muscular dystrophy thus they did not achieve recognition for their discovery. Successively in 1851, Meryon reported a family with four children, all of whom were affected by significant muscle changes without abnormality in the central nervous system. Meryon hypothesised that the cause of the disorder could be the sarcolemmal defect and hypothesised furthermore that the disorder is genetically transmitted through females but affects only males[4]. The first comprehensive case was described by Guillaume Duchenne in 1868[5]. He described 13 patients with the same disease, which he called it "paralysie musculaire pseudo-hypertrophique". However, the term "muscular dystrophy" was first used by Wilhelm Heinrich Erb in 1891[6]. He was the first who finally convinced the medical opinion that dystrophies are due to primary degeneration of muscle, totally distinct from disorders in which muscular wasting is secondary to anterior horn cell disease in the spinal cord.

Over the next 100 years, genetic and phenotypically distinct forms of muscular dystrophy were recognised. The first gene responsible of muscular dystrophy was discovered later in 1970, when genetic studies linked the Duchenne gene to chromosome Xp21 and the cDNA and protein product were discovered in 1987[7,8]. So far, many other genes have been sequenced and mutations ascribed to specific forms of dystrophy. The genetically distinct forms of muscular dystrophies result from mutations in genes encoding:

- Integral structural proteins of the sarcolemmal membrane (α, β, γ and δ sarcoglycan, integrin α7)
- Structural proteins associated with the inner surface membrane (dystrophin, plectin)
- Structural proteins associated with the outer surface membrane (laminin α2, collagen type VI)
- Protein localised in specialised regions of the sarcolemma (caveolin 3)
- Proteins associated with the inner nuclear membrane (emerin, laminin A/C, plectin)
- Muscle specific proteins kinases (myotonic dystrophy protein kinase)
- Muscle-specific proteases (calpain)
- Protein involved in muscle membrane repair machinery (dysferlin)

The muscular dystrophies have been commonly subdivided on clinical criteria, particularly the pattern of muscle involvement, the age of onset and mode of inheritance.

<u>X-linked muscular dystrophy</u>:

-Duchenne muscular dystrophy
-Backer muscular dystrophy
-Emery-Dreifuss syndrome

<u>Autosomal recessive muscular dystrophy</u>:

-Congenital muscular dystrophy

-Limb girdle muscular dystrophy

<u>Autosomal dominant muscular dystrophy</u>:

-Limb girdle muscular dystrophy

-Facioscapulohumeral muscular dystrophy

1.2 LIMB GIRDLE MUSCULAR DYSTROPHIES

Limb girdle muscular dystrophies (LGMDs) is a group of neuromuscular disorders inherited with dominant and recessive manner. Initially, LGMDs were described as a distinct phenotype but now it is well known their heterogeneity. LGMDs are a clinically and genetically heterogeneous group of hereditary myopathies that are usually characterised by predominant weakness and atrophy with the prevalent involvement of proximal muscles of the hip and shoulder girdles. Moreover, elevated serum CK levels and dystrophic findings on muscle biopsy are commonly found and facial muscles are generally not involved. Muscle biopsy shows necrosis and regeneration of the muscle fibres with various levels of fibrosis and adipose tissue infiltration. The clinical spectrum of LGMDs can vary from very mild forms which allow patients to maintain a fairly normal life to much more severe deterioration of proximal limb muscles, that cause dramatic physical weakness along with a short-ended life span[9–11]. Then, progression, distribution of weakness and wasting, respiratory involvement and heart damage vary considerably among individuals and genetic subtypes[12,13]. The onset of symptoms may occur at almost any age, with the exception of infancy, which would indicate the presence of a congenital muscular dystrophy[14]. The clinical involvement is usually limited to skeletal muscle, in some cases cardiac and respiratory involvement can be variably presented and cognitive impairment is usually absent. Clinical variability is presented among patients of the same subtype, or even within the same family. For the clinical heterogeneity has been challenging to reach a classification of LGMDs, which it has been revised continuously[15–17].

The first description of LGMDs goes back to late 18th century when Erb and Mobius described patients having a weakness with primary involved the shoulder and the hip girdle. Only in 1954 LGMDs were defined as a distinct nosological entity by John Walton and Federick Nattrass. They separated these disorders from the X-linked dystrophinopathies such as Duchenne and Becker Muscular Dystrophies (DMD/BMD)[15]. They defined LGMDs thanks to a few common features:

- Transmission usually via an autosomal recessive gene.
- Usually late onset.
- Muscular weakness in shoulder and/or pelvic girdle, without any facial weakness.
- Relatively slow course.

In the late 1980s, the development of immunostaining methods helped to distinguish between LGMDs and other conditions such as BMD[18]. Moreover, in 1995 the European Neuromuscular Centre

Workshop established more precise criteria for diagnosis and classification of LGMDs according to their onset, progression and mode of inheritance[19]. Over the last sixty years, the molecular genetic investigations and the increased in knowledge of the structure and function of the muscle have allowed a more precise classification and improvements in understanding of the pathogenesis of LGMD subtypes. In the last five years, new causative genes were identified thanks to high-throughput technologies such as whole-exome sequencing (WES) and next-generation sequencing (NGS)[20–24]. To date, over 30 distinct subtypes of LGMD have been identified. LGMDs are classified into two main groups depending on the inheritance pattern:

- LGMD type 1 (LGMD1), autosomal dominant (AD) forms
- LGMD type 2 (LGMD2), autosomal recessive (AR) forms

Among these two groups, each LGMD type is indicated with a letter given in the order of gene mapping. Of 31 LGMD subtypes recognised to date, 8 dominant forms and 23 recessive forms, each corresponding to a different causative gene[10,25,26] (Tab.1).

LGMD form	Locus/Gene	Protein	Onset	Protein properties	MIM	Prevalence
			Autosomal dominant LGMD			
LGMD1A	5q31.2 (TTID)	Myotilin	IV-VI decade	Component of Z-band	159000	<1/1 000 000
LGMD1B	1q21.2 (LMNA)	Lamin A/C	I-II decade	Nuclear lamina protein	159001	1-9/1 000 000
LGMD1C	3q25.3 (CAV3)	Caveolin 3	> 5 years	Component of caveola	607801	<1/1 000 000
LGMD1D	7q36.3 (DNAJB6)	HSP40	III-VI decade	Co-chaperon	603511	<1/1 000 000
LGMD1E	2q35 (DES)	Desmin	II-III decade	Intermediate filament	125660	<1/1 000 000
LGMD1F	7q32 (TNO3)	Transportin 3	I-V decade	Nuclear transport receptor	608423	<1/1 000 000
LGMD1G	4p21 (HNRNPDL)	Heterogeneous nuclear ribonucleoprotein D-like protein		Transcriptional regulator	609155	<1/1 000 000
LGMD1H	3p23-p25	--			613530	<1/1 000 000
			Autosomal recessive LGMD			
LGMD2A	15q15.1 (CAPN3)	Calpain 3	I-III decade	Protease	253600	1-9/100 000
LGMD2B	2p13.2 (DYSF)	Dysferlin	II-III decade	Membrane regeneration and repair	253601	1-9/1 000 000
LGMD2C	13q12.12 (SGCG)	γ-Sarcoglycan	I decade	DAP compex	253700	1-9/1 000 000
LGMD2D	17q21.33 (SGCA)	α-Sarcoglycan	I-II decade	DAP compex	608099	Unknown
LGMD2E	4q12 (SGCB)	β-Sarcoglycan	I-II decade	DAP compex	604286	1-9/1 000 000
LGMD2F	5q33.3 (SGCD)	δ-Sarcoglycan	I decade	DAP compex	601287	1-9/1 000 000
LGMD2G	17q12 (TCAP)	Telethonin	I-II decade	Sarcomere protein	601954	<1/1 000 000
LGMD2H	9q33.1 (TRIM 32)	Tripartite motif containing 32	I-II decade	Protein degradation	254110	Unknown
LGMD2I	19q13.32 (FKRP)	Fukutin realted protein	I-III decade	Putative glycosyltransferase	607155	1-9/1 000 000
LGMD2J	2q31.2 (TTN)	Titin	I-III decade	Sarcomere protein	608807	<1/1 000 000
LGMD2K	9q34.13 (POMT1)	Protein-O-mannosyl transferase 1	I decade	Putative glycosyltransferase	609308	<1/1 000 000
LGMD2L	11p14.3 (ANO5)	Anoctamin 5	II-IV decade	Putative calcium-activated chloride channel	611307	Unknown
LGMD2M	9q31.2 (FKTN)	Fukutin	< 6° month	Putative glycosyltransferase	611588	<1/1 000 000
LGMD2N	14q24.3 (POMT2)	Protein-O-mannosyl transferase 2	< 1° year	Putative glycosyltransferase	613158	<1/1 000 000
LGMD2O	1p32 (POMGnT1)	Protein-O-linked mannose beta 1,2 nacetylglucosaminyl transferase	II decade	Putative glycosyltransferase	613157	<1/1 000 000
LGMD2P	3p21 (DAG1)	Dystroglycan	I decade	DAP compex	613818	<1/1 000 000
LGMD2Q	8q24 (PLEC1)	Plectin	I decade	Cytolinker	613723	<1/1 000 000
LGMD2R	2q35 (DES)	Desmin	III decade	Sarcomeric microtubule-anchoring protein	615325	<1/1 000 000
LGMD2S	4q35 (*TRAPPC11*)	Transport protein particle complex 11	III decade	Involved in Golgi trafficking	615356	<1/1 000 000
LGMD2T	3p21(*GMPPB*)	GMPPB	II-III decade	Glycosylation pathways	615352	<1/1 000 000
LGMD2U	7p21(ISPD)	Isoprenoid synthase domain	Variabile	Required for protein O-linked mannosylation	616052	<1/1 000 000
LGMD2V	17q25.3 (GAA)	Alpha-1,4 glucoside	Variabile	Lysosomal enzyme		
LGMD2W	2q14 (LIMS2)	Lim and senescent cell antigen-like dmain 2	II decade	Modulating cell spreading and migration	616827	<1/1 000 000

Tab1: Description of limb girdle muscular dystrophies with autosomal and recessive inheritance known so far.

The LGMDs are rare disorders, whose different forms range mainly from 1 in 14500 to 1 in 123000, depending on the population. The autosomal recessive forms are much more common compared to the autosomal dominant forms that probably represent less than 10% of all LGMD cases; LGMD2A is the most frequently diagnosed among LGMDs worldwide followed by LGMD2B[25,27].

From our current state of knowledge and considering the diversity of proteins involved in the different forms of LGMDs, no universal treatment can be envisioned. Case by case approaches is necessary.

1.2.1 Limb girdle muscular dystrophy type 2A

Limb girdle muscular dystrophy type 2A (LGMD2A; OMIM 253600), also known as calpainopathy. It was the first form of limb girdle muscular dystrophy with a recessive inheritance identified and the

first disorder discovered caused by a deficiency of a non-structural protein. *CAPN3* is the gene responsible for LGMD2A and it was mapped to chromosome 15 in 1990[28]. *CAPN3* gene encodes a muscle-specific nonlysosomal proteolytic enzyme. Mutations lead to absence or reduction in this enzyme, but how this result in the dystrophic process is not fully understood. *CAPN3* gene was linked with the first form of LGMD2 thanks to a molecular genetic study performed in a group of families from the island of La Réunion and it has been classified as LGMD2A, in 1991[29]. Only two years later, linkage analysis in Brazilian families provided the first evidence of genetic heterogeneity for LGMD2A[30].

LGMD2A is the most common subtype between LGMDs and is up to 30% of all recessive LGMD cases taken together. The age of onset of muscle weakness range from two to 40 years but the disease usually first appears in the second or third decade of life[31]. The course of the disease is highly variable and a wide intra- and iter-familiar clinical variability is reported. Commonly, early onset means more severe phenotypes but there are cases in which a later onset leads to rapid progression and inability to walk by the third decade[32]. Use of wheelchair could become necessary during the second through the fifth decades, but some patients may remain ambulatory after the sixth decade[33]. Calpainopathy is characterised by the symmetrical and progressive weakness of proximal muscles more than distal muscles limbs, trunk and periscapular area. Laxity of the abdominal muscles, joint contractures, scapular winging, scoliosis and Achille tendon shortening. Moreover, patients with LGMD2A have the tendency to walk on tiptoe, difficulty in running, climbing stairs and in raising the arms above the head. Facial, tongue and neck muscles are not usually involved, therefore swallowing problems are unlikely. Heart problems and intellectual disability are not reported in this condition. Life expectancy is generally a normal range since heart and breathing muscles are usually not affected.

There is currently no treatment or cure available for LGMD2A, however, some interventions may be helpful for certain symptoms of the disease. For example, stretching and physical therapy may help prevent or lessen muscle contractures. Technical aids can also compensate for the loss of certain motor abilities and surgical intervention may be required for correction of orthopaedic complications. There are some studies ongoing to reach a therapy for LGMD2A disease but so far, research is still on mouse model. Bartoli *et al.*, evaluated the potential of recombinant adeno-associated virus (rAAV) vectors for gene therapy in a murine model for LGMD2A. In this study a muscle-specific promoters was used to avoid calpain 3 cell toxicity and it was reported an efficient and stable transgene expression in muscle with the restoration of the proteolytic activity and without evident toxicity[34]. Another research group succeeded, thanks to gene editing tools, to correct a LGMD2A patient-

specific iPS cell line with a homozygous deletion of exon 17 to 24 in the *CAPN3* gene. Now some transplantation studies are ongoing with corrected iPS cells in order to evaluate the potential of this new cell therapy[35].

1.2.2 Limb girdle muscular dystrophy type 2B

Limb girdle muscular dystrophy type 2B (LGMD2B; OMIM #253601) is due to autosomal recessive mutations in *DYSF* gene localised in chromosome 2p13.2. This disorder is caused by reduction or absence in skeletal muscle of dysferlin protein encoded by *DYSF* gene. Dysferlinopathies are the first class of diseases identified where it has been suggested a fault in the process of membrane repair rather than a structural weakness of the sarcolemma[13,36,37]. The classification of different dysferlinopathies is based on the initial symptoms of patients. Clinical presentations most commonly associated with dysferlinopathies include: classical Limb Girdle Muscular Dystrophy type 2B (LGMD2B) characterised by the initial involvement of proximal muscles, Miyoshi Myopathy (MM; #OMIM 254130) with the involvement of distal muscles and Distal Myopathy with Anterior Tibial onset (DMAT; #OMIM 606768) characterised by predominant involvement in the anterior compartment[38,39]. These are the three main clinical phenotypes associated with mutations in *DYSF* gene. However, even other clinical phenotypes may be produced by similar mutations[40,41].

The existence of a recessive form linked to chromosome 2p was first shown in 1994 and later confirmed by Passos-Bueno *et al.* in other three families. This disorder has been classified as LGMD2B[42,43]. In 1995, the linkage of Miyoshi Myopathy to the same region of chromosome 2p as LGMD2B was noted[44]. But only in 1998, thanks to *DYSF* gene cloning was clearly demonstrated that these two disorders were caused by the same gene defect[45]. The following year Bushby coined the term "dysferlinopathy" in order to group these disorders[46]. While for the third form of dysferlinopathy (DMAT) the linkage to chromosome 2p was noted only in 2001[38].

The onset of LGMD2B is usually between the second and the third decades. After the first year of evolution, the disease mainly affects lower limbs but later even paravertebral and proximal upper girdle muscles are affected. Generally, head, neck and cardiac muscles are spared. No central nervous system involvement has been described and respiratory involvement is mild and occurs in the later stage of the disease. Patients have difficulty standing on their tiptoes, mild scapular winging may be evident at any stage of the disease and patients are characterised by elevated levels of serum CK. One of the distinctive phenotype is the early loss of the Achilles' tendon reflexes, is peculiar of this disorder since in other forms of LGMDs is the most preserved reflex. Loss of ambulation usually occurs on average after years of disease course[47–49]. Disease progression is slow, although several

patients show abrupt onset and rapid progression to a non-ambulatory state. Moreover, LGMD2B show intra- and interfamilial variability in the age of onset, the pattern of muscle involvement and rate of disease progression[50–53].

To date there are no specific treatments for LGMD2B, the only therapeutic approaches are palliative in order to allow patients to delay loss of ambulation. Possible future therapies, as in LGMD2A, including gene therapy replacement of affected gene and use of stem cells[54–58]. A Phase 1 clinical trial (NCT02710500) is ongoing in LGMD2B patients to evaluate the safety and tolerability of intramuscular injection of rAAVrh.74.MHCK7.DYSF.DV. This is a rAAV carrying a dysferlin transgene under control of a muscle specific promoter.

1.2.3 Causative genes in LGMD2A and LGMD2B

Limb girdle muscular dystrophy type 2A is caused by mutations in *CAPN3* gene. It is located on chromosome 15, covers a genomic region of 52.8kb and consists of 24 exons. *CAPN3* gene (NC_000015.10) is expressed as a 3.5kb transcript and is translated into a 94kDa protein composed of 821 amino acids with proteolytic function. The human *CAPN3* gene produces many splicing variants and only the full-length isoform is tissue specific and it is expressed exclusively in skeletal muscle. The other isoforms are characterised by the splicing of exons 6, 15 and 16[59]. In human lymphoblastoid cell lines in contrast to adult muscle tissue, exon 15 is systematically spliced out. White blood cells are characterised by the presence of four different *CAPN3* transcripts generating by the alternative splicing of exons 6, 15 and 16[60]. So far, 497 variants have been identified, according to the Leiden Muscular Dystrophy database (www.dmd.nl) and they are spread in the entire length of the gene with the absence of mutational hot spots.

Limb girdle muscular dystrophy type 2B is caused by mutations in the *DYSF* gene (NC_000002.12), it encodes for dysferlin protein, a member of the ferlin protein family. *DYSF* gene spreads over 150kb of genomic DNA and is located on chromosome 2p13[42]. *DYSF* gene encodes for a 237kDa single pass transmembrane protein composed of 2080 amino acids and produces seven different transcripts characterised by exon 17 skipping and the insertion of exon 5a and 40a[61]. These transcripts are differentially expressed in skeletal muscle and in white blood cells. The skipping of exon 17 is specifically found in white blood cells and does not occurs in skeletal muscle in which in 90% of cases the full-length isoform is present. The insertion of exon 5a and 40a happen in both tissues[62]. So far, 511 variants have been identified, according to the Leiden Muscular Dystrophy database (www.dmd.nl) and they are spread over the entire length of the gene without mutational hot spots.

1.3 OVERVIEW OF PROTEINS INVOLVED IN LGMD2A AND LGMD2B DISEASE

1.3.1 Calpain 3

Calpain 3 is a member of calpain protease family. Calpain family was described for the first time in 1964 by Gordon Guroff. In 1981, the term "calpain" was used by Takashi Murachi but it became the official name of this protease family only ten years after[63–65]

Calpains belong to the papain superfamily of cysteine proteases, which are a family of complex multi-domain intracellular enzymes. Calpains are not degradative enzymes but instead, carry out cleavage of target proteins in response to calcium signalling. Therefore, calpain is considered to be a modulator protease rather than a degrading protease such as proteasomal and lysosomal proteases[66]. The mammalian calpain gene superfamily contains 16 known genes: 14 of these genes encode proteins that contain cysteine protease domains while the other two genes encode smaller regulatory proteins. These small regulatory proteins join with some of the catalytic subunits to form heterodimeric proteases[67,68]. Until 1989, the two best-studied calpains in mammalian were called "conventional" calpains. These are the most abundant and constitutively expressed in all tissues, whereas many other isoforms demonstrate tissue-specific expression pattern. These conventional calpains are μ-calpain and m-calpain or even called CAPN1 and CAPN2. The term μ- and m- was given for the difference in their calcium requirement. The designations μ- and m- are short forms of micromolar and millimolar Ca^{2+} required for their activation. CAPN1 and CAPN2 are heterodimers composed of one catalytic subunit of about 80kDa encoded by *CAPN1* or *CAPN2* genes and a common smaller subunit of about 28kDa encoded by *CAPN4* gene that is the regulatory subunit[69]. The primary structure of the conventional calpain catalytic subunit was determined in 1984[70]. The structure consists of four domains (I-IV): domain I represents the N-terminal region and it has a regulatory role, domain II contains a catalytic site called calpain-type cysteine protease conserved (CysPc) domain and it is responsible for proteolytic functions. The CysPc domain is composed of two protease core subdomains (PC1 and PC2)[71–74]. This domain is characterised by the presence of the catalytic triad residues cysteine[129], histidine[334] and asparagines[358-75]. Domain III is known to interact with Ca^{2+} and phospholipids, previously called C2-domain-like (C2L) but now is called calpain-type beta-sandwich (CBSW) domain. The last domain represents the Ca^{2+} binding module containing five EF-hand motifs (PEF domain)[74,76]. Calpains with the same structure as CAPN1 and CAPN2 are called "classical or conventional" calpains, while the rest are "non-classical". In addition to sharing the same structure, conventional calpains have a specific endogenous inhibitor: calpastatin, which contains four repetitive inhibitory units, each of which inhibits an equimolar amount of conventional calpain[77] (Fig.1).

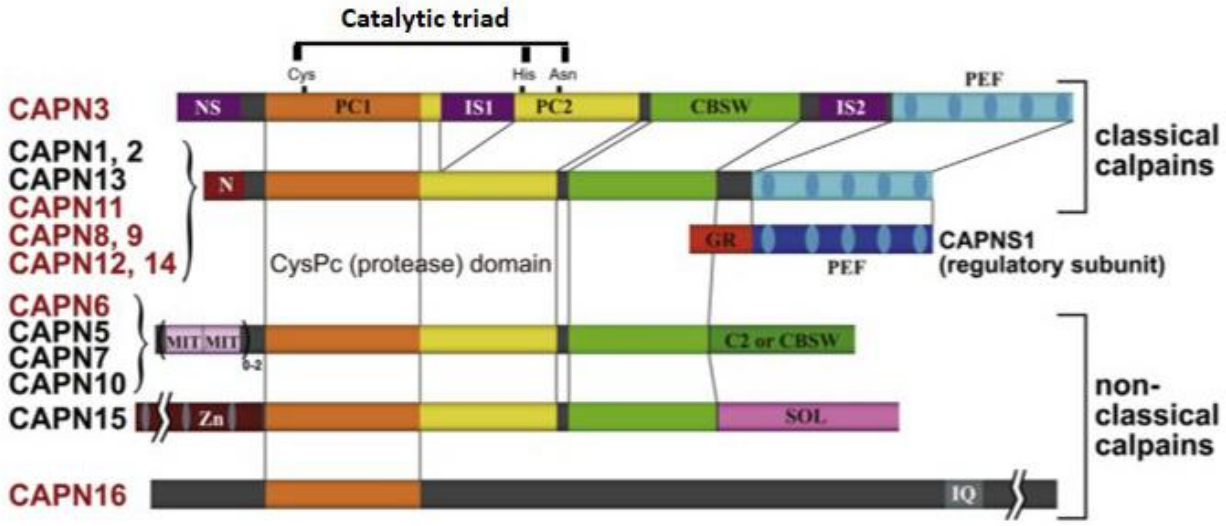


Fig1: Schematic shows the structure of human calpains. The calpains shown in red are predominantly expressed in specific tissues or organs while those in black are more widely expressed. (Image taken by Ono *et al.*, 2016[74]).

Genetics studies have revealed that individual calpains have important and widely divergent physiological functions. The calpain family of proteases has been implicated in a variety of calcium-regulated cellular processes such apoptosis, proliferation, differentiation, signal transduction and cellular migration[69,78]. For example, CAPN1 and CAPN2 are involved in a variety of processes whose the best documented role is the remodelling of cytoskeleton for processes such as cell movement[67,79,80]. Before the influx of Ca^{2+}, calpains exist in the cytosol as an inactive enzyme but in response to increases in the cellular Ca^{2+} levels, the heterodimer undergoes multiple structural changes to form the active calpain enzyme. Structural events, such as autoproteolysis, subunit dissociation, intradomain/interdomain rearrangement and phospholipid binding, are suggested to be involved in the complex regulation of activation[68,81–83]. The influx of Ca^{2+} is responsible for calpains activation. Calpains are inactive until their protease domain, which comprises two core domains (PC1 and PC2), fuse to form a functional protease, secondary to the binding with Ca^{2+}. When protease domain is formed, the activated calpains are able to cleave the target proteins[84]. Proper for the huge number of cellular processes in which calpains are involved, not only a deficiency in CAPN3 leads to a LGMD2A but even a deficiency in other calpains lead to numbers of disorders called calpainopathies (Tab.2).

Gene	Chr.	Protein	Other names	Tissue distribution	Protease activity	Association with small subunity	Deficiency phenotype
CAPN1	11q12–13.1	Calpain 1	μ-calpain large subunit	Ubiquitous	+	+	Platelet dysfunction
CAPN2	1q32-41	Calpain 2	m-calpain large subunit	Ubiquitous	+	+	Embryonic lethal
CAPN3	15q15	Calpain 3	p94, nCL-1	Skeletal muscle	+	-	Muscular dystrophy
CAPN5	11q14	Calpain 5	hTRA-3, nCL-3	Ubiquitous	+	-	Vitreoretinopathy
CAPN6	Xq23	Calpain 6	CANPX	Embryonic muscle, placenta	-	-	Hypergenesis
CAPN7	3p24	Calpain 7	PalBH	Ubiquitous	+	-	
CAPN8	1q32–41	Calpain 8	nCL-2	Gastointestinal tract	+	-	Gastric ulcer
CAPN9	1q42.1–43	Calpain 9	nCL-4	Gastointestinal tract	+	+	Gastric ulcer
CAPN10	2q37.3	Calpain 10	_	Most cells	ND	ND	Type 2 diabetes
CAPN11	6p12	Calpain 11	_	Testis	ND	ND	
CAPN12	19q13	Calpain 12	_	Hair follicle	ND	ND	
CAPN13	2p21–22	Calpain 13	_	Ubiquitous	ND	ND	
CAPN14	2p21–22	Calpain 14	_	Esophagus	ND	ND	Eosinophilic Esophagitis
CAPN15	16p13.3	Calpain 15	SOLH, Calpain small subunit, 30K	Ubiquitous	ND	ND	
CAPNS1	19q13	CAPNS1	CAPN4	Ubiquitous	-	+	

Tab.2: List of human calpains. (Table modified by Suzuki *et al.*,2004[68]).

1.3.1.1 Calpain 3 protein

Calpain 3 (NP_000061.1), indicated also as CAPN3, p94, nCL-1 and nCANP, was the first non-ubiquitously expressed calpain to be identified. Discovered in 1989 by Sorimachi *et al.*, but is still not completely characterized, due to its extraordinary instability at least *in vitro*[85].

Calpain 3 was originally described as a skeletal muscle specific calpain isoform[86,87]. More precisely, calpain 3 is expressed predominantly, rather than exclusively, in skeletal muscle. It is expressed more strongly in skeletal muscles, especially in the fast (type II) fibres[88]. In the heart, CAPN3 mRNA has been detected but not its protein. While the protein is expressed in in peripheral blood mononuclear cells, in which there are four different splice variants of CAPN3[54,55,83,84]. Structurally, it is categorised as a classical calpain and it is approximately 50% identical to CAPN1 and CAPN2. One of the features of CAPN3 that distinguish it from CAPN1 and CAPN2 is the insertion of additional regions (known as NS, IS1 and IS2). This is the reason why the molecular weight of CAPN3 is 94kDa and not 80kDa as the other calpains. NS is an additional N-terminal sequence with unknown function, IS1 is localised between the catalytic cysteine[126] and histidine[334] and IS2 is upstream of the first EF hand motif and it contains nuclear translocation signal[85] (Fig.1). These sequences are found only in the full-length isoform of CAPN3. Several studies have revealed other unique characteristics of this protein that diverge greatly from those of the conventional calpains[91]. For instance, CANP3 shows extremely rapid Ca^{2+} -independent autolytic activity. The extreme rapidity of autolysis depends on IS1 and IS2

sequences and seems to be independent of Ca^{2+} but is activated by Na^{+}. This feature is unique among known calpains since Na^{+} dependence is generally considered a property of extracellular enzymes and not intracellular enzyme as calpains[86,92]. So far, CAPN3 is the only Na^{+} dependent intracellular enzyme known. It is activated by physiological Ca^{2+} but with levels much lower than other classical calpains. CAPN3 is activated additively by both ions and the Na^{+} and Ca^{2+} required for autolysis are 100 and 0.1mM, respectively[74]. This unique features present in calpain 3 depend on the presence of IS1 and IS2 insertion. The deletion of these sequences completely abolishes Na^{+} dependent activation without changing its Ca^{2+} dependent activity[92]. The lower levels of Ca^{2+} required for CAPN3 activation could be explained by the presence of IS1 and IS2 insertions which probably add more flexibility to the protein, leading to the formation of the catalytic active site with less Ca^{2+} levels[91]. Thanks to these two activation signals, CAPN3 molecules undergoes intramolecular autolysis in IS1 insertion. This event is followed by activation of CAPN3 molecules which received the activation signals, these fully activated molecules can thereafter activate other CAPN3 molecules which have not received the activation signal by intermolecularly autolysis. This ultimate step generates an amplification cascade leading to global activation of CAPN3 pool[93]. The intra- or intermolecular cleavage leads to the formation of two fragments that remain together by non-covalent binding to form the active enzyme. For the disassembling of autolyzed fragments is needed cleavage of IS2 sequence[94].

As in other PEF domain proteins, the EF-hands of calpain 3 are paired: EF1 paires with EF2, EF3 paires with EF4 and EF5 of one subunit pairs up with EF5 of the other subunit to form an homodimer[95,96]. This is another distintive fature of CAPN3 since EF5 domain of other calpains is used exclusively to form heterodimers with small regulatory subunits. Moreover EF5 of calpain 3 compare to the other calpains is able to bind Ca^{2+} ion[97]. Finally, CAPN3 differs to other calpains by the interaction with connectin/titin, the giant sarcomere protein and by the presence of different inhibitors compared to the other conventional calpains. For example, calpastatin has no effect on CAPN3 autolysis. However, other inhibitors exist to control CAPN3 activity since proteolysis, which leads to irreversible modifications, is a widely used and powerful means to regulate cellular functions. Uncontrolled proteolysis may be deleterious for the cell. Then, multiple strategies exist to control such events. Given the very low calcium requirement of CAPN3 and its high instability, it has been hypothesized that cellular activators and inhibitors of CAPN3 must exist in order to modulate its function. It has been shown that the interaction with the sarcomere protein titin suppresses IS2 autolysis suggesting that titin is important for the stabilizing and the regulating of CAPN3[98]. For non-sarcomere associated CAPN3, the titin function is performed by the protein PLEIAD, which

moderates the protease activity of the enzyme[99]. Recently, it has been shown the presence of a positive regulator of CAPN3 activity, calmodulin (CaM) that is able to binds and facilitates CAPN3 autolytic activation[100].

1.3.1.2 Main functions of CAPN3

The specific role of CAPN3 and its link to deficits in muscle adaptation is still under investigation. The research of specific substrates of calpain 3 has been so far unsuccessful and despite intense research, the physiological functions of calpain 3 and its pathophysiological implications remain elusive. CAPN3 protein is located at several different cellular sites and may possess both proteolytic and non-proteolytic functions and it is important for homeostasis of skeletal muscle. CAPN3 is associated with the sarcomere in skeletal muscle, thanks to the binding with titin. However, calpain 3 has been observed in other cellular compartments beside the sarcomere, such as the membrane in myogenic cells or the triad associated protein complex, suggesting that calpain 3 has multiple functions within the muscle tissue[74,101,102].

The most well known function of calpain 3 is its role in the maintenance of sarcomere integrity via the regulation of sarcomeric protein turnover. This function is performed by the pool of CAPN3 associated with the sarcomere, which is the major pool of calpain 3. It may play a role in signalling pathways and in releasing damaged proteins to facilitate myofibrillar proteins turnover in order to maintain an effectively functioning muscle[101]. The full-length mature CAPN3 is incorporate into myofibrils via titin only once the sarcomere structures are fully mature[103]. In response to external stress, CAPN3 is redistributed from the M-line to the N2A region of titin thanks to its protease activity. The redistribution of the protein leads to dissociation of muscle ankyrin repeat proteins (MARPs) from titin. MARPs are a family of titin-associated stress responsive molecules and are putative transducers of stretch-induced signalling in skeletal muscle[104]. Their dissociation from titin is followed by the translocation to the nucleus to transmit signals of mechanical perturbation[101,103,105,106]. In addition of this, CAPN3 seems to be involved in myofibrillar protein turnover. Myofibrillar proteins must be removed from the myofibril before they can be turned over metabolically and CAPN3 seems to be involved in this step acting upstream of ubiquitin-proteasome pathway by the cleavage and release of myofibrillar proteins[107].

A portion of cellular CAPN3 is concentrated at the triad, in which is necessary for the triad-associated protein complex[74,102]. In this case, CAPN3 has a structural role rather than a proteolytic one, since it is necessary for proper recruitment of two components of the triads: the glycolytic enzyme Aldolase A (AldoA) and ryanodine receptors (RyRs). CAPN3, in this case, doesn't proteolyze ryanodine

receptors or Aldo A but plays a role in the recruitment and for the proper localization of these two triad components[108]. Indeed, in the absence of CAPN3, the concentration of both proteins in triad fractions is reduced and the calcium release upon activation is significantly decreased.

Calpain 3 is also found in the membrane of myogenic cells where is involved in the myogenic differentiation process participating in the generation of a pool of reserve cells. This due by decreasing the transcriptional activity of the key myogenic regulator MyoD via proteolysis independently of the ubiquitin-proteasome degradation pathway. This means a potential rule of CAPN3 in the muscular regeneration process by promoting the renewal of the satellite cell compartment[109]. Regeneration of adult skeletal muscle is a process requiring the activation, proliferation and fusion of satellite cells, to form new muscle fibres[110]. The myogenic differentiation of satellite cells and their myoblast progeny is regulated by various transcriptional factors including MyoD[111,112]. In the early stage of myogenic differentiation, calpain 3 is involved in the regulation of the myogenic regulatory factor MyoD by inducing its destabilization and leading myoblasts to quiescence. Calpain 3 induces the proteolysis of MyoD and promotes the generation of a pool of reserve cells.

CAPN3 seems even to participate in membrane repair machinery with dysferlin protein, that is one of the proteins involved in this. In normal myofibers CAPN3 is inactive through interaction with titin, but with the injury of the membrane, Ca^{2+} influx leads CAPN3 activation. Active calpain 3 may contribute to the patch repair response by cleaving proteins to mediate disassembly of the damaged actin cytoskeleton. After that, intracellular vesicles are aggregated to form an endomembrane patch that is trafficked to the site of disruption. Dysferlin molecules present on repair vesicles and on the plasma membrane mediate the docking and fusion of the patch[113,114].

1.3.2 Dysferlin

Dysferlin belongs to the ancient ferlin family. Ferlins are detected in all eukaryotic kingdoms, including unicellular phytoplankton and protozoans[115]. They are a group of single-pass transmembrane proteins that possess multiple tandem cytosolic C2 domain, from five to seven, variable numbers of Fer and DysF domains and a short C-terminal extracellular domain. There are six known human ferlins including dysferlin, otoferlin and myoferlin (Fig.2).

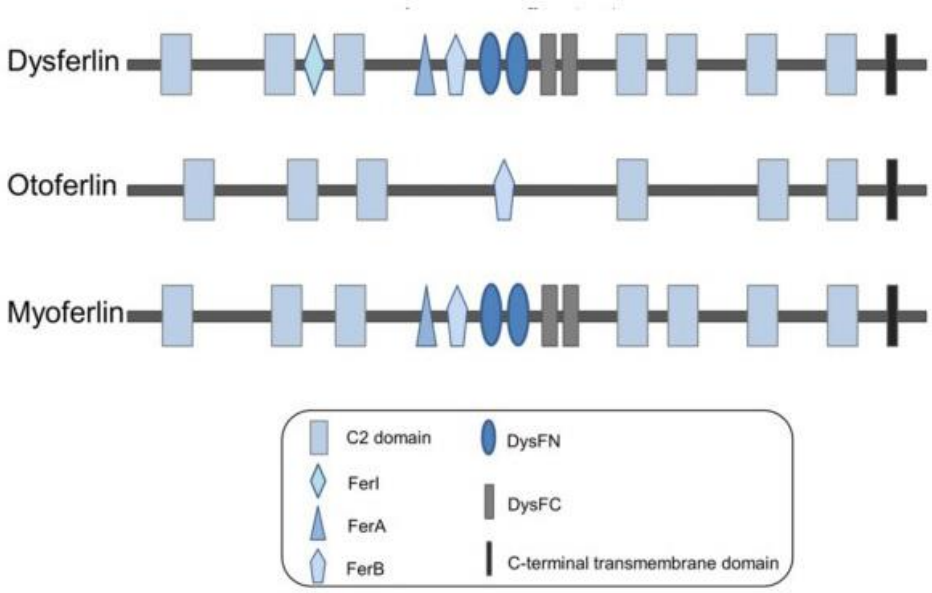


Fig.2: Structure of three of the six known human ferlins. The structure is characterized by a short C-terminal transmembrane domain, five to seven tandem C2 domains and a variable number of DysF and Fer domains. (Image taken by Cardenas *et al.,* 2016[53]).

The function of Fer and DysF domains remains unknown. Most actions of ferlins appear to be mediated by their C2 domains, which mediate lipid and protein binding. Moreover, many C2 domains coordinate calcium ions, within a negative charged binding pocket, to regulate calcium-dependent events[123,124].

The ferlins were identified and named based on homology to a *Ceanorhabditis elegans* spermatogenesis factor Fer-1, required for fusion of vesicles with the spermatid cell membrane[116]. Much currently remains unknown about the fundamental function of this gene family, however, mutations in the two most well-characterized members, dysferlin and otoferlin, have been implicated in human disease[43,117]. Dysferlin like myoferlin is abundantly expressed in the plasma membrane, and both proteins are involved in membrane repair and muscle regeneration[119–122]. Whereas otoferlin mainly localized in intracellular compartments where it plays a role in exocytosis[53,118].

Dysferlin is ubiquitously expressed with high expression in skeletal muscle, heart and brain. Dysferlin is a 230 kDa type II anchored membrane protein with a single C terminal helix buried in the membrane. The structure of the protein consists of seven C2 domains (C2A to C2G), three Fer domains (FerA, FerB and FerI), two couple of DysF domains and a C terminal trans-membrane domain. In dysferlin, C2A domain binds phospholipids in a calcium dependent manner and it is essential for phosphatidylserine accumulation at the membrane repair site[125]. The other C2 domains show calcium independent binding to phospholipids[126,127]. Dysferlin is a sarcolemmal protein distinct

from the dystrophin-glycoprotein complex (DGC) in human muscle. It is not only localised in the sarcolemma but it shows localisation to intracellular and endosomal vesicles and there is even a minor portion associated with the T-tubule system as well[102,128,129]. Dysferlin in T-tubule membranes contributes to the maintenance of the T-tubule system and it colocalizes with dihydropyridine receptor, caveolin-3, MG53, annexin A1 and AHNAK[128,130]. Dysferlin contributes to the biogenesis and remodelling of the T-tubule systems since it is able to recruit the phosphatidylinositol 4,5-bisphosphate (PI(4,5)P2), that is an important component of the T-tubule complex[131].

1.3.2.1 Main function of DYSF

Dysferlin has been implicated in various cellular processes, such as muscle cell-cell fusion during regeneration, muscle growth and cell adhesion[132–137]. An increasing number of dysferlin interacting proteins have been described including caveolin-3, affixin, annexins A1 and A2, calpain 3, myogenin and AHNAK[114,130,132,138–141]. However, the major function of dysferlin is its role in sarcolemma repair machinery since is involved in patch formation to reseal the membrane lesion. Dysferlin is a critical component of muscle membrane repair system. Mechanical and chemical stress lead to plasma membrane distruption[142]. Tiny membrane injuries can be repaired spontaneously while for larger injury is required the help of an active membrane repair machinery. It is proper in the active membrane repair machinery that dysferlin plays a key role. The membrane disruptions cause a massive entry of Ca^{2+} to the cell, with the consequent recruitment of vesicles that fuse with each other and with the plasma membrane to form the patch for resealing membrane lesion[143].

The exact mechanism of membrane repair is still unclear, what we know so far is the presence of two membrane repair machinery, one Ca^{2+}-dependent and the other Ca^{2+}-independent. The first step after membrane injury is the accumulation of mitsigumin-53 (MG53) at injury sites, not in a Ca^{2+}-dependent manner but in an oxidation-dependent manner [144,145]. This step is essential for active trafficking of intracellular vesicles to the sarcolemma and is required for the movement of dysferlin to the site of cell injury during membrane repair patch formation[146,147]. During the Ca^{2+}-dependent phase of membrane repair, the influx of ions leads to the activation of m-calpain and μ-calpain, but not calpain 3, and strong depolarization of L-type calcium channels (L-type VGCC). Activated calpains cleave dysferlin releasing a C-terminal fragment called mini-dysferlin $_{C27}$[149]. The calpain cleavage site in dysferlin is predicted to residue in exon 40a[150]. After that cleavage mini-dysferlin $_{C27}$ isoforms are rapidly recruited to sites of membrane injury and fuse with MG53 in a process coordinated by L-type VGCC. In the site of injury dysferlin molecules present on repair vesicles and

on the membrane plasma membrane mediate docking and fusion of the patch, sealing the membrane breach and preventing a further influx of Ca^{2+} (Fig.3).

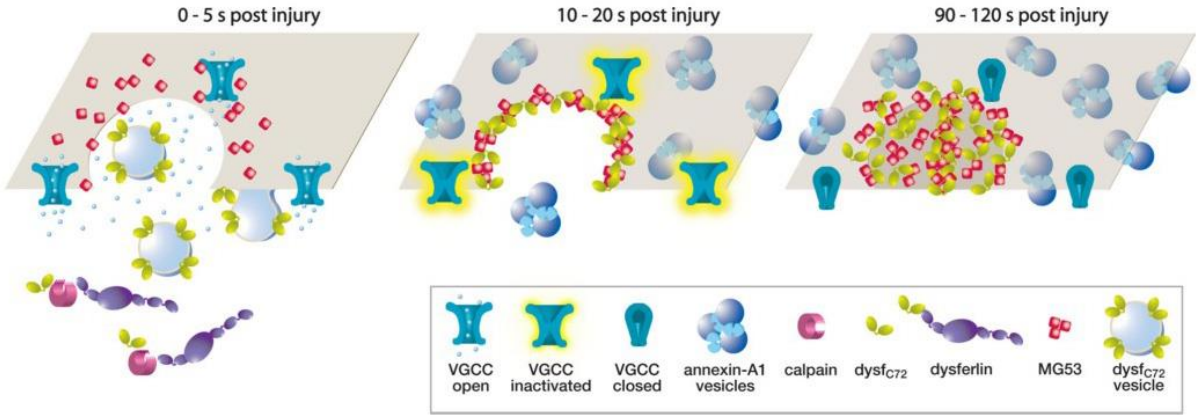


Fig.3: Model of membrane repair after injury. The first five seconds after the membrane injury are characterised by a promiscuous influx of calcium at the sites of membrane injury that leads the local activation of calpains and strong depolarization of L-type VGCCs. In the meantime, MG53 is mobilized and targeted to the injury site in a Ca^{2+}-independent manner. With the activation of calpains, dysferlin is cleaved releasing the mini-dysferlin $_{C27}$. Mini-dysferlin $_{C27}$ are recruited inside the vesicles to the sites of injury and fuse with MG53 in a calcium-dependent process coordinated by L-type VGCCs. Between 2 and 10 s after injury, mini-dysferlin $_{C27}$ fuses into the plasma membrane and undergoes calcium-dependent phospholipid binding via its C2 domains. While between 90 and 120 s mini-dysferlin $_{C27}$ and MG53 compartments infiltrate the plasma membrane surrounding the membrane injury, forming a lattice to stabilize the plasma membrane as it extends to reseal the membrane injury. This process is helped by annexin A1. (Image taken by Lek *et al.*, 2013[149]).

1.4 GENOTYPE-PHENOTYPE CORRELATION

In LGMD2A is difficult to establish genotype-phenotype correlations for the high phenotypic variability (intra- and inter-familial variability) of symptom and severity, even among patients with the same *CAPN3* genotype[31,151]. There is no correlation between the age of onset and the time in which the patient became wheelchair bound. Moreover, the type of mutations in *CAPN3* gene seems to be not a risk factor for becoming dependent on a wheelchair at a determined age. However, it has been shown that patients with two null mutations or homozygous null mutation have a greater risk to have a severe phenotype and to loss of ambulation earlier than patients with at least one missense mutation. Anyway, the precise moment of needing a wheelchair appears to be independent of the type of muatation[151,152]. The amount of calpain 3 protein appears to have a relative prognostic value, then patients carried mutations that lead to degradation of the protein usually are characterised by early onset of the disease. However, this is not always true since absent or markedly reduced protein levels were also found in patients with late or adult onset. On the contrary, almost all patients with normal levels of calpain 3 protein had late or adult onset[152]. This high variability shows in LGMD2A, probably implies the existence of additional factors, that could be genetic and/or environment, that

might regulate the expression of the phenotype, even within the same family. Intra-family variability has been shown since 1998, in which two siblings carried the same two mutations showed two different phenotypes. The brother presented a classical phenotype of LGMD while the sister a metabolic myopathy phenotype[153]. Another case was observed in two siblings with the same mutations but with 10-year difference in becoming wheelchair bound[151]. Moreover, the clinical-molecular correlation is complicated even for the presence of double heterozygosity. In fact, two different mutations have a different functional impact on the activity of the protein.

In conclusion, for LGMD2A we can try to study a genotype-phenotype correlation only with mutations that lead to the absence of the protein, that in almost all cases are null mutations. Patients with two null mutations had a tendency toward slightly earlier onset than patients with two missense mutations. While no define correlation can be drawn from two missense mutant alleles, either homozygote or compound heterozygote characterised by a variable clinical features[151,152,154].

Even LGMD2B is characterised by high variability, just considering the fact that different mutations in the same gene give rise different phenotypes and even the same mutations may lead to a phenotypically different muscular dystrophy[155,156]. Moreover, the same mutation is even associated with a wide inter- and intra-familiar variation in clinical phenotype. Then, the type of mutation does not correlate with phenotypic severity. All attempts to correlate phenotypes with *DYSF* gene mutations have been unsuccessful. It is still not found any genotype-phenotype correlation either with regard to the nature of the mutation and severity or with the phenotype and age at presentation[157,158].

1.5 DIAGNOSTIC APPROACH IN LGMD

Considering the wide clinical and genetic variability of LGMDs, achieving a precise diagnosis might be difficult and requires access to multiple specialities. Clinical examination, laboratory testing, muscle imaging and histological appearances of muscle are required[9,11]. In addition to the clinical and genetic variability, there is even an overlap with a number of other diseases including myofibrillar myopathy, pompe disease and congenital muscular dystrophy.

For a certain diagnosis is required first a clinical assessment for directing further investigations. The medical evaluation includes family history, physical examination, serum CK levels and muscle MRI, CT scan and electromyogram (EMG) are used to determine distinct patterns of muscle involvement[159]. Medical evaluation is followed by muscle biopsy analysis, even if is not crucial for all forms of LGMD. For example, deficiency of CAPN3 is not evaluated with immunohistochemistry or in some cases, protein deficiency documented by immunohistochemistry in muscle may be a

secondary deficiency due to mutations in other genes. However, muscle biopsy still represents an important step in the diagnostic process in order to investigate the presence or absence of specific muscle proteins (such dystrophin, dysferlin, sarcoglycans and glycosylated α-dystroglican). All these exams just described lead in many cases to targeted genetic testing. In fact, diagnosis is confirmed by detection of pathogenic variants in the causative gene of the disease[160,161]. However, even if direct sequencing of the coding regions of the genes responsible for the different form of LGMDs is the most common method, it remains often insufficient due to the size and numbers of genes involved. Today with next generation technique is possible to sequencing genome or exome, and there are even commercially kits which include loci for the diagnosis of neuromuscular disorders. This is a big step forward for the diagnosis of LGMDs, thanks to that is possible to analyse a big number of genes at a reasonable cost without invasive muscle biopsies[26,162]. Unfortunately, there is a limitation that is related to insufficient exon coverage, leading to a potential lack of representative examination of all exons[163]. However, for the approach to heterogeneous hereditary conditions with huge phenotypic variability and large size of many associated genes such as muscular dystrophies, NGS techniques represent a powerful tool to identify genetic variations[164]. So far, exome and genome sequencing strategies are not the first test used for skeletal muscle disorder but thanks to the technological improvements and the cost decreased will soon replace the single gene testing.

1.5.1 Diagnostic approach in LGMD2A

Diagnosis of LGMD2A is challenging for the high phenotypic variability, lack of precise protein analysis in muscle biopsies and absence of mutational hot spots in the *CAPN3* gene. Clinical evaluation and the clinical tests help to direct the further analysis. One of the clinical tests is the examination of serum creatine kinase (CK) concentration that is always elevated (5-80 times normal) from early infancy on, particularly during the active stage of the disease. Serum CK concentration decreases with disease progression, as muscles become more and more atrophic[165]. Another clinical test is the muscle imaging with CT-scan and EMG that is used to discover the pattern of muscle involved. After the initial clinical work up and exclusion of the more common conditions with overlapping phenotypes, laboratory tests are performed. Unlike the other types of LGMDs, in LGMD2A is not possible to perform immunohistochemistry analysis for diagnosis. Anyway, the muscle biopsy of most individuals shows the typical features of an active dystrophic process (increased fiber size variability, increased fibrosis, regenerating fibers, degenerating and necrotic fibers), others have mild and non-specific myopathic features[166–168]. For evaluated the expression level of the protein western blot is performed. However, western blot results need to be interpreted with caution for the lack of specificity, since calpain 3 protein levels can be partially reduced in other

muscular dystrophies such as LGMD2B/2I/2J[140,151,169]. Moreover, calpain 3 is highly susceptible to degradation. Then, it is important to avoid technical artefacts that could be introduced when the biopsy sample is processing and might lead to protein degradation[170]. For example is important to exclude Na^+ from any buffer since CAPN3 rapidly autodegrades in solutions containing Na^+. Anyway, loss of all calpain 3 bands means with a high probability mutations in *CAPN3* gene but it is not a certanty[152]. It has been shown cases in which absence of CAPN3 protein in western blot was not linked to mutations in CAPN3 gene[151]. Moreover, it has been described patients with two mutations that showed normal full-sized calpain 3 protein[171]. Therefore, western blot analysis is not a diagnostic test but in some cases help to obtain a complete diagnosis.

Genetic analysis is necessary to confirm or not the clinical suspect, but unfortunately, the lack of defined mutational hot spots makes the analysis of the 24 exons and the flanking intronic sequences laborious and expensive. Moreover, in ~23% of patients believed to have LGMD2A only one mutation can be detected[151,152]. In these cases is essential even cDNA analysis in order to investigate the presence of a deep intronic variant. So far, the absence of the second mutation is hard to explain. Last year, Duno's research group speculated about the possibility of an in frame deletion of 21bp in heterozygous state could behave as dominant mutation, but not all scientific community agree with him[172,173]. This explanation is not sharing in the scientific community even because it has been shown that heterozygous deletion of one entire allele of *CAPN3* doesn't lead to LGMD2A symptoms, indicating that half-normal levels of CAPN3 are sufficient to maintain skeletal muscle[174].

1.5.2 Diagnostic approach in LGMD2B

Diagnosis of LGMD2B is challenging as well. This is for the high phenotypic variability, the absence of mutational hot spots in *DYSF* gene and for the large size of the gene. Mutation screening of *DYSF* gene is labour intense, expensive and time-consuming.

Like LGMD2A, the first step is the clinical evaluation and clinical tests. After that, differently to the LGMD2A, it is possible to perform immunohistochemistry in order to evaluate the presence of dysferlin in the muscle. Immunostaining could reveal absent or decreased levels of dysferlin at the membrane with an increased cytoplasmic staining. However, the reduced sarcolemmal immunostaining can be secondary and seen in other types of LGMD[175]. Then, western blot needs to be performed on the muscle or white blood cells to confirm a primary deficiency. It is possible to perform western blot in both tissues since the expression of dysferlin in peripheral blood monocytes correlates with dysferlin expression levels in skeletal muscle in both healthy and dysferlinopathy patients[176,177]. Western blot in blood sample represents a noninvasive method of confirming the

diagnosis. Moreover, like in LGMD2A, muscle biopsy shows the typical features of an active dystrophic process (variation in fiber size, necrotic events and regenerating fibers). There is often evidence of inflammation, sometimes leading to a misdiagnosis of polymyositis[38,178]. If these tests show an absence of dysferlin protein, in the majority of cases indicate the presence of mutations in *DYSF* gene. However, in order to confirm that, molecular and genetic studies are needed to obtain a definite diagnosis.

2. AIM OF THE PROJECT

Limb girdle muscular dystrophies are a clinically and genetically heterogeneous group of disorders. Then, diagnosis in most of the cases can be really challenging.

The aim of this project was to obtain a certain diagnosis for 90 patients with LGMD2 clinical features. We concentrated on the two most frequent forms of limb girdle muscular dystrophies with autosomal recessive inheritance. The gold standard for LGMD diagnosis is the study of proteins involved either by immunoblot or immunohistochemistry based on clinical phenotypes, followed by a mutation study in candidate genes. For that reason, it was set up genetic screening for *CAPN3* and *DYSF* gene, immunoblot and immunohistochemistry analysis. Both genetic and biochemical analysis resulted essentially for obtaining a diagnosis in patients in which only one mutation has been identified.

In addition to obtaining a certain diagnosis, this project was carried out for increase the mutational spectrum in LGMD2A and LGMD2B and to contribute to the phenotype and genotype correlation studies.

3. MATERIAL AND METHODS

3.1 ETHICS

All patients in this study provided written informed consent for muscle, skin biopsy and genetic analyses. All the procedure are in accordance with the Helsinki Declaration of 1975.

3.2 CELL CULTURE MAINTENANCE

Skin biopsies from patients and healthy controls were washed twice and broken into small pieces under sterile condition for primary fibroblasts isolation. After 15' incubation with trypsin at 37°C, tissues were placed in 25ml plastic flasks with 1 ml of DMEM (D5671, Sigma) supplemented with 10% fetal bovine serum (D8537, Sigma), 1% GlutaMAX^{TM} Supplement (35050061, Thermo Fisher Scientific) and 1% Penicillin-Streptomycin (P4458, Sigma). The flasks were incubated vertically overnight at 37° C with 5% CO_2 and after that were returned to horizontal position. Then 5ml of DMEM 10% was added to plastic flasks and cells were grown in standard conditions.

3.3 PRIMER DESIGN

Before primers design, repeated sequences have been identified using RepeatMasker Web Server (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker), while SNPs, small deletions and duplication using 1000 Genomes Project (http://grch37.ensembl.org/index.html). PCR primers were designed using Primer 3 (http://primer3.ut.ee/) with the following conditions:

- Melting temperature between 56°C and 63°C
- Less than 2°C difference in melting temperature between the two primers
- GC content between 40-60%
- Length between 18-22bp
- Avoiding secondary structure or sequence complementary at 3'ends

Finally, the specificity of primer pairs was tested with UCSC In Silico PCR (http://rohsdb.cmb.usc.edu/GBshape/cgi-bin/hgPcr). Nucleotide sequences are listed in Table 3.

A

Specific primers for *DYSF* gene			Specific primers for *DYSF* gene		
Exon	Fw	Rv	Exon	Fw	Rv
1	CAGAAGCCCCTGTTCTCGG	AGAAGCGCCATTCAGACACC	29	GCTATTTCACCCAGGGACGC	GAGTGTGATCTGTGGGTGTTCC
2	TGTAAGATTTCCCTTGGCGACC	TGCAGCAGACAGAGAGAAAAGG	30	ACTAGTGTGGGAGCTGGTGG	CAGACTGGGGCTGCTTTCC
3	AGAGGGGCCTGTGAAAGACC	GCTAATTGAAACCCCAGCTCCC	31-32	TCTAACTCTCTGGGCTAGTCTCC	CCCGAAAGCCAGATGTCTCC
4	CCAGGGATGGTGCATGGTGG	GTGCCCCTCCAGACACTGC	33	TCTTGTGCCCATTTGCCAGG	CTAGGGCCCGTGGAGTTAGG
5	TGCAGGTGAGGTTGGATCCC	CTGGGGAGGGCAGCAAAGG	34	AGTTCAGAAGGCACATGTAGACC	TCAGAAAATGGGTCTCAGCAGG
6	CTGGCTCACTGATGGGGAGG	AGGATCCCACACACATCTTCCC	35	GCTGCAAACCATGGACTGTC	TCCTAACCCAGAGAGATGCCA
7	TCTGACTCTTGGGGTGGATGG	GAGCAGCAGGAGAAGTAAGAGC	36	CCCTTCCAACCCCTCTCACC	AGGTCTCCATGTGACAGATGCC
8	CCCAGAACTGTGCCCAATCC	TCAGGCTTTAACACGTCTCCC	37-38	TCTGTTTCTCCCTGCACTTGG	CGGGGACATGTAGAACTTCTAGC
9	CTGGGCCTTGGCTCTTCC	TCCCAAAGCAGAGAAACACAGG	39	TGGCTGTGGGGATTATCTGC	CTAGGAAGGGTAGCTTGGCC
10	TTGACTGCCTGTGTTTCC	GACAGTGAGGACCCCAAA	40	AGAGCATGTGAAGGAGGATGGG	CTTGCCCAGATGAGCACACC
11	TCTGAGTCCTGAGCTCATGG	ACACCTTTCTACATGTTCCACA	41	GGTTTCCTCTCTCTGCTCCCC	GGTCTAGCAGCATTCATCACCC
12	GCAGCATTTATGTGGCGAAG	TGTAACTCCACCAGCACTGA	42	TTTCCCCTCCTTCTCTTGTTTCC	TGGTCACCAACATCAGCACC
13	TCTGAAAGGCTTTGAATGCAGGG	TCTCACTCTGTCTTGCTCTCACC	43	CAAGTTGAGATTTGGGTGGGG	AGGCAGAGTTTGTGTCCAGG
14-15	AAGATTCCCTGTGTATGAACCCC	CAACAAGCAGGTGTGAGTAAGC	44	GGACACAGCCCACATCTCA	TGCCAGTCACATCTCCCC
16	CCAAAGTGAAGACCAGAAGCGC	TAGAGCACAGTCGGCCAAGG	45	GCCCAAGGAAAGAAGACTCCC	GTGGTGTGTTGTCAGAGAAGC
17	GGCTCATGGGGTGTTGGGG	TCCTCTTGGCCTCCTCCTGG	46	AAGCAAGGAGTGGGCAATGC	GCTGCTGTGGGTTATGTCCG
18-19	GGTGGGCTCAGGTTGGAAGG	GCGTGTGGGCAGAGCTTCC	47	GCCAAATCTAGGAGTAAAGGTTGC	GTGAAGAAATGGGTCAAGGAAGG
20	GCCTGGGTGCCTTTCTTTGC	CCAGGAGCTGCATCTGTGGG	48	CCCTCCATAACTGAACGGTGC	CAGGCCCAAACGAGAACTGT
21	GCACTCTTAGCACCCCATCC	ATCCCCTCTTTTCTCTTCCAAGC	49-50	CCAGGCAGGCATCGAGTCC	CGTGGTGGTCAGGTCAGGC
22	TAGCAGTGACAAGGCCTGG	CCAGTTGAGCCAGGAGAGG	51	TCCTACTCCTCTGCCTCCTCC	ACACATGACCCTGAGCTCCC
23	GGGCCTGCTGTGAGAACC	GGAACCTGGGGAAGCTGG	52	AGACAATGGGTAGGGAGGTGG	CGGGAATTGGGAGGAAAGAGG
24	CCTTATGCTCTGTTCTCTGGGG	TGGGAGAAAGGCCTTATGTTGG	53	CAGGTGCTGCTTTTATGTTCAGA	CGTGGGCTTTGTGACTTCTACA
25-26	GGTGTCAGCCTGCTCTACCC	TGGAAAGGACTGCTGCCTGG	54	GGCTAGTTTAGGGACAGGGCC	ATCCTCATCAGACCAGGGGC
27-28	CATTGTTGGTTGGGGTGAGGC	GAGAGCTTGCCGGGGATGG	55	GAGTGTTGAGCCCCACATGG	AGAGATACACACACTGACACTGC

B

Specific primers for *CAPN3* gene		
Exon	Fw	Rv
1	ACTCCAACTTCCCCTTTGCAGTTGCT	GAGGTGCTGAGTGAGAGGACTCCT
2	CTATAGTCCCAGCTACTCGAGAGGCT	GTAGGAGGAAATGAAGCAATCGCCAGC
3	CCTTCCCTGGGTTGACACCTCTGT	TGCTTCCTCCCAGGCTTGAAGCCT
4	ACAGGCACCCAGATGCAGAGTCCA	GAGCTGTTGTTGCCTGGAACTGATCTC
5	TGTTTCCATCCCATGAGCTCATAGGG	CCAGTCCTCAAAGTTGAGAGCCAG
6	GTATCACTTTTCTCCGTCTTTCCTCC	GCCTTGGCTGAGCAAGTTATTTCAAC
7	TCTGATGGTCAGGACAGAGCCTTC	CCTTGAACTGAATGTCCTGATTGGAG
8	CTGCCATTCCCCAGCACACTTGTG	TGTCAGCAACCTCTCTGGGTCCC
9	GGATGTGAGAAGTCATTGCTATTGGG	AAGGCTCCTCTCAGAACCTCATCC
10	GGGACCCTGACCAAGTTCCTGTGA	CAGCTCCCTATTGCAAACTTCTCAGC
11_12	CCTAAGCACCTAGATGTAGGGAATAG	ACCAGGAAGCTTCTCGACGAGC
13	GTAGGGAGGCTATTTAAGCCTTGGG	TAATGTGGCCAGTGGCTCAGGAG
14	GGTAGGTGGCTGGGTCATGCCTTT	CCCCGGTGCTTAAGGCCGATTACT
15	GGTGTCTGGGCATGGAGGAGAAAG	TCTCATAGGGCTGTTGTGAGGAAGGA
16	GCTTGTCAGTGGCAGAGATAGAGCTT	CACCCAAGCAGCAGCCCAAATTTC
17	ATCTCACTCCAGCTCACCAGGTCT	GAAGCTGGGGCTTTTCTGAGGAAG
18_19	TTTTTGCTCTGAGCTGCCCACGGT	CAGGGTCTCCTTAAGTTTCCCTGG
20_21	GTGGAGTGGAGGGGAGGGTTAAATA	AGGCCTTGGGCCTGCCTTCTATTT
22_23	AGAGAAAGGAGAGGGAAAGGGCTTCT	CGTTGACCCCTCCACCTTGAGTTTT
23_24	TAACTGGCCTCTGGCCTGTGCATT	TAGGCATGACCGATCAAGGATGGGTA

Tab.3: Panel A: specific primers designed for amplified all coding exons and flanking regions of *DYSF* gene. Panel B: specific primers designed for amplified all coding exons and flanking regions of *CAPN3* gene.

3.4 DNA ANALYSIS

Total genomic DNA was extracted from peripheral blood samples with MasterPure Complete DNA Purification Kit (MCD85201, Epicentre), according to manufacturer's instructions. The quantity and quality controls were carried out with the spectrophotometer NanoDrop-1000 (Thermo Scientific). The exons and adjacent intron regions of *CAPN3* gene (NC_000015.10) were amplified in a pool of 20 different PCR. While exons and adjacent intron regions of *DYSF* gene (NC_000002.12) were amplified by 55 PCR with specific primers.

PCR reactions were performed in 50 µl reaction volumes containing 100 ng of template DNA, 0.2 pmol forward and reverse primers, 0.4 mM deoxynucleotide triphosphate (dNTP), 1.5 mM $MgCl_2$, 1X reaction buffer, and 1.25 U Red Hot DNA Polymerase (AB0406B, Thermo Fisher Scientific), using a DNA thermal cycler (PTC-200; MJ Research). The products of individual PCR reactions were separated electrophoretically on 1.5% agarose gels (16500500, Thermo Fisher Scientific) and revealed with ethidium bromide staining (E1510, Sigma). Polymerase chain reaction products were sequenced by automated sequencer ABI 3730 (Applied Biosystems). Obtained results were analysed using Sequence Scanner software (Technelysium) and compared with human *CAPN3* gene sequence (NG_008660.1) or human *DYSF* gene sequence (NG_008694.1).

Novel variants have been identified, a new extraction and amplification were done in order to be sure that no contamination events have occurred.

3.4.1 Real time qPCR

Real time qPCR was performed in LGMD2B patients to confirm the homozygosity of mutations and exclude the presence of large deletions. Six assays were designed: exon 19, exon 27, exon 37, exon 42, exon 51 and exon 54 of the *DYSF* gene. Before carried on the experiments, validation of each assay was performed in order to check the reaction efficiency. For validation, DNA template from a healthy control was used and the standard curve prepared with four serial dilutions. The construction of a standard curve enabling the determination of the efficiency, linear dynamic range and reproducibility of a qPCR assay. Selected assays were those with 90-105% range of efficiency, the R^2 value close to one and similar Cq values in the replicates. The qPCR reactions have been performed with CFX Connect™ Real-Time PCR Detection System using SsoAdvanced ™ SYBR® Green Supermix (1725271, Bio-Rad) according to the manufacturing instructions. Nucleotide sequences are listed in Table 4.

qPCR primers (*DYSF*)		
Exon	Fw	Rv
19	CACAGTGAACAGAAGGTGGAGGAC	GCGTGTGGGCAGAGCTTCC
27	CCGGTCACCTGAGCTTCGTGG	ACCCCACACTTTCCCTTTGG
37	AGTTACCAGCTGGCCAACATCTC	GAAGAGGGTGCAGATGTCAAAGTT
42	TGTCTGACTTTTGTAACACCTTCAA	TGGTCACCAACATCAGCACC
51	CTGGACAAGACTGAGAGCAAAATC	AAAATCATCAAAGGAGAACTTGTCA
54	CCCGACACCTCCTTCCTGTGG	ATCCTCATCAGACCAGGGGC

Tab.4: Primers designed for quantitative PCR on *DYSF* gene.

Three reference genes (*Sel1L*, *RBM11* and *ZNF80*) were evaluated as endogenous controls and *ZNF80* was selected for data normalization, as it showed the most constant expression and stability

in all the tested samples. The reactions were performed in 15µl containing 20ng of DNA. In the present study, data normalization was performed with the Bio-Rad CFX Manager 3.1 Software by using both the Normalized Expression Method (ΔΔCt). Data shown in the present study represent the mean of three different experiments, calculated by the Gene Study tool of the CFX Manager Software.

3.4.2 Restriction fragment length polymorphism analysis

Restriction fragment length polymorphism (RFLP) analysis was used to exclude the possibility that new mutations identified were small nuclear polymorphisms normally present within the population. RFLP was carried on in 320 chromosomes of controls. For the identification of restriction enzymes sites that cleave the region of our interest NEBcutter V2.0 (http://nc2.neb.com/NEBcutter2/) was used.

For c.526G>A (p.Val176Met) in *CAPN3* gene, the reaction was performed in 40 µl reaction volumes containing 5U of NlaIII restriction enzyme (R00125S, New England BioLabs®Inc), 1X reaction buffer and 5 µl of PCR product. The samples were incubated at 37°C overnight. The enzyme cut the wild-type sequence two times generating three fragments of 355bp, 55bp, 30bp. While in the mutant allele, the enzyme cut three times generating four fragments of 195bp, 160bp, 55bp and 30bp. The digested samples were examined by electrophoresis through 2% agarose gel. In order to recognize species-specific banding patterns, the yielded restriction profiles were compared with molecular weight marker VIII (11209264001, Sigma).

For c.2458T>C mutation (p.Tyr820His) in *CAPN3* gene, no restriction enzymes were identified with NEBcutter V2.0, then modified oligonucleotide was designed in order to generate a restriction site for RsaI enzyme (R0167S, New England BioLabs®Inc) within the wild-type sequence. The reaction was performed in 40 µl reaction volumes containing 5U of RsaI restriction enzyme, 1X reaction buffer and 5 µl of PCR product. The samples were incubated at 37°C for 1h. The digested samples were examined by electrophoresis through 2% agarose gel. The enzyme recognized only one restricted site in the wild-type allele generating two fragments of 140bp and 45bp. In order to recognize species-specific banding patterns, the yielded restriction profiles were compared with Gene Ruler™ 50bp DNA Ladder (SM0373, Thermo Fisher Scientific).

3.4.3 *In silico* analysis

The possible effect of the variants identified was assessed with software tools: Mutation Taster (http://www.mutationtaster.org/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), PROVEAN

(http://provean.jcvi.org/index.php), snSNPAnalyzer (http://snpanalyzer.uthsc.edu/), SIFT (http://sift.jcvi.org/), MutPred (http://mutpred.mutdb.org/) and PMut (http://mmb.pcb.ub.es/PMut/).

3.4.5 Conservative analysis

Conservative analysis was performed with multiple alignment of amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of *Macaca mulatta* (XP_001103220.1), *Mus musculus* (NP_031627.2), *Rattus norvegicus* (NP_058813.1), *Bos taurus* (NP_776685.1), *Alligator mississippiensis* (XP_006258549.1), *Danio rerio* (NP_001004571.1), *Gallus gallus* (NP_001004405.2) and *Xenopus leavis* (NP_001087053.1) using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

While conservative analysis was performed with multiple alignment of amino acid sequence of human DYSF (NP_003485.1) with the orthologous proteins of *Macaca mulatta* (XP_014968118.1), *Mus musculus* (XP_006506233.1), *Rattus norvegicus* (XP_017448122.1), *Bos taurus* (XP_015328883.1), *Alligator mississippiensis* (XP_019351176.1), *Danio rerio* (XP_005172350.1), *Gallus gallus* (XP_015141556.1) and *Xenopus tropicalis* (XP_017946387.1) using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

3.4.6 3D modeling

CAPN3 3D structure was predicted by utilizing Automated Mode of SWISS-MODEL software (http://swissmodel.expasy.org/). The 3D structure was generated with the alignment of human m-calpain (NP_001739.2, 1kfuL) and it was displayed by PyMol (PyMOL v.1.7.0.1; Python v.2.7.2). Only regions with perfect alignment were analysed in order to obtain a good reliability of the prediction.

3.5 RNA EXTRACTION AND RETROTRANSCRIPTION

Total RNA, isolated from fibroblast or form PBLs, was processed immediately using RNeasy Mini Kit (74104, QIAGEN), following the manufacturer's instruction. RNA was treated with DNAse I for 15 minutes at RT (79254, Qiagen). The concentration and purity of total RNA samples were quantified using the spectrophotometer NanoDrop-1000 (Thermo Scientific).

For each sample, 1 µg of total RNA was reverse transcribed using ImProm-II ™ Reverse Transcriptase (A3800, Promega) and random oligonucleotides in a 20µl volume. cDNA was used for both qualitative and quantitative analysis.

3.5.1 Qualitative analysis

cDNA was amplified into five overlapping regions in order to analyse the whole coding region of the *CAPN3* gene with primers designed by Blazquez *et al.*, 2008[60]. Then, cDNA was compared with the *CAPN3* human isoform 1 (NM_000070.2). While the whole coding region of *DYSF* was amplified into thirteen overlapping regions with primers shown in table 5 and cDNA was compared with the *DYSF* human isoform 8 (NM_003494.3). The products of individual PCR reactions were separated electrophoretically on agarose gels or polyacrylamide gels and revealed with ethidium bromide staining.

Specific primers DYSF cDNA

Exon	Fw	Rv
1	CTCTAAGCCAGGAGCCAGAGATT	CATTGAAGCTGGCGGACAGACTA
2	GTGGTCAAAGACCATGAGACGAT	AGTCTCATTGAAGAGTGGGCTGTTT
3	ACATCTAGAAAGCTGCTGTCAGAC	AAAGATCTGTTTCACGTTGTCCAT
4	AGCCCACTTCTGCCTGAAGGTCT	AGTACTTGCGCCTCCTAAGGTAC
5	CACAGTGAACAGAAGGTGGAGGAC	CTCAGTGATTTGGCTCAGGTGATG
6	GTGACATCCATGAGACACCCTCTG	AGAGTCTTCTCCGGACACACGAAC
7	CTCACCTACCCCAAGTTTTCTGAC	GAAGGCATCTGTCTTGCGGTACTCG
8	AGCCAAATGGAAGCACTGAAAAG	ATAAACTCGTCTGCACCATAAGTGT
9	GTGGTGGTGAAGAACACCCTTAAC	GAAGAGGGTGCAGATGTCAAAGTT
10	AGTTACCAGCTGGCCAACATCTC	GTCTTCTGGGAGGGGATAAATTTTG
11	TGTCTGACTTTTGTAACACCTTCAA	TGCCAGCCTCTATCTCTTCAATG
12	ATAGAGTCAAGGCACCTGTGTACC	AAAATCATCAAAGGAGAACTTGTCA
13	CTGGACAAGACTGAGAGCAAAATC	CAATCAGGCCTAGGAGGTCTGGAG

Tab.5: Specific primers designed for *DYSF* cDNA.

3.5.2 RT-qPCR

The expression levels of mRNA in each sample were investigated by a quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) using SsoAdvanced ™ SYBR® Green Supermix with CFX Connect™ Real-Time PCR Detection System. Each experiment was performed in three independent experiments with triplicate. The reaction specificity was confirmed by single peak in melt curve analysis and relative RNA expression was performed according to the $2^{-\Delta\Delta Ct}$ method, with internal control genes.

3.5.2.1 Dysferlin RT-qPCR

The expression levels of dysferlin transcript were evaluated in blood samples and 5ng of cDNA was used for RT-qPCR. Three assays were designed in order to cover the middle and the end of the transcript: between exon 42 and 43, exon 54 and the last assay should be around exon 34. The assay around econ 34 was designed and validated by Bio-Rad (qHsaCID0012070) and then the exact localization is not known. Before carried on the experiments, validation of the first two assays was performed in order to check the reaction efficiency as done for qPCR. Experiments have been performed with CFX Connect™ Real-Time PCR Detection System using SsoAdvanced ™ SYBR®

Green Supermix according to the manufacturing instructions. Nucleotide sequences are listed in Table 6.

RT-qPCR primers (*DYSF*)		
Exon	Fw	Rv
~34	qHsaCID0012070, Bio-Rad	
42-43	TGTCTGACTTTTGTAACACCTTCAA	GTCTTCTGGGAGGGGATAAATTTTG
54	CCCGACACCTCCTTCCTGTGG	ATGCTGGAGGGGACCCCACGG

Tab.6: Primers designed for RT-qPCR to evaluated the expression levels of dysferlin transcript.

The experiments were carried out with cDNA from a healthy control that was used as calibrator. While hypoxanthine phosphoribosyltransferase 1 (HPRT1, qHsaCID0016375) and TATA box binding protein (TBP)-associated factor (TAF15, qHsaCID0021632) were used as endogenous controls to normalize each sample. Reference genes were provided and validated by Bio-Rad.

3.5.2.2 Calpain 3 RT-qPCR

The expression levels of calpain 3 transcript were evaluated in skin samples and 5ng of cDNA was used for RT-qPCR. Three assays were designed and before carried on the experiments, validation of them were performed in order to check the reaction efficiency as done for qPCR. Experiments have been performed with CFX Connect™ Real-Time PCR Detection System using SsoAdvanced ™ SYBR® Green Supermix according to the manufacturing instructions. Nucleotide sequences are listed in Table 7.

RT-qPCR primers (*CAPN3*)		
Exon	Fw	Rv
Exon 5/6	GATGATGGCACGAACATGAC	TCTGAGCCTCTGGGGTC
Exon 9/10	GTGGACAAAGATGAGAAGGCC	TGAGGTTGCAGATCTCCAACT
Exon 10/11	GAACAGCAACAATTCCGGAAC	GAACCCGTGTGTCTTCAGGT

Tab.7: Primers designed for RT-qPCR to evaluate the expression levels of calpain 3 transcript.

The experiments were carried out with cDNA from a healthy control that was used as calibrator. While hypoxanthine phosphoribosyltransferase 1 (HPRT1, qHsaCID0016375) and beta glucuronidase (GUSB, qHsaCID0011706) were used as endogenous controls to normalize each sample from skin samples. Reference genes were provided and validated by Bio-Rad.

3.6 PROTEIN EXTRACTION

Muscle samples were lysed on ice with Tissue Protein Extraction Reagent T-PER™ (78510, Thermo Fisher Scientific) and Complete Mini Anti-protease Cocktail Tablets (11836153001, Roche) according to the manufacturers' instructions. Fibroblast and lymphocyte samples were lysed with Mammalian Protein Extraction Reagent M-PER™ (78501, Thermo Fisher Scientific) and Complete Mini Anti-protease Cocktail Tablets according to the manufacturers' instructions. The protein concentration was determined by Lowry method.

3.6.1 Western blot analysis

Calpain 3 and dysferlin expression were evaluated by western blot analysis. 90µg protein from muscle samples, 40µg protein for skin samples and 50 µg protein from lymphocyte samples were used. Samples were separated by 7% SDS-polyacrylamide gel electrophoresis. Gels were run at 150 V for 1 hour and another hour at 120 V. Gels were electrophoretically transferred to nitrocellulose membrane (1620115, Bio-Rad) for 1h at 350mA. Membranes were then blocked with 1.5%BSA (A7030, Sigma) for CAPN3, 3% BSA for DYSF and 5% BSA for GAPDH in TBS/Tween 0.05% (P1379, Sigma) for 1 hour at RT followed by incubation overnight at 4°C of primary antibodies.

- CAPN3 antibody (1:25) (NCL-CALP-12A2, Leica Biosystems).
- DYSF antibody (1:200) (JAI-1-49-3, Abcam).

The following day, membranes were washed four times for 5 minutes with PBS-T and then incubated with secondary antibodies for 1 hour at RT.

- Anti-rabbit secondary antibody, HRP (1:15000) (HAF008, R&D Systems)
- Anti-mouse secondary antibody, HRP (1:15000) (HAF018, R&D Systems)

In each experimental session, a patient and a control were analysed in parallel gel/blots and GAPDH (ZG003, Thermo Fisher Scientific) was used as internal reference to normalize protein expression. The amount of proteins were detected by Clarity ™ Western ECL Substrate (1705060, Bio-Rad) according to the manufacturer's instructions. The relative amounts of CAPN3 and DYSF protein were analysed by densitometry using ImageJ software.

3.7 IMMUNOSTAINING

Dysferlin immunofluorescence was performed in muscle sections from patients and controls. The muscle sections of 10µm were fixed in 4% paraformaldehyde (15710, Electron Microscopy Sciences) in PBS 1X at room temperature (RT) for 15 minutes. In fixed cells were added ice cold methanol for

5 minutes at -20°C. Then cells were washed three times with PBS 1X for 5 minutes each and blocked in 5% horse serum (AB7484, Abcam) diluted in PBS 1X for 1 hour at RT. After blocking, cells were incubated with RabMab Rabbit monoclonal (1:100) (JAI-1-49-3, Abcam) diluted in blocking solution overnight at 4°C. The day after the cells were washed with PBS 1X three times for 5 minutes each and incubated with Alexa 594 goat anti-rabbit (1:200) (R31632, Thermo Fisher Scientific) secondary antibody for 1 hour at RT. After that, three washes for 5 minutes each with PBS 1X followed by Hoechst (33342, ImmunoChemistry) staining for nuclei.

The cells were examined by fluorescence microscopy (ZEISS AxioCam MRc5).

4. RESULTS

4.2 DYSF PATIENTS

4.2.1 D1_Clinical features

The first patient (D1), a 55-year-old man presented weakness of the pelvic and shoulder girdle. Muscle biopsy revealed the presence of dystrophic features with increased variation in fiber size, substitution of muscle with fibro-adipose tissue, increased number of internal myonuclei and fiber necrosis.

4.2.1.1 D1_Genetic analysis

Direct sequencing of *DYSF* gene showed the already reported missense mutation c.6124C>T (p.Arg2042Cys) in exon 54 in homozygous state (Fig.4). Presumed homozygosity could be formally confirmed by sequence analysis of parental DNA, unfortunately, the parents of this patient were not available for sequence analysis. Therefore, the homozygous state or the possible large deletions in the other allele were detected by real-time PCR assays (Fig.6).

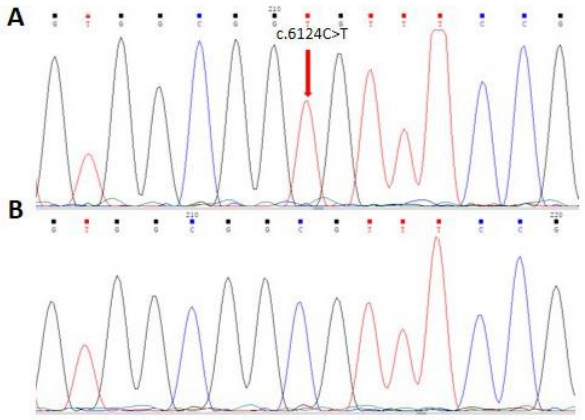


Fig.4: Panel A: Electropherogram shows c.6124C>T variant in the homozygous state in exon 54. Panel B: Electropherogram of a negative control.

4.2.1.2 D1_Conservative and *in silico* analysis

Multiple alignments of the amino acid sequence of human DYSF (NP_003485.1) with the orthologous proteins of nine different species have shown that arginine in exon 54 (p.Arg2042) is highly conserved among all species (Fig.5A).

In silico analysis shows the pathogenic nature of p.Arg2042Cys (Fig.5B).

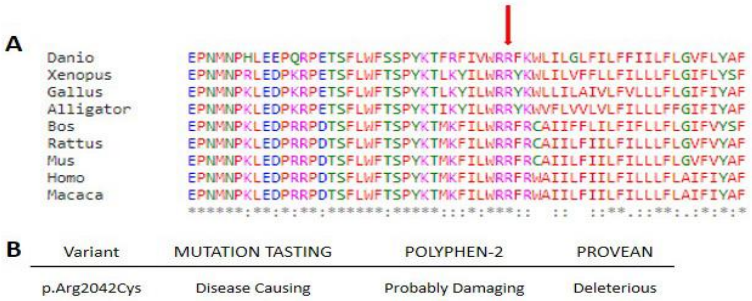


Fig.5: Panel A: Multiple alignment of amino acid sequence of human DYSF (NP_003485.1) with the orthologous proteins of *Macaca mulatta* (XP_014968118.1), *Mus musculus* (XP_006506233.1), *Rattus norvegicus* (XP_017448122.1), *Bos taurus* (XP_015328883.1), *Alligator mississippiensis* (XP_019351176.1), *Danio rerio* (XP_005172350.1), *Gallus gallus* (XP_015141556.1) and *Xenopus leavis* (XP_017946387.1). p.Arg2042 is highlighted with the arrow. Panel B: *In silico* analysis results for the amino acid substitution p.Arg2042Cys.

4.2.1.3 D1_qPCR and qRT-PCR analysis

qPCR in D1 patient was performed since relatives were not available for the genetic testing, moreover D1 showed all SNPs in homozygous state in *dysferlin* gene. It could be possible that the second allele has not been analysed with the only gene screening. Then, the homozygosity of the variant c.6124C>T was verified with qPCR in genomic DNA of the proband and a healthy control using five assays. No large deletions were detected and homozygosity of the variant was confirmed (Fig.6).

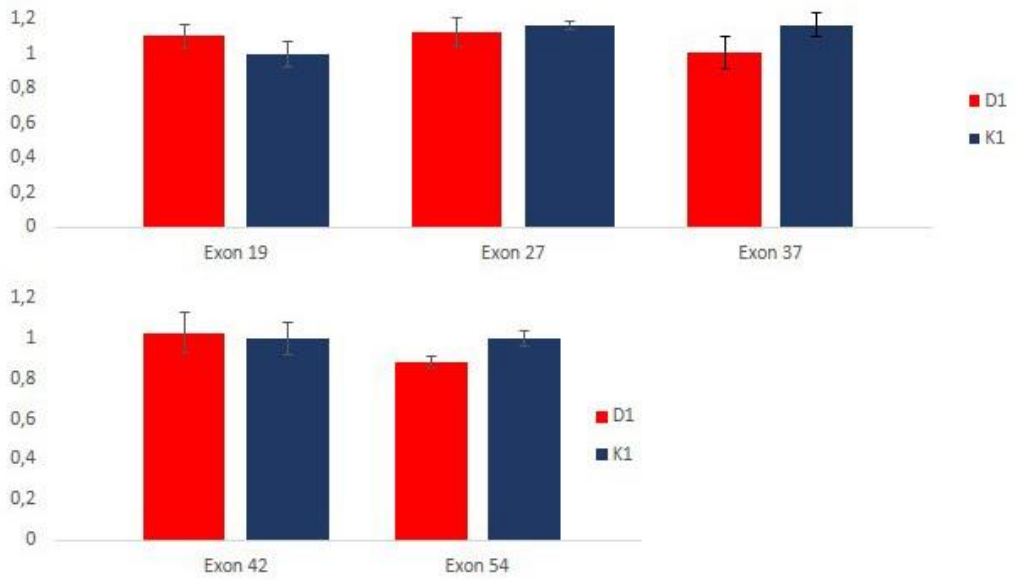


Fig.6: Results of qPCR show no large deletion in *DYSF* gene in D1 patient (shown in red) compared to the healthy control (K1, shown in blue). ZNF80 was used for data normalization.

qPCR was followed by the analysis of the expression levels of DYSF mRNA by qRT-PCR using three different assays. As shown in figure 7, DYSF transcript in blood sample was significantly

increased compared to healthy control. Currently, no muscle biopsy was available for western blot analysis.

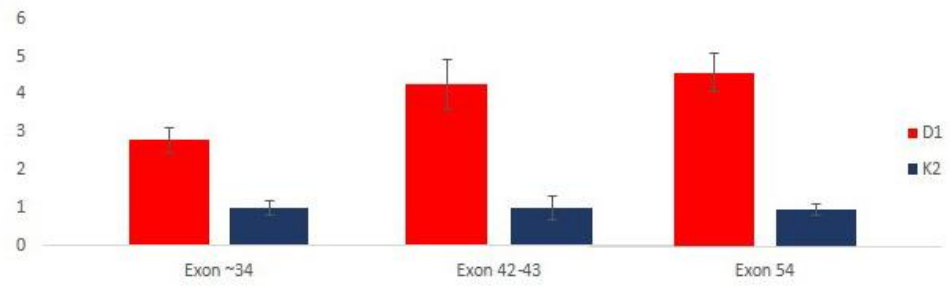


Fig.7: Histograms show the expression levels of dysferlin in blood samples. In red is shown D1 patient while in blue the healthy control (K2). Dysferlin mRNA is significantly increased in D1 patient compared to healthy control.

4.2.1.4 D1_DYSF immunofluorescence

Dysferlin immunofluorescence in muscle sample showed a significantly decreased of DYSF protein at the sarcolemma. Dysferlin protein is still present at sarcolemma but levels are extremely decreased compared to a healthy control (Fig.8).

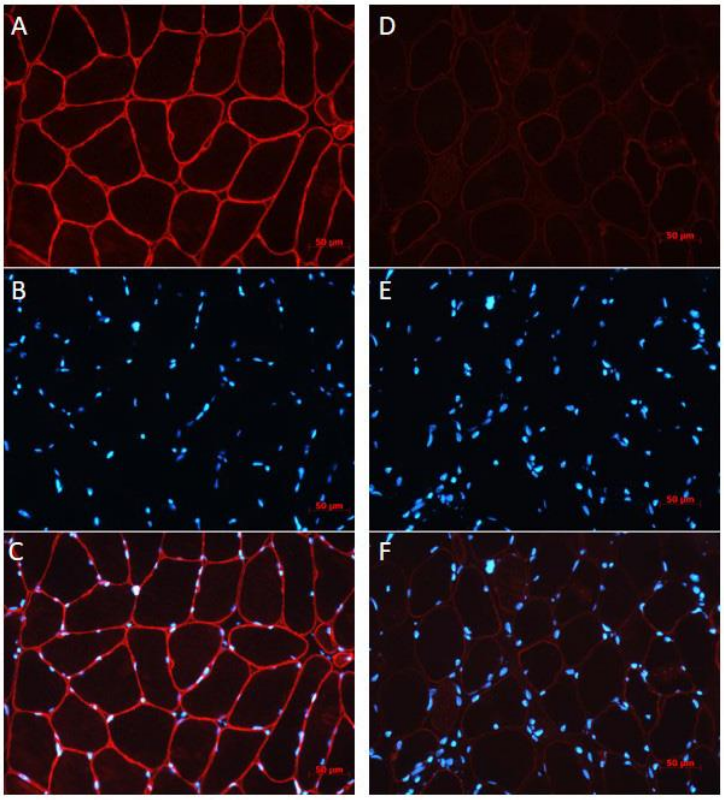


Fig.8: Panel A-B-C: Dysferlin staining (A), Hoechst staining (B) and merge image (C) in the sarcolemma of healthy muscle. Panel D-E-F: Dysferlin staining (D), Hoechst staining (E) and merge image (F) in the sarcolemma of D1. Images took with the same setting at 20X magnification (scale bar = 50µm).

4.2.2 D2_Clinical features

The second patient (D2), a 48-year-old man presented limb-girdle weakness. The first symptoms started at 24 years of age with weakness and ipertrofic features of the pelvic and shoulder girdle muscles. Last clinical examination showed hyperlordosis, waddling gait and severe weakness of proximal muscles. EMG showed myopathic changes and TC showed adipose degeneration of limb girdle, thigh and bilateral leg. There was no cardiac or respiratory involvement. Muscle biopsy showed a complete substitution of muscle with fibro-adipose tissue and increased numbers of internal myonuclei.

4.2.2.1 D2_Genetic analysis

In the second patient, a homozygous missense variant was identified c.5509G>A (p.Asp1837) in exon 49 (Fig.9A). Genetic testing was extended to the father (D2_A) which carried the same variant in heterozygous state (Fig.9B).

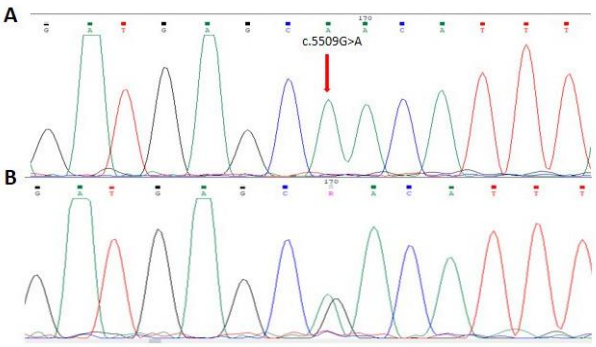


Fig.9: Panel A: Electropherogram shows c.5509G>A variant in homozygous state in exon 49 in D2 sample. Panel B: Electropherogram of D2_A shows c.5509G>A in the heterozygous state.

4.2.2.2 D2_Conservative and *in silico* analysis

Multiple alignments of the amino acid sequence of human DYSF (NP_001123927.1) with the orthologous proteins of nine different species have shown that aspartic acid in exon 49 (p.Asp1837) is highly conserved among all species (Fig.10A).

In silico analysis shows the pathogenic nature of p.Asp1837Asn (Fig.10B).

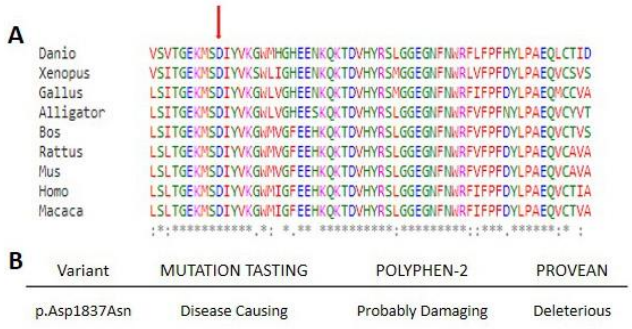


Fig.10: Panel A: Multiple alignment of amino acid sequence of human DYSF (NP_003485.1) with the orthologous proteins of *Macaca mulatta* (XP_014968118.1), *Mus musculus* (XP_006506233.1), *Rattus norvegicus* (XP_017448122.1), *Bos taurus* (XP_015328883.1), *Alligator mississippiensis* (XP_019351176.1), *Danio rerio* (XP_005172350.1), *Gallus gallus* (XP_015141556.1) and *Xenopus leavis* (XP_017946387.1). p.Asp1837 is highlighted with the arrow. Panel B: *In silico* analysis results for the amino acid substitution p.Asp1837Asn.

4.2.2.3 D2_qPCR analysis

For D2 patient only the father was available for genetic *CAPN3* screening. However, it is not sufficient to confirm the homozygosity state of the variant, moreover D2 showed all SNPs in the homozygous state in dysferlin gene. Then, the homozygosity of the variant c.5509G>A was verified with qPCR. The analysis was performed in the proband, the father and a healthy control. Six assays were used and no large deletions were detected confirming the homozygosity of c.5509G>A (Fig.11).

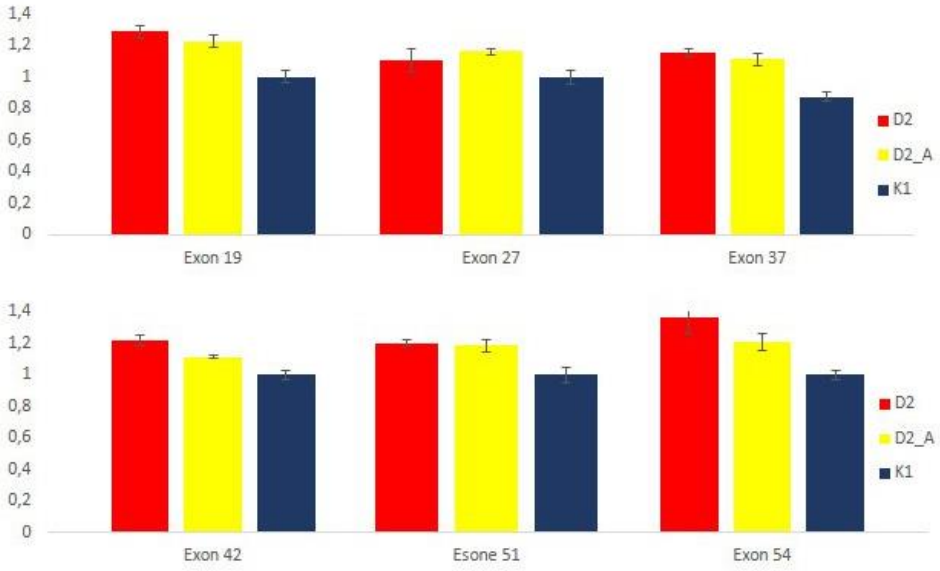


Fig.11: Results of qPCR showing no large deletion in *DYSF* gene in D2 patient (shown in red) compared to the father (D2_A, shown in yellow) and the healthy control (K1, shown in blue).

4.2.2.4 D2_DYSF immunofluorescence

Dysferlin immunofluorescence in D2 muscle sample showed a significantly decreased of DYSF protein at the sarcolemma. Dysferlin protein is not present in all fibres and DYSF left at the sarcolemma is extremely decreased compared to a healthy control (Fig.12).

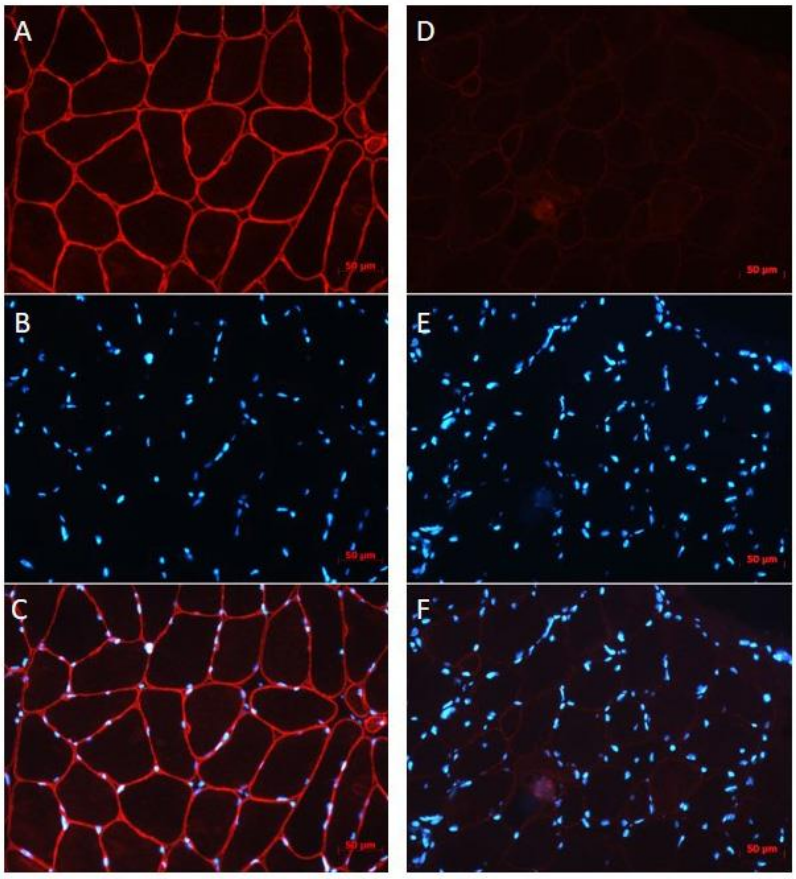


Fig.12: Panel A-B-C: Dysferlin staining (A), Hoechst staining (B) and merge image (C) in the sarcolemma of healthy muscle. Panel D-E-F: Dysferlin staining (D), Hoechst staining (E) and merge image (F) in the sarcolemma of D2. Images took with the same setting at 20X magnification (scale bar = 50µm).

4.2.3 D3_Clinical features

The third patient (D3), a 55-year-old woman presented limb-girdle weakness. The first symptoms started at childhood with difficulty in raising the arms above the head and got worse at 40 years of age with lower back pain and recurrent falls for muscular weakness. Last clinical examination showed waddling gait, severe weakness of proximal muscles and increased serum CK (743 U/L). Muscle biopsy revealed the presence of dystrophic features with increased connective tissue, increased variation in fiber diameter and inflammatory infiltrate. Immunohistochemical analysis revealed normal staining of dystrophin, sarcoglycans and telethonin.

4.2.3.1 D3_Genetic analysis

D3 patient, a heterozygous missense variant was identified c.6139A>G (p.Ile2047Val) in exon 54 (Fig.13A). Genetic testing was extended to the father (D3_B) and the daughter (D3_A). The father did not show the variant while the daughter carried the same heterozygous variant even if she didn't show dysferlinopathy phenotypes (Fig.13B and C).

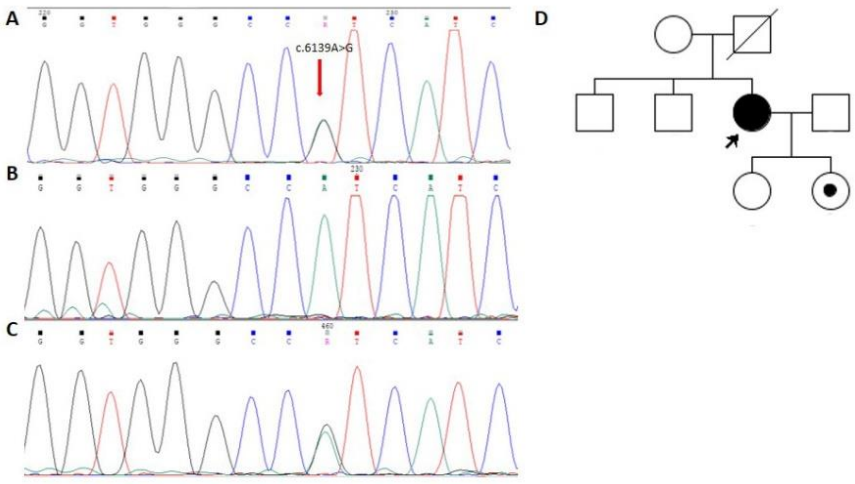


Fig.13: Panel A: D3 electropherogram shows c.6139A>G variant in the heterozygous state in exon 54. Panel B: D3_B electropherogram does not show c.6139A>G variant. Panel C: D3_A electropherogram shows c.6139A>G variant in the heterozygous state. Panel D: D3 family pedigree. D3 is highlighted with the arrow and is clinically affected carrying only one heterozygous mutation. D3_A is shown with a black dot since she is asymptomatic even if she carrying c.6139A>G variant.

In the proband, genetic screening was extended to *CAPN3* gene as well. No pathogenic variants were detected.

4.2.3.2 D3_Quantitative analysis

The entire cDNA, retrotranscribed from a blood sample, was analysed to detect the presence of a possible second variant not observe with *DYSF* screening. No splicing defects were detected in D3 and D3_A. The variant c.6139A>G in both cDNA is presented in heterozygous state (Fig.14).

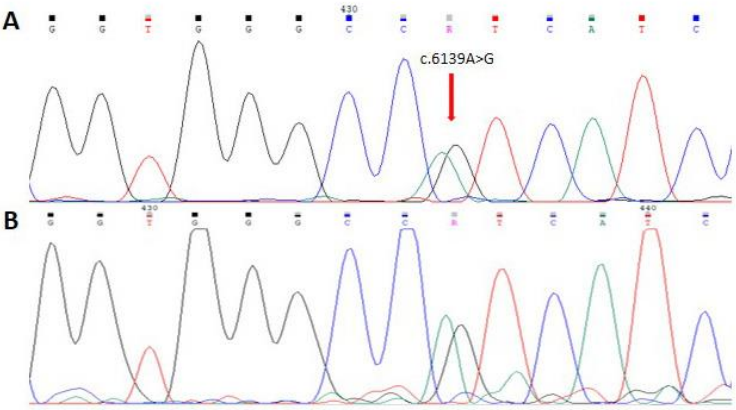


Fig.14: Panel A: Electropherogram of proband cDNA shows c.6139A>G variant in the heterozygous state. Panel B: Electropherogram of daughter cDNA shows c.6139A>G variant in the heterozygous state as well.

4.2.3.3 D3_Conservative and *in silico*

Multiple alignments of the amino acid sequence of human DYSF (NP_003485.1) with the orthologous proteins of nine different species have shown that isoleucine in exon 54 (p.Ile2047) is conserved among the species, except for two species (Fig.15A).

In silico analysis shows conflicting results for the nature of p.Ile2047Val (Fig.15B).

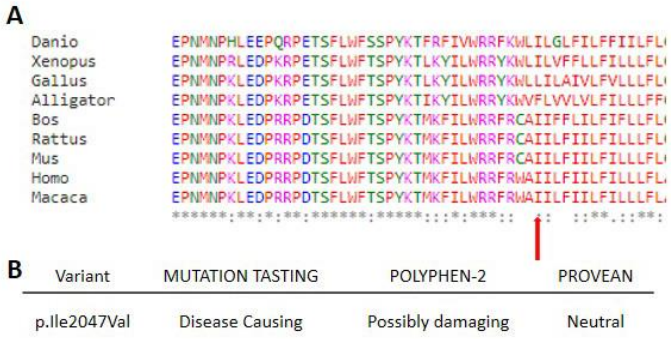


Fig.15: Panel A: Multiple alignment of amino acid sequence of human DYSF (NP_003485.1) with the orthologous proteins of *Macaca mulatta* (XP_014968118.1), *Mus musculus* (XP_006506233.1), *Rattus norvegicus* (XP_017448122.1), *Bos taurus* (XP_015328883.1), *Alligator mississippiensis* (XP_019351176.1), *Danio rerio* (XP_005172350.1), *Gallus gallus* (XP_015141556.1) and *Xenopus leavis* (XP_017946387.1). p.Ile2047 is highlighted with the arrow. Panel B: *In silico* analysis results for the amino acid substitution p.Ile2047Val.

4.2.3.4 D3_qRT-PCR analysis

The expression levels of DYSF mRNA, in the proband and in the asymptomatic daughter, were analysed in blood samples. No differences were observed in both samples compared to healthy control (Fig.16).

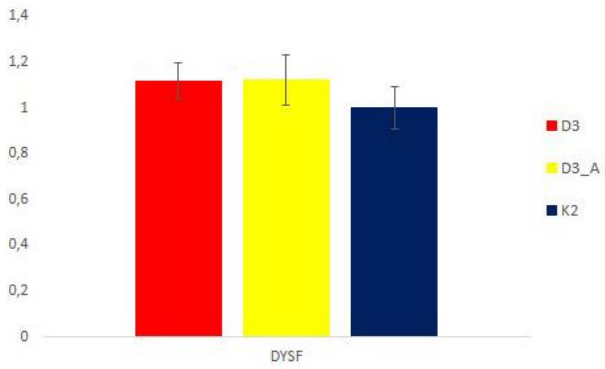


Fig.16: Histograms showing *DYSF* transcript levels in lymphoblastoid cells of the proband shown in red, the daughter shown in yellow and a healthy control shown in blue

4.2.3.5 D3_Dysferlin immunofluorescence

Dysferlin immunofluorescence in D3 muscle sample showed a reduction of DYSF staining compared to the control. Dysferlin protein left at the sarcolemma is decreased compared to a

healthy control. The majority of the fibers showed a weaker staining and regenerating fibers are characterized by dysferlin accumulation within the cytoplasm (Fig.17).

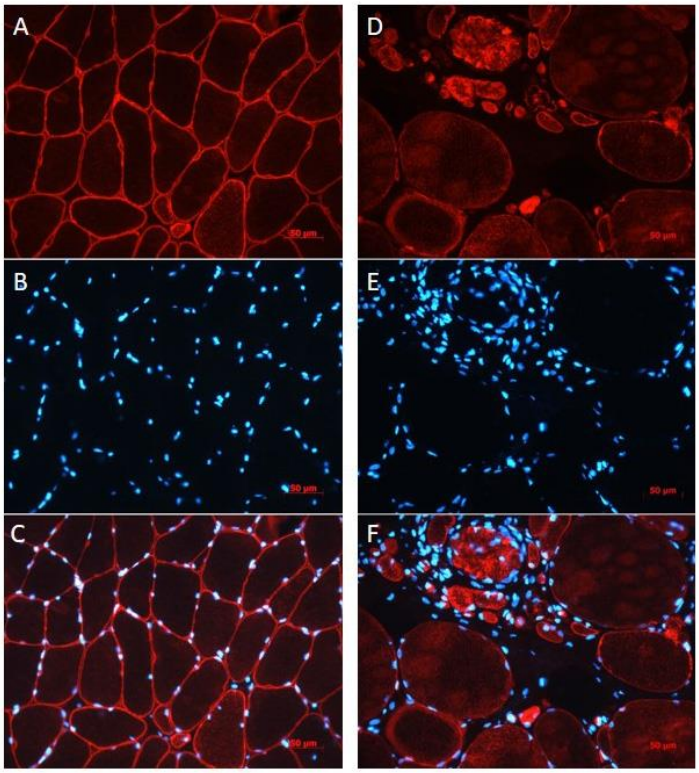


Fig.17: Panel A-B-C: Dysferlin staining (A), Hoechst staining (B) and merge image (C) in the sarcolemma of healthy muscle. Panel D-E-F: Dysferlin staining (D), Hoechst staining (E) and merge image (F) in the sarcolemma of D3. Images took with the same setting at 20X magnification (scale bar = 50µm).

4.2.3.6 D3_Western blot analysis

Contrary to immunofluorescence results, western blot analysis on lymphocyte samples highlighted an overexpressed amounts of DYSF protein in D3 and D3_A compared to normal control. D3 patient and her daughter showed about 2-fold increase in dysferlin protein (Fig18).

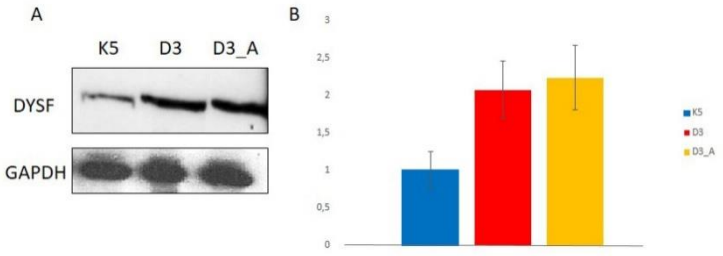


Fig.18: Panel A**:** Western blot analysis of DYSF performed on lymphocyte samples of normal control (K5), D3 and D3_A. Panel B: DYSF expression normalized with respect to GAPDH expression.

4.3 CAPN3 PATIENTS

4.3.1 *In silico* analysis

Before starting *in silico* analysis, the software tools were checked for prediction of the effects of a collection of 160 known pathogenic variants, acquired from literature and databases, in order to attest the rate of correct predictions relatively to mutations affecting *CAPN3* gene. The test proved a sensibility of 97% for Mutation Taster, 90% for PolyPhen-2, 78% for PROVEAN, and 77% for snSNPAnalyzer while SIFT, Panther, MutPred and PMut showed a low sensibility.

4.3.2 C1_Clinical features

In C1 family, three siblings born to unrelated parents presented limb-girdle muscle weakness. In the family pedigree, the mother was reported to present muscle weakness. Patient C1 was a 59-year-old man presenting with gait difficulties since 20 years. The patient's early development was normal and the past medical history was unremarkable. Over the following years, he was repeatedly referred to our hospital due to easy fatigability and progressive muscle weakness. EMG showed myopathic changes. Serum CK was highly increased (2341 U/L). He showed waddling gait, calf hypertrophy and scapular winging. He presented severe and symmetrical weakness of shoulder or pelvic girdles. Pulmonary and cardiac functions were normal. Muscle biopsy was consistent with chronic dystrophic-like features including variability in myofiber diameters, atrophic and hypertrophic fibers, fiber splitting and increase in adipose and fibrous tissue. Degenerating and/or regenerating fibers were also present. Immunohistochemical studies showed a normal membranous pattern of staining with antibodies to dystrophin and sarcoglycans.

The 56-year-old sister (C1_A) presented progressive weakness in his right biceps brachii muscle starting at 40 years of age, followed by diffuse muscle weakness. EMG showed myopathic changes, serum CK was increased (2118 UI/l at the age of 48, 1413 Ul/l at the age of 51). In the last neurological examination, the patient showed a clinical condition similar to his brother. Muscle biopsy showed variation in fibers size with atrophic fibers and internal nuclei. There were no degenerating or regeneration fibers. Immunohistochemical studies showed a normal membranous pattern of staining with antibodies to dystrophin and sarcoglycans.

The 55-year-old sister (C1_B) referred progressive muscle weakness since 40 years. EMG showed myopathic changes and serum CK was elevated (1767 UI/l). Neurological examination showed limb girdle muscle weakness although less severe than her siblings. Muscle biopsy showed the presence

of atrophic and hypertrophic fibers, without necrosis. Immunohistochemical studies showed normal expression of dystrophin and sarcoglycans.

4.3.2.1 C1_Genetic analysis

C1 patient showed two known pathogenic variants in heterozygous status: c.755T>G (p.Met252Arg) in exon 5 and c.1746-20C>G in intron 14. Genetic testing was extended to siblings, one of them carried both mutations in heterozygous status (C1_A) whereas the other carried only the heterozygous c.755T>G variant (C1_B) (Fig.19). Both variants were detected even in cDNA in heterozygous state as well.

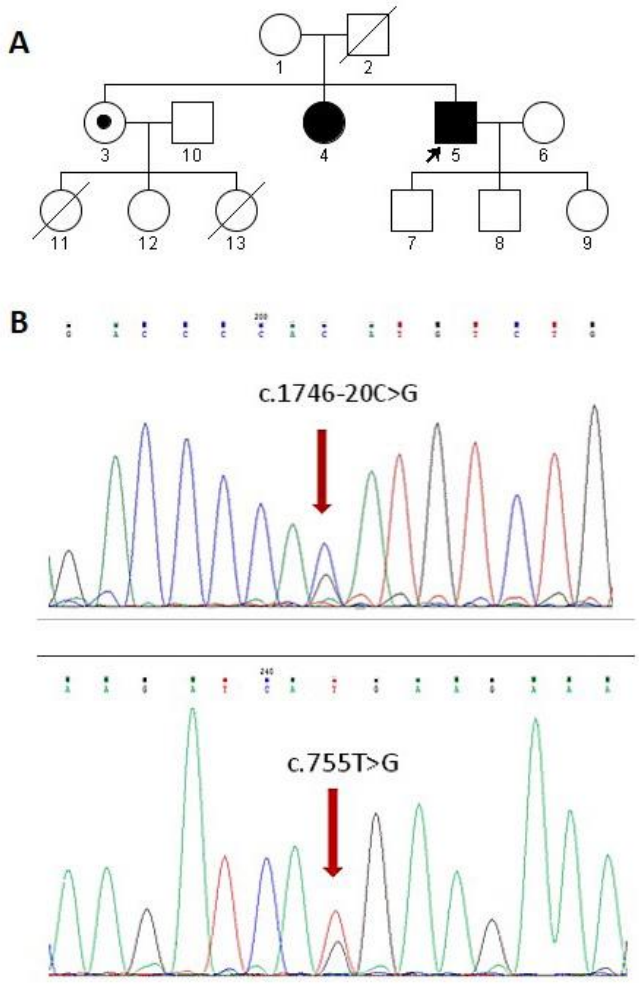


Fig.19: Panel A: Family pedigree. C1 is highlighted with the arrow Panel B: Electropherograms show c.1746-20C>G variant in intron 14 and c.755T>G variant in exon 5 both in the heterozygous state.

4.3.2.2 C1_Conservative, 3D and *in silico* analysis

Multiple alignments of the amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of nine different species have shown that p.Met252 is conserved among species except for *Xenopus leavis* (Fig.20A).

In silico analysis shows the pathogenic nature of p.Met252Arg (Fig.20B).

Within the 3D structure of CAPN3 p.Met252 is located in a α-helix within the 3D structure inside the domain II (Fig.20C).

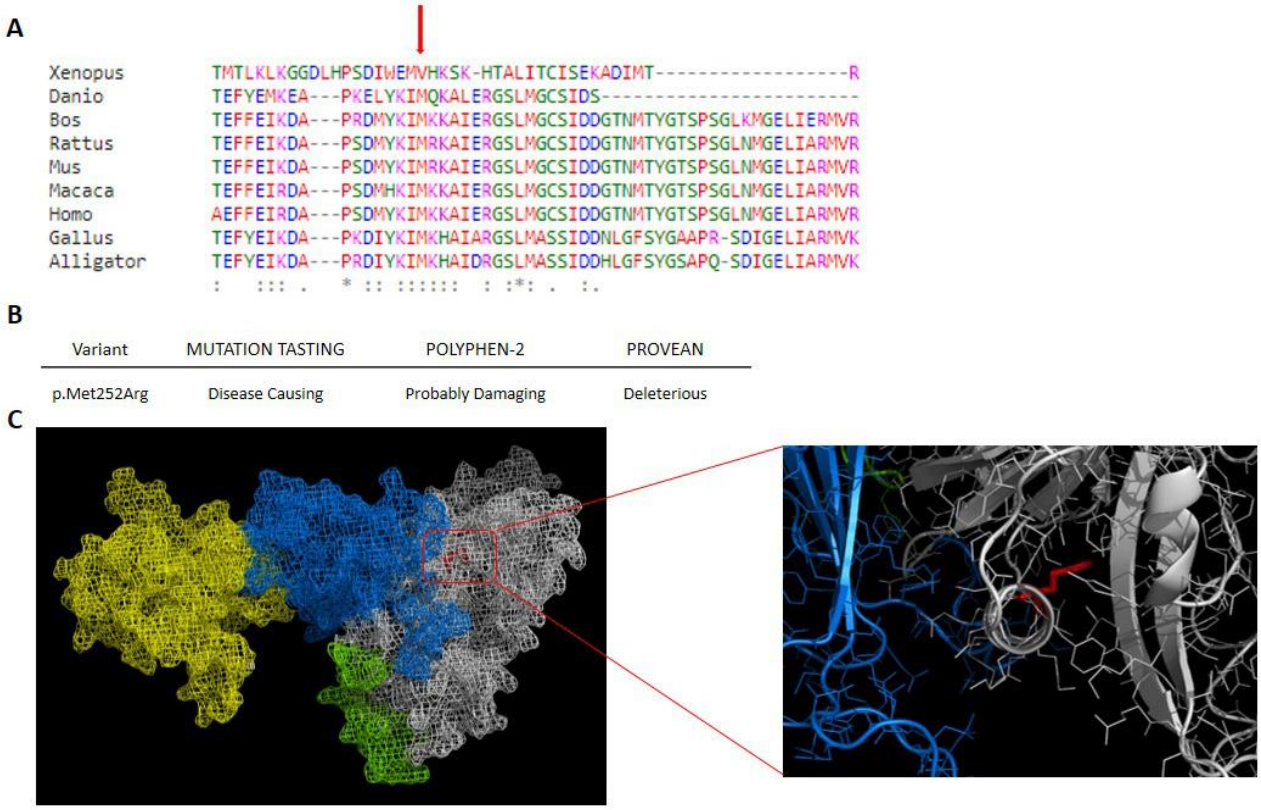


Fig.20: Panel A: Multiple alignment of amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of *Macaca mulatta* (XP_001103220.1), *Mus musculus* (NP_031627.2), *Rattus norvegicus* (NP_058813.1), *Bos taurus* (NP_776685.1), *Alligator mississippiensis* (XP_006258549.1), *Danio rerio* (NP_001004571.1), *Gallus gallus* (NP_001004405.2) and *Xenopus leavis* (NP_001087053.1). p.Met252 is highlighted with the arrow. Panel B: *In silico* analysis results for the amino acid substitution p.Met252Arg. Panel C: Representation of the 3D structure of CAPN3. In green is shown domain I, in white domain II, in blue domain III and in yellow domain IV. While in red is shown the localization of p.Met252 within the 3D structure of the protein.

Using the mutagenesis tool in PyMOL, it was possible to highlight that p.Met252Arg variation doesn't generate a steric hindrance for the similar dimension of both amino acids. However, the addition of the arginine positive charge within the 3D structure could be neutralized by a negative charge nearby. It is likely that the negative charge of aspartic acid in position 402 interacts with the

positive charge of arginine. Since the distance between the two amino acids is 3.5Å, which is compatible with electrostatic interaction and/or hydrogen bonding (Fig.21).

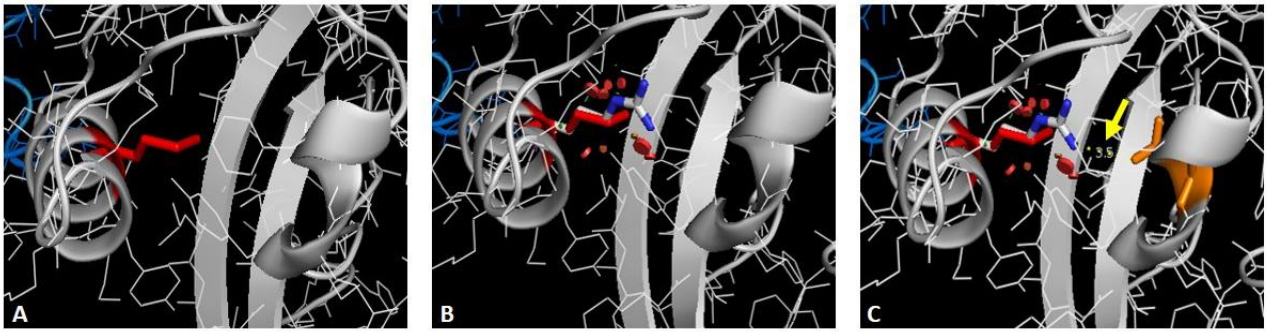


Fig.21: Panel A: p.Met252 shown in red within a α-helix of CAPN3. Panel B: Representation of p.Arg252 using mutagenesis wizard of PyMol. Panel C: Yellow arrow highlights the distance between p.Arg252 and p.Asp402 (p.Asp402 is shown in orange).

4.3.2.3 C1_RT-qPCR analysis

Analysis of the transcript levels were performed in proband, siblings and healthy control, with three couple of primers localized in different regions of the transcript. RT-qPCR highlighted increased levels of the transcript in C1 and C1_A compared to healthy control (K3). In all assays, transcript levels are almost two times more compared to control. Unlike C1 and C1_A, C1_B did not show a significant increase of the transcript. In exons 5/6 and exons 9/10 transcript levels are comparable to the healthy control. Only exons 10/11 showed an increase of the transcript even if lower than C1 and C1_A (Fig.22).

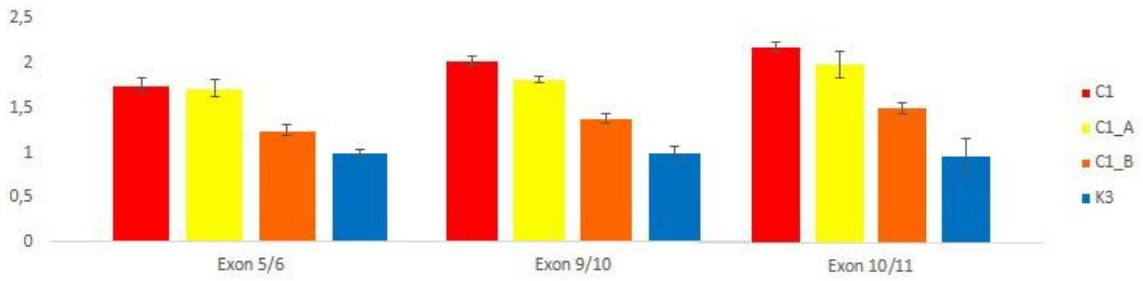


Fig.22: Relative normalized expression of calpain 3 transcript levels. Histograms represent total transcript level of calpain 3 in C1 shown in red, C1_A shown in yellow, C1_B shown in orange compared to a healthy control (K3) shown in blue.

4.3.3 C2_Clinical features

Patient C2 was a 74-year-old woman presented limb-girdle weakness. The first symptoms started at 45 years of age and she was faced with a gradually developing difficulty of going up and downstairs and standing. Clinical examination showed scapular winging, waddling gait severe weakness of upper and lower girdle muscles. EMG showed myopathic changes and serum CK was elevated (1169 UI/l). There was no cardiac or respiratory involvement. Muscle biopsy showed increased fiber size

variability with 10% central nuclei and diffuse increased connective tissue. Immunohistochemical studies showed normal expression of dystrophin, sarcoglycans, dysferlin.

4.3.3.1 C2_Genetic analysis

CAPN3 screening showed three previously reported pathogenic mutations: a single base pair mutation encoding a premature stop codon c.967G>T (p.Glu323*) in exon 7, a single amino acid deletion c.1401_1403delGGA (p.Glu467del) in exon 11 and a missense mutation c.2257G>A (p.Asp753Asn) in exon 21 (Fig.23). Segregation analysis allowed us to establish that mutations c.1395_1397delGGA and c.2257G>A were localized in the same allele. All variants were detected in cDNA in heterozygous state as well.

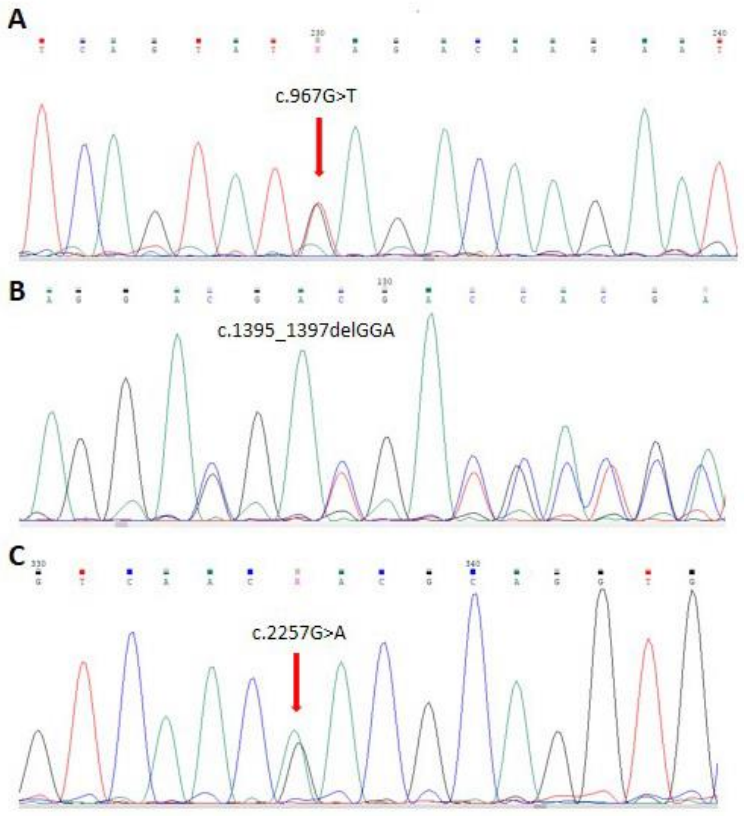


Fig.23: Panel A: Electropherogram shows c.967G>T variant in exon 7 in the heterozygous state. Panel B: Electropherogram shows the deletion of three nucleotides c.1401_1403delGGA in exon 11 in the heterozygous state. Panel C: Electropherogram shows the missense variant c.2257G>A in exon 21 in the heterozygous state.

4.3.3.2 C2_Conservative, 3D and *in silico* analysis

Multiple alignments of the amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of nine different species have shown that the glutamic acid deleted in exon 11 (p.Glu467) and the aspartic acid in exon 21 (p.Asp753) are conserved among species except for *Xenopus leavis* (Fig.24A and B).

In silico analysis was performed only for the missense mutation and the results of four reliable software are conflicting (Fig.24C).

The 3D analysis was performed in two of the three variants identified. The mutation c.1395_1397delGGA leads to deletion of glutamic acid and it occurs in the domain III of CAPN3. Deletion is located inside a loop pointed towards the external surface of the protein. While the amino acid substitution p.Asp753Asn is localized on the surface of CAPN3, inside one of the five EF-hand (PEF) domains (Fig.24D).

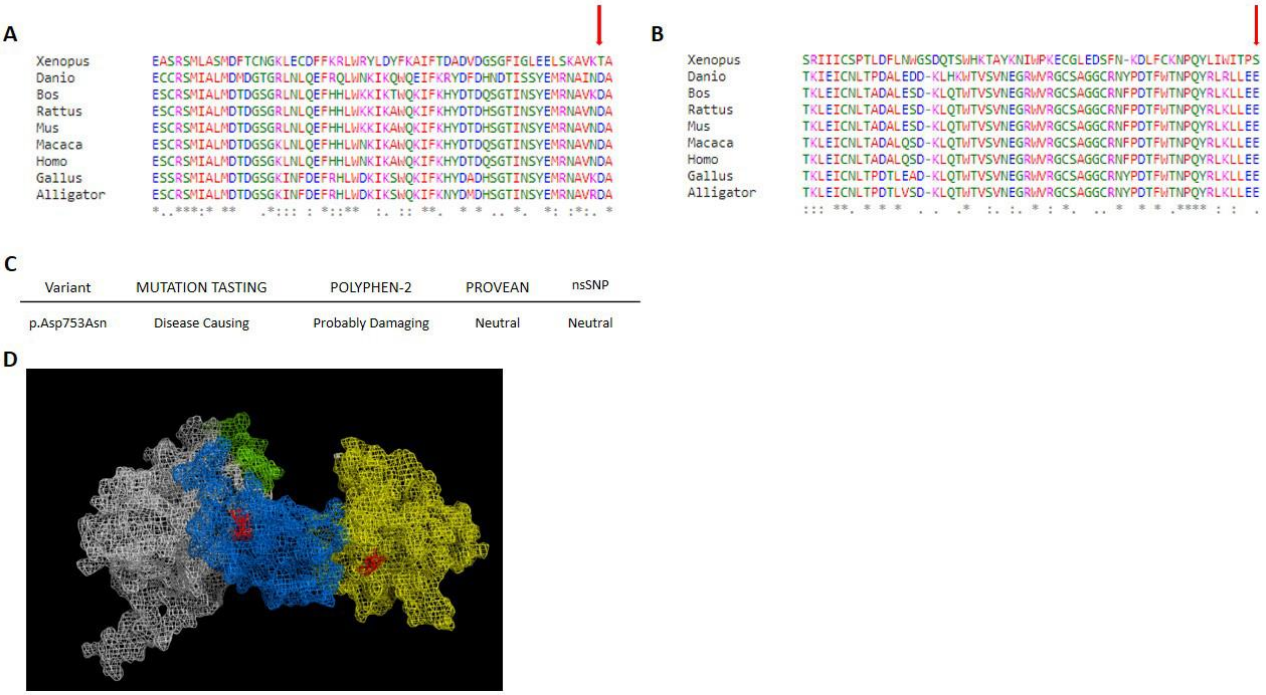


Fig.23: Panel A: Multiple alignment of amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of *Macaca mulatta* (XP_001103220.1), *Mus musculus* (NP_031627.2), *Rattus norvegicus* (NP_058813.1), *Bos taurus* (NP_776685.1), *Alligator mississippiensis* (XP_006258549.1), *Danio rerio* (NP_001004571.1), *Gallus gallus* (NP_001004405.2) and *Xenopus leavis* (NP_001087053.1). p.Glu467 is highlighted with the arrow. Panel B: Multiple alignment of amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of *Macaca mulatta* (XP_001103220.1), *Mus musculus* (NP_031627.2), *Rattus norvegicus* (NP_058813.1), *Bos taurus* (NP_776685.1), *Alligator mississippiensis* (XP_006258549.1), *Danio rerio* (NP_001004571.1), *Gallus gallus* (NP_001004405.2) and *Xenopus leavis* (NP_001087053.1). p.Asp753 is highlighted with the arrow. Panel C: *In silico* analysis results for the amino acid substitution p.Asp753Asn. Panel D: Representation of the 3D structure of CAPN3. In green is shown domain I, in white domain II, in blue domain III and in yellow domain IV. While in red are shown the localization of p.Asp753 inside domain IV and p.Glu467del inside domain III.

Using the mutagenesis tool in PyMOL, it was possible to highlight that p.Asp753Asn variation does not generate a steric hindrance, since the similar dimension of both amino acids and even because the substitution is found in the surface of the protein, characterised by more flexibility (Fig.25A). However, p.Asp753Asn is localized in EF3 domain that is crucial for the binding of Ca^{2+} ions. The changing of one charged amino acid with a polar one could modify that interaction. While the deletion

of glutamic acid (p.Glu467del) occurs inside a loop pointing to the external surface. Deletion of one amino acid in that region does not interfere with protein folding but it could disturb the linkage of calpain 3 with the sarcomere protein titin, essential for control the autolytic function of CAPN3. Since p.Glu467 is localized close to the binding site needed for the interaction with titin (Val575-Leu580). It is possible that the negative charge of the glutamic acid interacts with the charged p.Lys575, since the distance between the two amino acids, is compatible for interaction (Fig.25C).

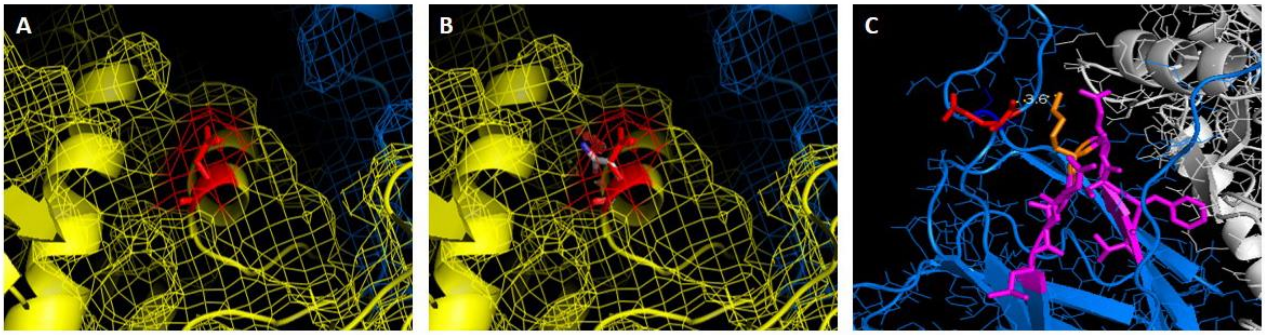


Fig.25: Panel A: p.Asp753 shown in red within the α-helix on the external surface of CAPN3. Panel B: Representation of p.Asn753 using mutagenesis wizard of PyMol. Panel C: Distance between p.Glu467 shown in red and p.Lys575 shown in orange. Potential interaction is shown as a yellow dots. While in magenta is shown the titin binding site (Val575-Leu580).

4.3.4 C3_Clinical features

Patient C4 was a 21 years old woman born to unrelated parents presented limb-girdle muscle weakness. Even her cousin shows a clinical pattern compatible with limb girdle muscular dystrophy. The first symptoms started at 16 years of age with hyposthenia especially of lower girdle muscles compared to upper girdle muscle and got worse after pregnancy. Moreover, she showed waddling gait, difficulty raising the arms above the head and serum CK was highly increased (3856 U/L in 2016 and 5018 U/L in 2017). Pulmonary and cardiac functions were normal. Muscle biopsy showed atrophic and hypertrophic fibers, internal nuclei and necrosis. Immunohistochemical studies showed a normal membranous pattern of staining with antibodies to dystrophin and sarcoglycans.

4.3.4.1 C3_Genetic analysis

CAPN3 screening showed a single missense mutation c.2257G>A (p.Asp753Asn) in exon 21 in heterozygous state (Fig.26). Same variant detected in C2 patient.

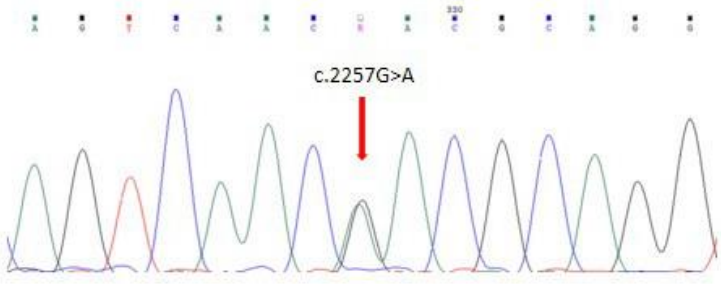


Fig26: Electropherogram shows c.2257G>A variant in exon 21 in the heterozygous state.

4.3.4.2 C3_Conservative, 3D and *in silico* analysis

Conservative analysis of p.Asp753 is shown in figure 23B, *in silico* analysis in figure 24C while 3D analysis in figure 24D, 25A and B.

4.3.4.3 C3_Western blot analysis

Western blot analysis, in muscle sample, showed no modification in CAPN3 protein levels, while an increased DYSF protein levels compared to the healthy control (Fig.27).

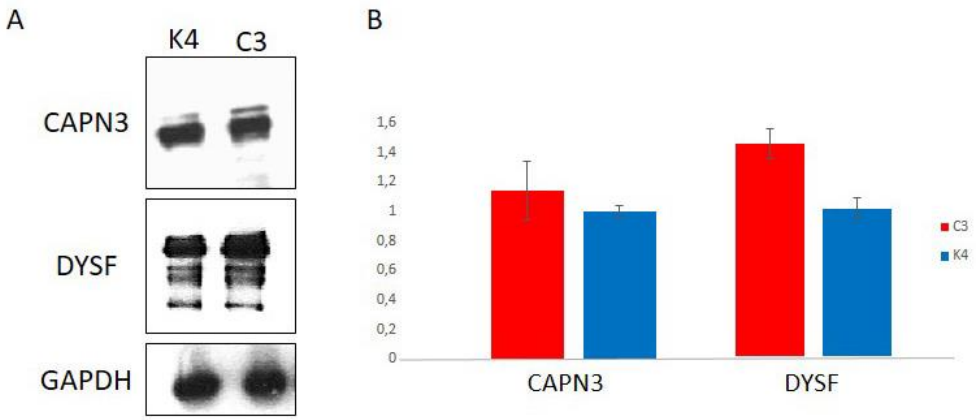


Fig.27: Panel A**:** Western blot analysis of CAPN3 and DYSF on skeletal muscle of healthy control (K4) and C3 patient. Panel B: CAPN3 and DYSF protein expression normalized with respect to GAPDH expression.

4.3.4.4 C3_DYSF immunofluorescence

Dysferlin staining in C3 patient showed increased levels of DYSF compared to control. This result is an accordance with western blot analysis. Dysferlin protein is not only localized at the sarcolemma level but even accumulated into the cytoplasm (Fig.28).

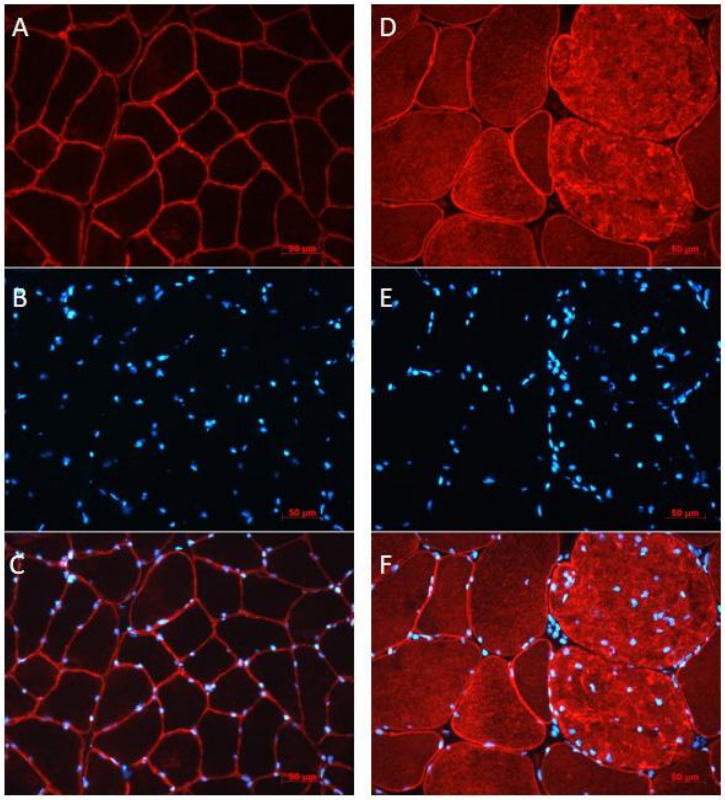


Fig.28: Panel A-B-C: Dysferlin staining (A), Hoechst staining (B) and merge image (C) in the sarcolemma of healthy muscle. Panel D-E-F: Dysferlin staining (D), Hoechst staining (E) and merge image (F) in the sarcolemma of C3. Images took with the same setting at 20X magnification (scale bar = 50µm).

4.3.5 C4_Clinical features

Patient C4 was a 60-year-old man presenting easy fatigability and progressive limb-girdle muscular weakness since adolescence (15 years). Last clinical examination showed waddling gait with weakness in the shoulder and pelvic girdles. Serum CK was 191 Ul/l. Muscle biopsy showed mild variation in fiber size and internal nuclei. Immunohistochemical analysis revealed normal stain of dystrophin and sarcoglycans.

4.3.5.1 C4_Genetic analysis

CAPN3 screening in C4 patient showed only one missense mutation in exon 4, c.526G>A (p.Val176Met) in heterozygous state (Fig.29B). Genetic testing was extended to close asymptomatic family members. This variant has been detected in mother, sibling, daughter and niece in heterozygous state as well. p.Val176Met is a new variant not yet reported, then PCR-RFLP analysis was performed to evaluate the presence of this new variant within 320 chromosomes. PCR-RFLP analysis showed the absence of p.Val176Met in 320 chromosomes (Fig.29C-D).

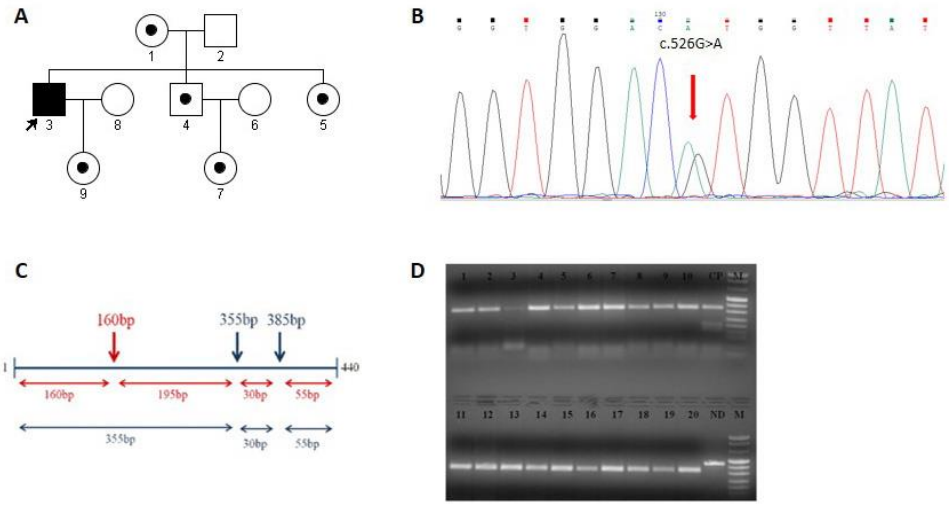


Fig.29: Panel A: C4 pedigree Panel B: Electropherogram shows c.526G>A variant in heterozygous state. Panel C: Schematic representation of sites for restriction enzyme NlaIII. In blue is shown the two sites recognized on wild-type allele and the three fragments generating (355bp, 55bp, 30bp), while in red is shown the adding site recognized on mutant allele and the four fragments generating (195bp, 160bp, 55bp and 30bp). Panel D: Electrophoresis through 2% agarose gel of 20 samples with C4 (positive control, CP) and not digested sample (ND). The yielded restriction profiles were compared with molecular weight marker VIII.

The genetic screening was extended even for *DYSF* gene but no pathogenic variants were detected.

4.3.5.2 C4_Conservative, 3D and *in silico* analysis

Multiple alignments of the amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of nine different species have shown that valine in exon 4 (p.Val176) is highly conserved among all species (Fig.30A).

In silico analysis shows the pathogenic nature of p.Val176Met (Fig.30B).

The 3D analysis shows that p.Val176Met is located in a beta sheet within the 3D structure inside the domain II of the protein (Fig.30C).

A

```
Xenopus     VVIDDRLPFVNGDFFSVRPSCTKEYWPCLLEKAYAKLLGGYQNLHWGDPAEAMVNFTGAL
Danio       VVVDDRIPTFNNQLVFTKSAERNEFWSALLEKAYAKLHGSYEALKGGNTAEGMEDFTGGV
Bos         VVIDDCLPTYNNQLVFTKSNHRNEFWNALLEKAYAKLHGSYEALKGGNTTEAMEDFTGGV
Rattus      VVIDDCLPTYNNQLVFTKSNHRNEFWSALLEKAYAKLHGSYEALKGGNTTEAMEDFTGGV
Mus         VVIDDCLPTYNNQLVFTKSNHRNEFWSALLEKAYAKLHGSYEALKGGNTTEAMEDFTGGV
Macaca      VVIDDCLPTYNNQLVFTKSNHRNEFWSALLEKAYAKLHGSYEALKGGNTTEAMEDFTGGV
Homo        VVIDDCLPTYNNQLVFTKSNHRNEFWSALLEKAYAKLHGSYEALKGGNTTEAMEDFTGGV
Gallus      VIIDDCLPTYNNQLVFTKSSQRNEFWSALLEKAYAKLHGSYEALKGGNTTEAMEDFTGGV
Alligator   VVIDDYLPTYNNHLIFTKSSQRNEFWSALLEKAYAKLHGSYEALKGGNTTEAMEDFTGGV
            *::** :*  *.:. .:    :*:* .******** *.*: *: *: :*.* :***.:
```

B

Variant	MUTATION TASTING	POLYPHEN-2	PROVEAN	nsSNP
p.Val176Met	Disease Causing	Probably Damaging	Deleterious	Disease

C

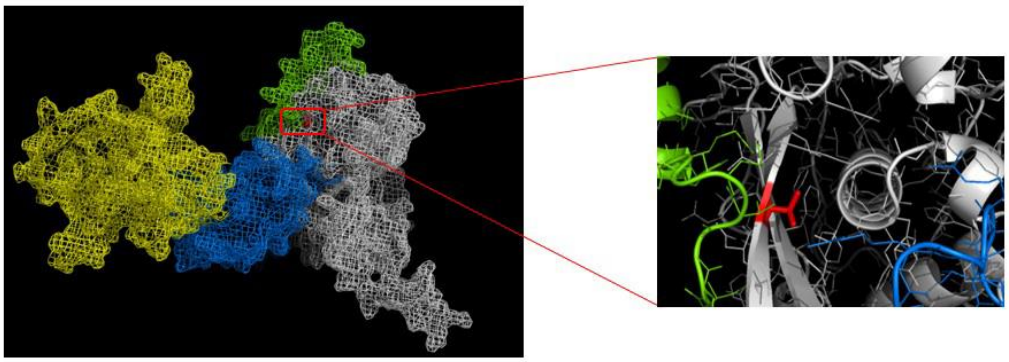


Fig.30: Panel A: Multiple alignment of amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of *Macaca mulatta* (XP_001103220.1), *Mus musculus* (NP_031627.2), *Rattus norvegicus* (NP_058813.1), *Bos taurus* (NP_776685.1), *Alligator mississippiensis* (XP_006258549.1), *Danio rerio* (NP_001004571.1), *Gallus gallus* (NP_001004405.2) and *Xenopus leavis* (NP_001087053.1). p.Val176 is highlighted with the arrow. Panel B: *In silico* analysis results for the amino acid substitution p.Val176Met. Panel C: Representation of the 3D structure of CAPN3. In green is shown domain I, in white domain II, in blue domain III and in yellow domain IV. While in red is shown the localization of p.Val176 within domain II on a β-sheet pointing to a central α-helix.

Using the mutagenesis tool in PyMOL, it was possible to highlight that p.Val176Met variant is localised within a well defined 3D structure. There is a central α-helix surrounded by other secondary structures. Substitution of valine versus methionine generates a steric hindrance since the dimension of these two amino acids is completely different (Fig.31).

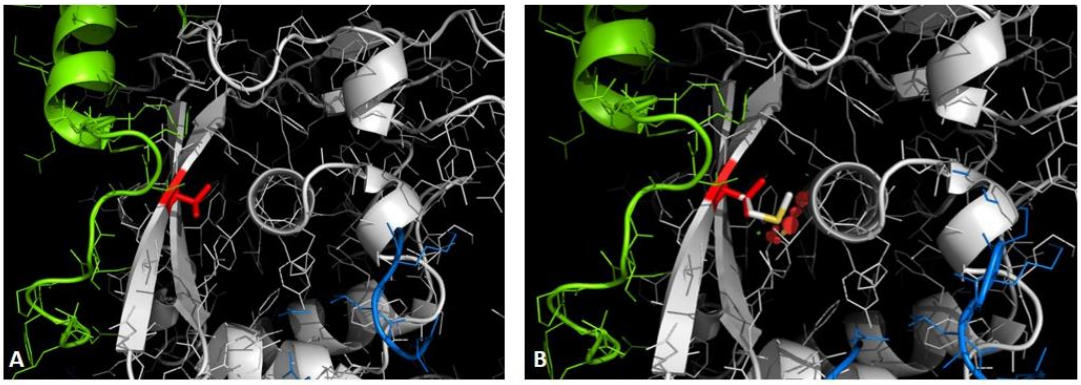


Fig.31: Panel A: In red is shown pVal176 within a β-sheet of CAPN3 domaini II. Panel B: Representation of p.Met176 using mutagenesis wizard of PyMol.

4.3.5.3 C4_Qualitative analysis

The entire cDNA, retrotranscribed from skin sample, was amplified into five overlapping regions to detect the presence of a possible second variant not observe with *CAPN3* screening. However, no alterations were observed and the variant c.526G>A has been found in the heterozygous state even in the transcript (Fig.32).

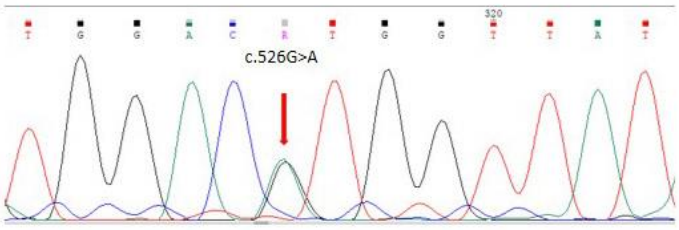


Fig.32: Electropherogram of the transcript shows c.526G>A variant in heterozygous state.

4.3.5.4 C4_DYSF immunofluorescence

Even if no pathogenic variants were detected with dysferlin screening, immunofluorescence was performed to evaluate the presence of a possible secondary reduction of DYSF protein. No differences were observed between C4 and healthy control in dysferlin staining, the patient did not show reduced plasma membrane expression of dysferlin protein (Fig.33).

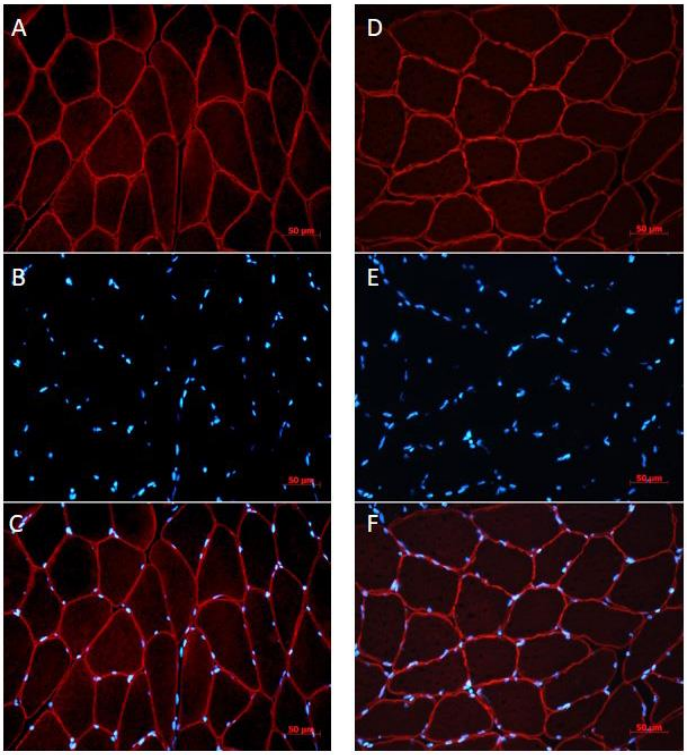


Fig.33: Panel A-B-C: Dysferlin staining (A), hoechst staining (B) and merge image (C) in the sarcolemma of healthy muscle. Panel D-E-F: Dysferlin staining (D), hoechst staining (E) and merge image (F) in the sarcolemma of C4. Images took with the same setting at 20X magnification (scale bar = 50µm).

4.3.5.5 C4_Western blot analysis

Western blot analysis in muscle sample showed a reduction of CAPN3 protein levels in patient C4 compared to the healthy control. Calpain 3 showed a reduction of 30%. Interestingly, CAPN3 expression was evaluated even in patient's skin sample, which showed the same reduction of 30% compared to healthy control (Fig.34).

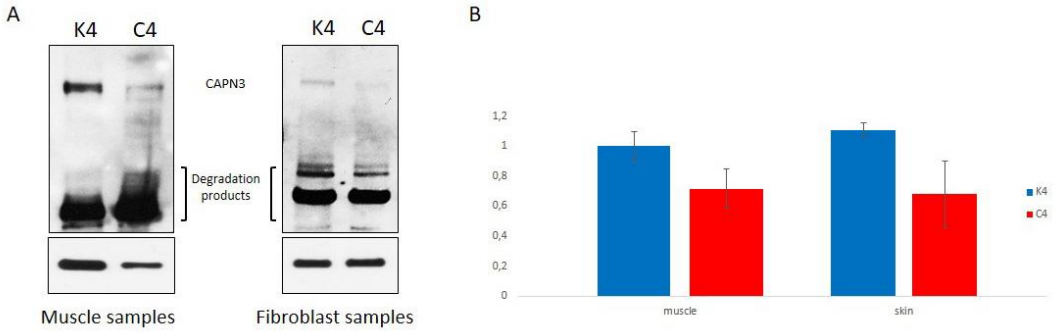


Fig.34: Panel A**:** Western blot analysis of CAPN3 on muscle and fibroblast samples of healthy control (K4) and patient C4. Panel B: CAPN3 expression normalized with respect to GAPDH expression.

4.3.6 C5_Clinical features

C5 was a 51-year-old man presenting with gait difficulties since the age of 35 years and progressive limb-girdle muscular weakness. Clinical examination showed waddling gait with diffuse muscle wasting, scapular winging. Severe weakness in the shoulder and pelvic girdles were also present. Pulmonary and cardiac functions were normal. Muscle biopsy revealed the presence of dystrophic features with increased connective tissue, increased variation in fibers size, internal nuclei and several lobulated fibers. There were also necrosis and degenerating/regenerating fibers. Immunohistochemical analysis revealed normal stain of dystrophin, sarcoglicans, dysferlin, caveolin.

4.3.6.1 C5_Genetic analysis

C5 patient showed one novel missense mutation in exon 24 in the heterozygous state: c.2458T>C (p.Tyr820His) (Fig.35A). Since the non-availability of the sample, it has not been possible carry on additional molecular analysis. PCR-RFLP analysis confirmed the absence of amino acid variation in 320 chromosomes (Fig.35B-C).

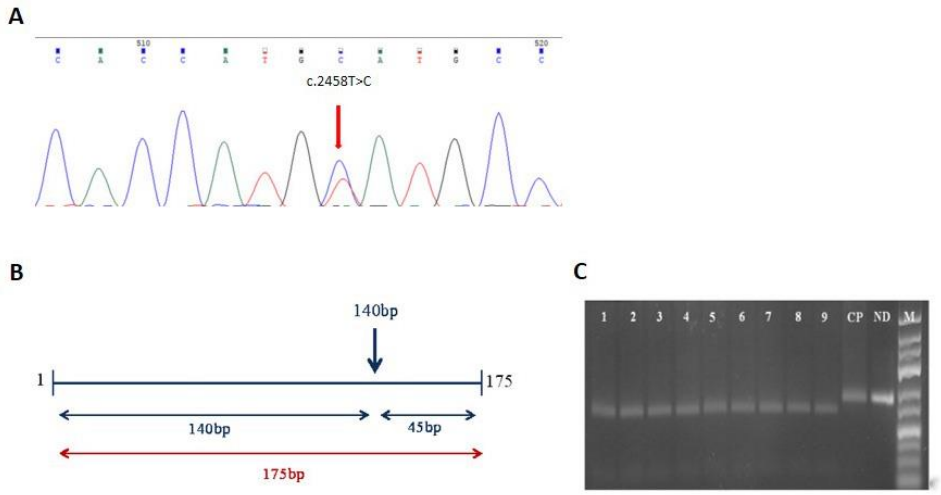


Fig.35: Panel A: Electropherogram shows c.2458T>C variant in heterozygous state. Panel B: Schematic representation of sites for restriction enzyme RsaI. In blue is shown the site recognized on wild-type allele and the two fragments generating (140bp, 45bp), while the enzyme does not recognize any restriction site on mutant allele. Panel C: Electrophoresis through 2% agarose gel of 9 samples with C5 (positive control, CP) and not digested sample (ND). The yielded restriction profiles were compared with molecular weight marker VIII.

4.3.6.2 C5_Conservative, 3D and *in silico* analysis

Multiple alignments of the amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of nine different species have shown that tyrosine in exon 24 (p.Tyr820) is conserved among species except for Xenopus leavis (Fig.36A).

In silico analysis with the four reliable software shows in three software the pathogenic nature of p.Tyr820His (Fig.36B).

For the last patient, it was not possible to analyze the variant p.Tyr820His into the 3D structure of CAPN3 since the amino acid substitution is located in a region not perfectly aligned with human m-calpain. However, the amino acid before the mutation in position 819 is localised in the domain IV inside the EF5-hand. Therefore is highly possible that p.Tyr820His is localised in the EF5-hand as well (Fig.36C).

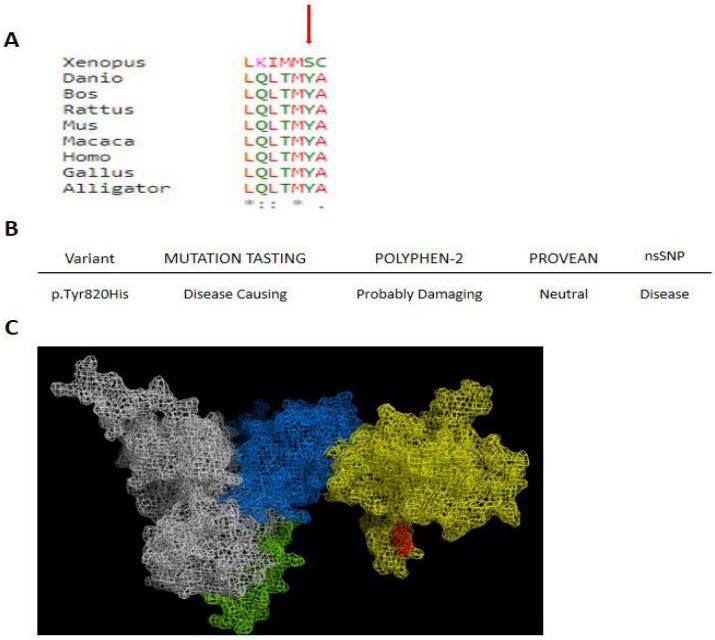


Fig.36: Panel A: Multiple alignment of amino acid sequence of human CAPN3 (NP_000061.1) with the orthologous proteins of *Macaca mulatta* (XP_001103220.1), *Mus musculus* (NP_031627.2), *Rattus norvegicus* (NP_058813.1), *Bos taurus* (NP_776685.1), *Alligator mississippiensis* (XP_006258549.1), *Danio rerio* (NP_001004571.1), *Gallus gallus* (NP_001004405.2) and *Xenopus leavis* (NP_001087053.1). p.Tyr820 is highlighted with the arrow. Panel B: *In silico* analysis results for the amino acid substitution p.Tyr820His. Panel C: Representation of the 3D structure of CAPN3. In green is shown domain I, in white domain II, in blue domain III and in yellow domain IV. While in red is shown the localization of p.Met819 within EF5-hand of CAPN3 domain IV.

5. DISCUSSION

5.1 DYSFERLIN PATIENTS

5.1.1 D1 patient

DYSF screening highlighted, in a 55-year-old man presented limb-girdle weakness, one missense mutation (p.Arg2042Cys) in exon 54 in homozygous state. This variant c.6124C>T (p.Arg2042Cys) has been already reported in several papers but never in homozygous state[45,176,179–181] (Tab.8)

Reference	Gender	Age of onset	Phenotype	Genotype
Liu et al., 1998	M		MM, LGMD2B	p.Arg2042Cys; p.Ile1298Val
Cagliani et al., 2005	M	40	MM, LGMD2B	p.Arg2042Cys; p.Glu1994ArgfsX2
Therrien et al., 2006		15	LGMD2B	p.Arg2042Cys; ND
Gallardo et al., 2011	F	29	MM	p.Arg2042Cys; p.Glu1994ArgfsX3
Vincent et al., 2016	F		LGMD2B	p.Arg2042Cys; p.Asp1837Asn

Tab.8: Reported cases in which p.Arg2042Cys has been detected.

As shown in the table p.Arg2042Cys is correlated with two different phenotypes: LGMD2B and MM. In the present study p.Arg2042Cys was detected for the first time in homozygous state and it is correlated with LGMD2B phenotype.

We found that p.Arg2042Cys mutation significantly increased the expression of the gene (Fig.7) and this finding seems to be in contrast with immunofluorescence analysis, which showed a notably decreased of DYSF protein at the sarcolemma level (Fig.8). p.Arg2042Cys mutation is localized close to transmembrane domain (2045-2067) and could interfere with insertion of the protein in the sarcolemma, explaining low level of DYSF protein in this region. However, this is just one hypothesis since it was not possible to perform western blot analysis to quantify dysferlin protein. Western blot analysis would have been conclusive to assess if the big amount of dysferlin transcript correspond to a big amount of mutant proteins. Because it could be even possible that the increased transcriptional activity does not correspond to increase quantity of protein.

5.1.2 D2 patient

D2 patient, a 48-year-old man, showed the first clinical symptoms at 24 years old with weakness and ipertrofic features of the pelvic and shoulder girdle muscles. D2 harboured a pathogenic homozygous missense mutation p.Asp1837Asn already reported in several papers, even in homozygous state[50,176,179,181,182] (Tab.9).

Reference	Gender	Age of onset	Phenotype	Genotype
Matsuda et al., 2001			MM	p.Asp1837Asn; p.Tyr522X
Cagliani et al., 2005	M	26	MM	p.Asp1837Asn; ND
Rosales et al., 2010		17	LGMD2B	p.Asp1837Asn; p.Asp1837Asn
Gallardo et al., 2011	M	26	LGMD2B	p.Asp1837Asn; p.Trp1968X
Vincent et al., 2016	F		LGMD2B	p.Asp1837Asn; p.Arg2042Cys

Tab.9: Reported cases in which p.Asp1837Asn has been detected.

Dysferlin immunofluorescence in D2 muscle sample showed a severe reduction of dysferlin at sarcolemma level (Fig.12). This is in agreement with Rosales' paper, describing a patient carrying the same variant in homozygous state. Rosales reported a male, in which the first symptoms appeared at 17 years old with weakness in distal lower extremity. The patient showed LGMD2B phenotype characterized by the absence of dysferlin protein both in western blot and immunofluorescence analysis. The onset and the phenotype are similar to our patient, in which the first symptoms appeared at 24 years old with LGMD2B phenotype and weakness in both distal and proximal muscles.

5.1.3 D3 patient

DYSF screening showed in D3 patient only one missense mutation, p.Ile2047Val, in exon 54 in heterozygous state. Genetic testing was extended to the daughter (D3_A) and the father (D3_B) and the variant was detected in heterozygous state in the asymptomatic daughter. The presence of large deletions on RNA level was evaluated in order to check a possible second mutation not detected by Sanger sequencing. The entire cDNA was analysed but no splicing defects were detected in both samples. Moreover, we evaluated the protein expression levels by western blot analysis on lymphocyte samples of D3 and D3_A. Both samples showed 2-fold increase in dysferlin protein expression compared to healthy control (Fig.17). While, dysferlin staining in D3 muscle showed a reduction of the protein at sarcolemma level with an increased level of dysferlin within the cytoplasm (Fig.16). Unfortunately, the muscle sample of the daughter was not available for the staining, thus it has not been possible to detect dysferlin localization. The discrepancy observed between western blot results and dysferlin staining has been already reported. In a patient with overexpressed amounts of protein in western blot analysis and absence of dysferlin in muscle staining[50]. Probably the mutant protein is unable to insert itself into the muscle membane. p.Ile2047Val mutation is located within the transmembrane domain (2045-2067) and it could be

possible that 50% of mutant proteins generated from c.6139A>G variant fails to insert at sarcolemma level. However, in order to confirm this hypothesis is essential to perform immunofluorescence analysis even in the asymptomatic daughter. This analysis could help us to clarify the functional consequences of p.Ile2047Val mutation. If the daughter will show the same reduction and localization of the protein in muscle sample of D3 this will assess the pathogenic role of p.Ile2047Val. So far, functional studies about DYSF mutation within TM domain are scarce, then it will be interesting to carry on further analysis.

5.2 CALPAIN PATIENTS

5.2.1 C1 family

C1 family included three siblings born to unrelated parents presented progressive muscle weakness with elevated serum CK. The proband (C1) and one of his sister (C1_A) showed severe limb girdle muscular weakness. Muscle biopsy showed both in C1 and C1_A dystrophic features including variability in fibers diameters and atrophic and hypertrophic fibers. In addition to that, C1 showed other dystrophic features not presented in C1_A as degenerating and/or regenerating fibers, fiber splitting and increase in adipose and fibrous tissue. The younger sister C1_B showed limb girdle muscle weakness although less severe than her siblings. Less severe phenotype of C1_B was examined both during neurological examination and in muscle biopsy analysis.

Since, clinical evaluation and muscle biopsy analysis were compatible with a LGMD2A phenotype, *CAPN3* screening was carried out. Two known pathogenic variants (c.1746-20C>G, p.Met252Arg) in heterozygous state were detected in C1 and C1_A, while just one variant (p.Met252Arg) was detected in C1_B.

The intronic variant c.1746-20C>G, presents in C1 and in C1_A, has been described in several reports. In 2005, it was described for the first time in 2 individuals in simple heterozygous state and other 2 cases in compound heterozygous state[183]. The following years it was described in *trans* with c.598_612del in a male patient characterised by the absence of calpain 3 at western blot analysis[167]. This variant was described even in other papers and its deleterious effect has been demonstrated by Nascimbeni *et al.*, 2010 at translational and transcriptional levels as well[184]. The intronic variant c.1746-20C>G leads to the activation of the cryptic acceptor splice sites carrying the retention of 19, 86, 122 and 305 bp of intron 13. The pathogenic effect of this variant was observed only when it is associated with another mutant allele and it always resulted in severely reduced protein quantity.

The missense variant p.Met252Arg found in siblings, has been already reported, more than 10 years ago by Fanin's research group[152]. This variant was described in simple heterozygous state in an asymptomatic male of 48 years old, characterized by high serum creatine kinase levels and a reduction of 50% of calpain 3 in western blot analysis.

Western blot analysis was performed by another laboratory only in C1 patient, showing 50% reduction of calpain 3 levels, likely due to the missense mutation p.Met252Arg, as reported in literature.

In silico analysis showed the missense variant as likely pathogenic since the amino acid substitution is located in a highly conserved region (Fig.20B). Moreover, 3D analysis highlighted that p.Met252 is located internally in α-helix within the 3D structure of calpain 3. The substitution of nonpolar amino acid as methionine for a charged amino acid as arginine provides increased polarity in this region due to the addition of a positive charge. The addition of a charge within a 3D structure could interfere with protein folding. Moreover, it is even possible that the positive charge of arginine is neutralized by the negative charge of aspartic acid at position 402. Since the distance between the two amino acids is 3.5 Å, distance compatible with electrostatic interaction and hydrogen bond. In this case, loss of a negative charge could interfere with the correct function of the protein. Then, in both scenarios p.Met252Arg seems to have a pathogenic effect in 3D structure of calpain 3 (Fig.21).

The intronic variant, c.1746-20C>G has been found associated with another mutant allele, leading to a severely reduced protein quantity[184]. Then, this could explain the severe phenotype shown by C1 and C1_A, carrying the intronic variant associated with the missense pathogenic mutation. However we detected in C1 and C1_A significantly increased levels of the transcript compared to healthy control and C1_B (Fig.22). Higher transcript levels were probably due to the intronic variant c.1746-20C>G, which leads to the activation of the cryptic acceptor splice sites generating four alternative transcripts with the retention of some nucleotides of intron 13[184]. This finding could explain why C1_B had different calpain transcript levels compared her sibling.

In conclusion, our study confirms the pathogenic role of c.1746-20C>G and p.Met252Arg in patients C1 and C1_A. As reported by Nasmibeni's paper, showing the pathogenic effect of the intronic variant when it is associated with a second mutation. However, even if C1 and C1_A are characterized by the same *CAPN3* genotype they do not show the same phenotype and the same age of onset of the disease. C1_A showed a less severe phenotype compared to C1. However, this is common in LGMD2A which is characterised by high phenotypic variability, considering the fact that the first case was reported in 1998.

While in C1_B only one mutation was detected and compared to the other cases reported in literature with only one mutation in *CAPN3* gene the presence of a second variant seems unlikely. In almost 22% of LGMD2A, only one variant is found but is not excluded the presence of another variant. However our case is different, then it could be possible that other factors are involved in the pathogenesis of LGMD2A.

5.2.2 C2 patient

In patient C2, three mutations were detected in *CAPN3* gene, already reported as pathogenic. Nonsense mutation c.967G>T (p.Glu323*) and deletion of one amino acid c.1401_1403delGGA (p.Glu467del) have been described in compound heterozygous state and in homozygous state[183]. While the missense variant c.2257G>A (p.Asp753Asn) was described several times in simple heterozygous state[12,151,152,185]. This variant seems to be associated with a moderate phenotype and slow clinical progression.

In our patient, segregation analysis showed that p.Glu467del and p.Asp753Asn variants were localized in the same allele while p.Glu323* in *trans*. Even if the presence of two pathogenic variants in *cis* seemed unlikely, previous studies have confirmed the pathogenic role of both of them[18312,151,152,185]. Both p.Glu467 and p.Asp753 are highly conserved among species. p.Glu467 is localized close to the region required for the interaction with titin, and is essential for the autolytic CAPN3 function. Moreover, the negative charge of the glutamic acid may interacts with the charged p.Lys575, which is one of the amino acids involved in titin interaction (Fig.25C). Then, the deletion of glutamic acid could modify the ability of CAPN3 to interact with titin leading to CAPN3 autolysis. While, p.Asp753Asn is localized in EF3-hand that contains a Ca^{2+} bound, amino acid modification within this domain could interfere during Ca^{2+} signalling.

In conclusion our studies are in agreement to what reported in literature, in which both mutations are described as pathogenic.

5.2.3 C3 patient

Patient C3 came to our attention at 16 years of age, manifesting hyposthenia especially of lower girdle muscles compared to upper girdle muscle and increased levels of serum CK (5018 U/L). *CAPN3* direct sequencing showed one heterozygous mutation in exon 21 c.2257G>A (p.Asp753Asn). This variant in simple heterozygous state, has been reported in several papers[12,151,152,185], as shown in table 10.

Reference	Gender	Age of onset	Phenotype	Clinical progression	WB analysis
Fanin *et al.*, 2004	F	3	Moderate	/	CAPN3 absent
Saenz *et al.*, 2005	/	40	Mild	Very benign evolution	/
Todorova *et al.*, 2007	F	25 at pregnancy	Mild	Mild disease manifestation	/
Guglieri *et al.*, 2008	F	37	Mild	Slow clinical progression	50%

Tab.10: Reported cases in which p.Asp753Asn has been detected in single heterozygous state.

p.Asp753Asn mutation seems not correlate with calpain 3 protein levels since patients with the same variant showed complete absence, 50% reduction and normal protein expression (our case). Then, it is possible that the differences in CAPN3 expression are due to the presence of a second mutation not detected. Even the age of onset seems not correlate with the mutation but the variant seems link to a mild disease manifestation and slow clinical progression. Interestingly, as the patient described by Tudorova the first symptoms of C3 got worse after pregnancy.

Moreover, in the present study, we evaluated not only CAPN3 expression but also DYSF expression. Both western bot and immunofluorescence analysis showed increased levels of dysferlin protein compared to healthy control. Further informations are required to understand the functional implications of that.

5.2.4 C4 patient

CAPN3 and *DYSF* screening in C4 patient showed only one novel missense mutation in, p.Val176Met, in heterozygous state in *CAPN3* gene. The presence of large deletions have been excluded with the sequencing of the whole *CAPN3* transcript. The missense variant is reported in the public database ExAC with a low allele frequency, but has not been found in previously reported calpin patients and in our in house healthy control. p.Val176 is highly conserved amino acid and all reliable software showed the pathogenic nature of the variant. 3D analysis led us to evaluate the effect of amino acid variation within the protein structure. p.Val176 is located internally in a β-sheet pointed towards the central α-helix. The substitution of valine versus methionine generates a steric hindrance, which may lead to structural reorganization of the protein. Moreover, western blot analysis in muscle and skin samples showed a protein reduction of 30%.

All these datas seem to confirm a possible pathogenic role of this variant. However, the genetic testing was extended to asymptomatic family members and p.Val176Met variant has been detected in mother, sibling, daughter and niece in heterozygous state as well. Then, among all family members carrying p.Val176Met variant, C4 patient is the only one with clinical LGMD

manifestation. This could be explained, as for the intronic variant c.1746-20C>G, that the pathogenic effect of p.Val176Met is observed in association with another mutant allele, that, in our case, has yet to be found.

5.2.5 C5 patient

CAPN3 screening in C5 patient showed one heterozygous novel missense mutation in exon 24, c.2458T>C (p.Tyr820His). The missense variant was not reported in the literature, in public databases and in house healthy controls. *In silico* analysis showed the pathogenic role of p.Tyr820His and multiple alignments highlighted the conservatively of tyrosine among species except for *Xenopus leavis*. Moreover, p.Tyr820His is localized within the EF5 of the protein and this domain is essential for the homodimerization of calpain 3. Therefore, amino acid substitution inside EF5 domain could interfere with the binding and consequently with the function of the protein. C5 patient showed among our LGMD2A patients the most severe clinical phenotype, with severe weakness in the shoulder and pelvic girdles. The immunohistochemical analysis highlighted dystrophic features as variation in fibers size, internal nuclei, necrosis and degenerating/regenerating fibers. We didn't perform further molecular analysis, since the unavailability of muscle biopsy to study the CAPN3 expression.

6. CONCLUSION

Limb-girdle muscular dystrophies (LGMDs) is the generic name for a group of neuromuscular disorders inherited either with dominant (LGMD1) and recessive manner (LGMD2). LGMDs are characterized by predominantly weakness and atrophy with the prevalent involvement of proximal muscles of the hip and shoulder girdles. The autosomal recessive forms are much more common compared to the autosomal dominant forms, that represent less than 10% of all LGMD cases, and limb-girdle muscular dystrophy type 2A (LGMD2A) is the most frequently diagnosed among LGMDs worldwide followed by limb-girdle muscular dystrophy type 2B (LGMD2B). Both diseases show high phenotypic variability (intra- and inter-familial variability) of symptoms and severity, even among patients with the same genotype. Therefore, it is difficult to establish genotype-phenotype correlations among these conditions.

Our study allowed to amplify the mutation spectrum in two genes involved in the most frequent forms of LGMD2 with the finding of two novel mutations in *CAPN3* gene and one novel mutation in *DYSF* gene. Our study pointed out the difficulty to establish genotype-phenotype correlations, as reported in literature, especially for LGMD2A. These patients showed high phenotypic variability, even intra-family variability in C1 family. The type of mutations is not correlate with the age of onset or the phenotype. We concluded that is really challenging to correlate a certain phenotype with certain mutations, especially for missense ones. Interestingly, the presence of three mutations in one patient (C2) does not mean an earlier age of onset or a more severe phenotype. C2 became symptomatic at 45 years old and the patient is still ambulant at 70 years old. Even the presence of a single heterozygous mutation does not correlate with a less severe phenotype or later onset. For example, in C3 patient only one missense mutation c.2257G>A (p.Asp753Asn) was detected and the patient showed the first clinical symptoms at 16 years old with a moderate phenotype. Moreover, our study is in line with the literature in which numerous LGMD2A patients are reported with just one heterozygous mutation. In all these patients is likely the presence of a second mutation not detected. However in one of our cases (C1_B patient) the presence of a second mutation is unlikely. This result points out the possibility of involvement of other factors in the pathogenesis of LGMD2A.

The LGMD2B study allowed us to correlate two homozygous mutations with a certain phenotype in accordance with literature. Moreover, we found a novel mutation in *DYSF* gene in heterozygous

state in D3 patient. This is the first time in which has been found an amino acid substitution in the transmembrane domain of dysferlin protein. The functional studies in p.Ile2047Val are not yet finished but this variant seems to correlate with an increased protein expression and a reduction of dysferlin at sarcolemma level. Immunofluorescence analysis highlighted a reduction at plasma membrane of dysferlin protein with an accumulation of dysferlin inside the cytoplasm. Dysferlin reduction at sarcolemma level could be due to the generation of a mutant protein that fails in the insertion at the sarcolemma. However, we need further analyses for clarifying the functional effect of p.Ile2047Val variant. It would be interesting to perform immunofluorescence analysis in the asymptomatic daughter carrying the same variant in heterozygous state as well.

7. REFERENCES

1. Dubowitz V. Muscular Dystrophy and Related Disorders. *Postgrad Med J.* **41(476),** 332–346 (1965).

2. Kaplan JC & Hamroun D. The 2016 version of the gene table of monogenic neuromuscular disorders (nuclear genome). *Neuromuscul Disord.* **25(12),** 991–1020 (2015).

3. Leyden. Klink der Ruchenmarks-Krankheiten. *Hirschwald* **2,** 447 (1876).

4. Meryon E. On Granular Degeneration of the Voluntary Muscles. *Med Chir Trans.* **49,** 45–50 (1866).

5. Duchenne De Boulogne GBA. Recherches sur la paralysie musculaire pseudo-hypertrophique ou paralysie myo-sclerosique. *Arch Gen Med* **11,** 5–25 (1886).

6. Reb WH. Dystrophia muscularis progressiva. Klinische und pathologisch- anatomische Studien. *Dtsch Z Nervenheilkd* (1891).

7. Koenig M, *et al*. Complete cloning of the duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* **50(3),** 509–517 (1987).

8. Hoffman EP, Brown RH Jr, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* **51,** 919–28 (1987).

9. Monies D, *et al.* A first-line diagnostic assay for limb-girdle muscular dystrophy and other myopathies. *Hum Genomics* **10(1),** 32 (2016).

10. Nigro V & Savarese M. Genetic basis of limb-girdle muscular dystrophies: the 2014 update. *Acta Myol.* **33(1),** 1–12 (2014).

11. Narayanaswami P. *et al.* Evidence-based guideline summary: Diagnosis and treatment of limb-girdle and distal dystrophies. *Neurology* **83,** 1453–1463 (2014).

12. Guglieri M, *et al.* Clinical, Molecular, and Protein Correlatio nsin a Large Sample of Genetically Diagnosed ItalianLimb Girdle Muscular Dystrophy Patients. *Hum Mutat.* **29(2),** 258–266 (2008).

13. Laval SH & Bushby KM. Limb-girdle muscular dystrophies - From genetics to molecular pathology. *Neuropathol. Appl. Neurobiol.* **30(2),** 91–105 (2004).

14. Jia Y, *et al*. Early embryonic lethality of mice with disrupted transcription cofactor PIMT/NCOA6IP/Tgs1 gene. **129(9-12),** 193–207 (2012).

15. Walton JN & Nattrass FJ. On the classification, natural history and treatment of the myopathies. *Brain* **77(2),** 169-231 (1954).

16. Murphy AP & Straub V. The Classification, Natural History and Treatment of the Limb Girdle Muscular Dystrophies. *J Neuromuscul Dis.* **2(s2),** S7–S19 (2015).

17. Bushby KM & Beckmann JS. The limb-girdle muscular dystrophies - Proposal for a new nomenclature. *Neuromuscul Disord.* **5(4),** 337–343 (1995).

18. Bushby KM & Gaedner-Medwin D. The clinical, genetic and dystrophin characteristics of Becker muscular dystrophy. I. Natural history. *J Neurol* **240(2),** 98–104 (1993).

19. Bushby KM. Diagnostic criteria for the limb-girdle muscular dystrophies: Report of the ENMC consortium on limbgirdle dystrophies. *Neuromuscul Disord.* **5(1),** 71–74 (1995).

20. Tasca G, *et al*. Limb-girdle muscular dystrophy with α-dystroglycan deficiency and mutations in the ISPD gene. *Neurology* **80(10),** 963–965 (2013).

21. Carss KJ, *et al.* Mutations in GDP-mannose pyrophosphorylase b cause congenital and limb-girdle muscular dystrophies associated with hypoglycosylation of α-dystroglycan. *Am J Hum Genet.* **93(1),** 29–41 (2013).

22. Torella A, *et al.* Next-Generation Sequencing Identifies Transportin 3 as the Causative Gene for LGMD1F. *PLoS One* **8(5),** e63536 (2013).

23. Vieira NM, *et al.* A defect in the RNA-processing protein HNRPDL causes limb-girdle muscular dystrophy 1G (LGMD1G). *Hum Mol Genet.* **23(15),** 4103–4110 (2014).

24. Sarparanta J, *et al.* Mutations affecting the cytoplasmic functions of the co- chaperone DNAJB6 cause limb-girdle muscular dystrophy. *Nat Genet* **44(4),** 450-455 (2012).

25. Pantoja-Melendez CA, *et al*. Epidemiological and molecular characterization of a mexican population isolate with high prevalence of limb-girdle muscular dystrophy type 2A due to a novel Calpain-3 mutation. *PLoS One* **12(1),** e0170280 (2017).

26. Reddy HM, *et al*. The sensitivity of exome sequencing in identifying pathogenic mutations for LGMD in the United States. *J Hum Genet* **62(2),** 243–252 (2017).

27. Magri F, *et al*. The Italian LGMD registry: relative frequency, clinical features, and differential diagnosis. *Muscle Nerve* **55(1),** 55–68 (2017).

28. Ohno S, *et al*. Four genes for the calpain family locate on four distinct human chromosomes. *Cytogenet Cell Genet* **53(4),** 225–229 (1990).

29. Beckmann JS, *et al*. A gene for limb-girdle muscular dystrophy maps to chromosome 15 by linkage. *C R Acad Sci III* **312(4),** 141–148 (1991).

30. Passos-bueno MR *et al.* Evidence of genetic heterogeneity in the autosomal recessive adult forms of limb-girdle muscular dystrophy following linkage analysis with 15q probes in Brazilian families. *J Med Genet* **30(5),** 385–387 (1993).

31. Schessl J *et al.* Phenotypic variability in siblings with calpainopathy (LGMD2A). *Acta Myol.* **27,** 54–58 (2008).

32. Zatz M & Starling A. Calpains and Disease. *N Engl J Med* **352(23),** 2413–2423 (2005).

33. Angelini C, *et al*. The clinical course of calpainopathy (LGMD2A) and dysferlinopathy

(LGMD2B). *Neurol Res* **32(1),** 41–46 (2010).

34. Bartoli M, *et al.* Safety and Efficacy of AAV-Mediated Calpain 3 Gene Transfer in a Mouse Model of Limb-Girdle Muscular Dystrophy Type 2A. *Mol. Ther.* **13(2),** 250–259 (2006).

35. Selvaraj S, *et al.* Gene Correction of LGMD2A Patient-Specific iPS Cells for Targeted Autologous Cell Therapy. *Mol Ther.* **24(S1),** S125–S126 (2016).

36. Bansal D, *et al*. Defective membrane repair in dysferlin-deficient muscular dystrophy. *Nature* **423(6936),** 168–172 (2003).

37. Campbell KP. Three muscular dystrophies: Loss of cytoskeleton-extracellular matrix linkage. *Cell* **80(5),** 675–679 (1995).

38. Illa I, *et al*. Distal anterior compartment myopathy: a dysferlin mutation causing a new muscular dystrophy phenotype. *Ann Neurol.* **49(1),** 130–4 (2001).

39. Illarioshkin SN, *et al.* Clinical and molecular analysis of a large family with three distinct phenotypes of progressive muscular dystrophy. *Brain* **119(6),** 1895–1909 (1996).

40. Paradas C, *et al.* A new phenotype of dysferlinopathy with congenital onset. *Neuromuscul Disord.* **19(1),** 21–25 (2009).

41. Nguyen K, *et al*. Phenotypic Study in 40 Patients With Dysferlin Gene Mutations. *Arch Neurol.* **64(8),** 1176–1182 (2007).

42. Bashir R, *et al*. A gene for autosomal recessive limb-girdle muscular dystrophy maps to chromosome 2p. *Hum Mol Genet.* **3(3),** 445–447 (1994).

43. Passos-Bueno MR, *et al*. Confirmation of the 2p locus for the mild autosomal recessive limb-girdle muscular dystrophy gene (LGMD2B) in three families allows refinement of the candidate region. *Genomics* **27(1),** 192–195 (1995).

44. Bejaoui K, *et al*. Linkage of Miyoshi myopathy (distal autosomal recessive muscular dystrophy) locus to chromosome 2p12-14. *Neurology* **45(4),** 768–772 (1995).

45. Liu J, *et al.* Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nat Genet.* **20(1),** 31–36 (1998).

46. Bushby KM. Making sense of the limb-girdle muscular dystrophies. *Brain* **122(8),** 1403–1420 (1999).

47. Díaz J, *et al.* Broadening the imaging phenotype of dysferlinopathy at different disease stages. *Muscle and Nerve* **54(2),** 203–210 (2016).

48. Woudt L, *et al.* Toward an objective measure of functional disability in dysferlinopathy. *Muscle and Nerve* **53(1),** 49–57 (2016).

49. Takahashi T, *et al.* Clinical features and a mutation with late onset of limb girdle muscular dystrophy 2B. *J Neurol Neurosurg Psychiatry* **84(4),** 433–440 (2013).

50. Rosales XQ, *et al.* Novel diagnostic features of dysferlinopathies. *Muscle Nerve* **42(1),** 14–21

(2010).

51. Guglieri M, *et al*. Limb–girdle muscular dystrophies. *Curr Opin Neurol.* **21(5),** 576–584 (2008).

52. Ueyama H, *et al.* Clinical heterogeneity in dysferlinopathy. *Intern Med.* **41(7),** 532–536 (2002).

53. Cárdenas AM, *et al*. Dysferlin function in skeletal muscle: Possible pathological mechanisms and therapeutical targets in dysferlinopathies. *Exp Neurol.* **283,** 246–254 (2016).

54. Merrick D, *et al.* Muscular dystrophy begins early in embryonic development deriving from stem cell loss and disrupted skeletal muscle formation. *Dis Model Mech* **2(7-8),** 374–388 (2009).

55. Pryadkina M, *et al.* A comparison of AAV strategies distinguishes overlapping vectors for efficient systemic delivery of the 6.2 kb Dysferlin coding sequence. *Mol Ther Methods Clin Dev.* **2,** 15009 (2015).

56. Gushchina LV, *et al.* Treatment with Recombinant Human MG53 Protein Increases Membrane Integrity in a Mouse Model of Limb Girdle Muscular Dystrophy 2B. *Mol Ther*. **25(10),** 2360-2371 (2017).

57. Escobar H, *et al*. Full-length Dysferlin Transfer by the Hyperactive Sleeping Beauty Transposase Restores Dysferlin-deficient Muscle. *Mol Ther Nucleic Acids* **5,** e277 (2016).

58. Sondergaard PC, *et al.* AAV.Dysferlin Overlap Vectors Restore Function in Dysferlinopathy Animal Models. *Ann Clin Transl Neurol* **2(3),** 256–270 (2015).

59. Herasse M, *et al.* Expression and Functional Characteristics of Calpain 3 Isoforms Generated through Tissue-Specific Transcriptional and Posttranscriptional Events. *Mol Cell Biol.* **19(6),** 4047–4055 (1999).

60. Blázquez L, *et al.* Characterization of novel CAPN3 isoforms in white blood cells: An alternative approach for limb-girdle muscular dystrophy 2A diagnosis. *Neurogenetics* **9(3),** 173–182 (2008).

61. Anderson LV, *et al*. Dysferlin is a plasma membrane protein and is expressed early in human development. *Hum Mol Genet.* **8(5),** 855–861 (1999).

62. Pramono ZA, *et al.* Identification and characterisation of human dysferlin transcript variants: Implications for dysferlin mutational screening and isoforms. *Hum Genet.* **125(4),** 413–420 (2009).

63. Guroff G. A Neutral, calcium-activated proteinase from the soluble fraction of Rat Brain. *J Biol Chem.* **239(1),** 149–155 (1964).

64. Murachi T, *et al*. Intracellular Ca2+-dependent protease (CALPAIN) and its high-molecular-weight endogenous inhibitor (CALPASTATIN). *Adv Enzyme Regul.* **19,** 407–424 (1981).

65. Suzuki K. Nomenclature of calcium dependent proteinase. *Biomed Biochim Acta* **50(4-6),**

483–484 (1991).

66. Sorimachi H, Mamitsuka H, Ono Y. Understanding the substrate specificity of conventional calpains. *Biol Chem.* **393(9),** 853–871 (2012).

67. Goll DE, *et al*. The Calpain System. *Physiol Rev* **83(3),** 731–801 (2003).

68. Suzuki K, *et al*. Structure, activation, and biology of calpain. *Diabetes* **53(S1),** S12-8 (2004).

69. Branca D. Calpain-related diseases. *Biochem Biophys Res Commun.* **322(4),** 1098–1104 (2004).

70. Ohno S, *et al*. Evolutionary origin of a calcium-dependent protease by fusion of genes for a thiol protease and a calcium-binding protein? *Nature* **312(5994),** 566–570 (1984).

71. Maki M, Narayana SV, Hitomi K. A growing family of the Ca2+-binding proteins with five EF-hand motifs. *Biochem J.* **328(2),** 718–720 (1997).

72. Sorimachi H, Hata S, Ono Y. Impact of genetic insights into calpain biology. *Biochem J.* **150(1),** 23–37 (2011).

73. Campbell RL & Davies PL. Structure–function relationships in calpains. *Biochem. J* **447(3),** 335–351 (2012).

74. Ono Y, *et al*. An eccentric calpain, CAPN3/p94/calpain-3. *Biochimie.* **122,** 169–187 (2016).

75. Sorimachi H, Ishiura S, Suzuki K. Structure and physiological function of calpains. *Biochem J.* **328(3),** 721–732 (1997).

76. Khorchid A & Ikura M. How calpain is activated by calcium. *Nat Struct Biol.* **9,** 239–241 (2002).

77. Todd B, *et al.* A Structural Model for the Inhibition of Calpain by Calpastatin : Crystal Structures of the Native Domain VI of Calpain and its Complexes with Calpastatin Peptide and a Small Molecule Inhibitor. *J Mol Biol.* **2836(1),** 131–146 (2003).

78. Sorimachi H & Suzuki K. The structure of calpain. *J Biochem.* **129(5),** 653–64 (2001).

79. Lebart MC & Benyamin Y. Calpain involvement in the remodeling of cytoskeletal anchorage complexes. *FEBS J.* **273(15),** 3415–3426 (2006).

80. Franco SJ & Huttenlocher A. Regulating cell migration: calpains make the cut. *J Cell Sci.* **118(17),** 3829–3838 (2005).

81. Moldoveanu T, *et al*. A Ca^{2+}Switch Aligns the Active Site of Calpainl. *Cell* **108,** 649–660 (2002).

82. Reverter D, *et al.* Structural basis for possible calcium-induced activation mechanisms of calpains. *Biol Chem.* **382(5),** 753–766 (2001).

83. Tompa P, *et al*. Domain III of calpain is a ca2+-regulated phospholipid-binding domain. *Biochem Biophys Res Commun.* **280(5),** 1333–1339 (2001).

84. Suzuki K, *et al*. Calpain: novel family members, activation, and physiologic function. *Biol Chem Hoppe Seyler.* **376(9),** 523–9 (1995).

85. Sorimachi H, *et al*. Molecular cloning of a novel mammalian calcium-dependent protease distinct from both m- and mu-types. Specific expression of the mRNA in skeletal muscle. *J. Biol. Chem.* **264(33),** 20106–20111 (1989).

86. Sorimachi H, *et al.* Muscle-specific calpain, p94, is degraded by autolysis immediately after translation, resulting in disappearance from muscle. *J Biol Chem.* **268(14),** 10593–10605 (1993).

87. Kinbara K, *et al*. Skeletal Muscle-specific Calpain, p94. Structure and physiological function. *Biochem Pharmacol.* **56,** 415–420 (1998).

88. Jones SW, *et al*. Fibre type-specific expression of p94, a skeletal muscle-specific calpain. *J Muscle Res Cell Motil.* **20(4),** 417–424 (1999).

89. De Tullio R, *et al*. Characterization of a new p94-like calpain form in human lymphocytes. *Biochem J.* **375(3),** 689–696 (2003).

90. Fougerousse F, *et al.* Calpain3 expression during human cardiogenesis. *Neuromuscul Disord* **10(4-5),** 251–256 (2000).

91. Ono Y, *et al.* Possible Regulation of the Conventional Calpain System by Skeletal Muscle-specific Calpain , p94 / Calpain 3. *J Biol Chem.* **279(4),** 2761–2771 (2004).

92. Ono Y, *et al.* Skeletal muscle-specific calpain is an intracellular Na+- dependent protease. *J Biol Chem.* **285(30),** 22986–22998 (2010).

93. Taveau M, *et al.* Calpain 3 Is Activated through Autolysis within the Active Site and Lyses Sarcomeric and Sarcolemmal Components. *Mol Cell Biol.* **23(24),** 9127–9135 (2003).

94. Ono Y, *et al.* Suppressed Disassembly of Autolyzing p94 / CAPN3 by N2A Connectin / Titin in a Genetic Reporter System. *J Biol Chem.* **281(27),** 18519–18531 (2006).

95. Ravulapalli R, *et al*. Distinguishing between calpain heterodimerization and homodimerization. *FEBS J.* **276(4),** 973–982 (2009).

96. Ravulapalli R, *et al*. Homodimerization of calpain 3 penta-EF-hand domain. *Biochem. J.* **388(2),** 585–91 (2005).

97. Imajoh S, Kawasaki H, Suzuki K. The COOH-terminal E-F hand structure of calcium-activated neutral protease (CANP) is important for the association of subunits and resulting proteolytic activity. *J Biochem.* **101(2),** 447–452 (1987).

98. Ono Y, *et al.* Suppressed disassembly of autolyzing p94/CAPN3 by N2A connectin/titin in a genetic reporter system. *J Biol Chem.* **281(27),** 18519–18531 (2006).

99. Ono Y, *et al.* PLEIAD/SIMC1/C5orf25, a novel autolysis regulator for a skeletal-muscle-specific calpain, CAPN3, scaffolds a CAPN3 substrate, CTBP1. *J Mol Biol.* **425(16),** 2955–2972 (2013).

100. Ermolova N, Kramerova I, Spencer MJ. Autolytic activation of calpain 3 proteinase is facilitated by calmodulin protein. *J Biol Chem.* **290(2),** 996–1004 (2015).

101. Beckmann JS & Spencer M. Calpain 3, the 'gatekeeper' of proper sarcomere assembly, turnover and maintenance. *Neuromuscul Disord* **18(12),** 913–921 (2008).

102. Kramerova I, *et al.* Novel role of calpain-3 in the triad-associated protein complex regulating calcium release in skeletal muscle. *Hum Mol Genet.* **17(21),** 3271–3280 (2008).

103. Ojima K, *et al.* Myogenic stage, sarcomere length, and protease activity modulate localization of muscle-specific calpain. *J Biol Chem.* **282(19),** 14493–14504 (2007).

104. Wette SG, *et al*. Characterization of Muscle Ankyrin Repeat Proteins in human skeletal muscle. *Am J Physiol Cell Physiol.* **313(3),** C328-C339 (2017).

105. Hayashi C, *et al.* Multiple molecular interactions implicate the connectin/titin N2A region as a modulating scaffold for p94/calpain 3 activity in skeletal muscle. *J Biol Chem.* **283(21),** 14801–14814 (2008).

106. Ojima K, *et al.* Dynamic distribution of muscle-specific calpain in mice has a key role in physical-stress adaptation and is impaired in muscular dystrophy. *J Clin Invest.* **120(8),** 2672–2683 (2010).

107. Kramerova I, *et al*. Calpain 3 participates in sarcomere remodeling by acting upstream of the ubiquitin - Proteasome pathway. *Hum Mol Genet.* **16(8),** 1006 (2007).

108. Ojima K, *et al*. Non-proteolytic functions of calpain-3 in sarcoplasmic reticulum in skeletal muscles. *J Mol Biol.* **407(3),** 439–449 (2011).

109. Stuelsatz P, *et al.* Down-regulation of MyoD by calpain 3 promotes generation of reserve cells in C2C12 myoblasts. *J Biol Chem.* **285(17),** 12670–12683 (2010).

110. Cooper RN, *et al.* In vivo satellite cell activation via Myf5 and MyoD in regenerating mouse skeletal muscle. *J Cell Sci.* **112(17),** 2895–2901 (1999).

111. Chargé SB & Rudnicki MA. Cellular and Molecular Regulation of Muscle Regeneration. *Physiol Rev* **84(1),** 209–238 (2004).

112. Tapscott SJ. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132(12),** 2685–2695 (2005).

113. Glover L & Brown RH. Dysferlin in membrane trafficking and patch repair. *Traffic* **8(7),** 785–794 (2007).

114. Huang Y, *et al.* Protein studies in dysferlinopathy patients using llama-derived antibody fragments selected by phage display. *Eur J Hum Genet.* **13(6),** 721–730 (2005).

115. Lek A, *et al.* Phylogenetic analysis of ferlin genes reveals ancient eukaryotic origins. *BMC Evol Biol.* **10,** 231 (2010).

116. Bashir R, *et al.* A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is

mutated in limb-girdle muscular dystrophy type 2B. *Nat Genet.* **20(1),** 37–42 (1998).

117. Yasunaga S, *et al.* A mutation in OTOF, encoding otoferlin, a FER-1-like protein, causes DFNB9, a nonsyndromic form of deafness. *Nat Genet* **21(4),** 363–369 (1999).

118. Redpath GM, *et al*. Ferlins Show Tissue-Specific Expression and Segregate as Plasma Membrane/Late Endosomal or Trans-Golgi/Recycling Ferlins. *Traffic* **17(3),** 245–266 (2016).

119. Doherty KR. *et al.* Normal myoblast fusion requires myoferlin. *Development* **132(24),** 5565–5575 (2005).

120. Leung C, *et al*. Expression of myoferlin in human and murine carcinoma tumors: Role in membrane repair, cell proliferation, and tumorigenesis. *Am J Pathol.* **182(5),** 1900–1909 (2013).

121. Han R & Campbell KP. Dysferlin and muscle membrane repair. *Curr Opin Cell Biol.* **19(4),** 409–416 (2007).

122. Chiu YH, *et al.* Attenuated muscle regeneration is a key factor in dysferlin-deficient muscular dystrophy. *Hum Mol Genet.* **18(11),** 1976–1989 (2009).

123. Südhof TC & Rizo J. Synaptotagmins: C2-domain proteins that regulate membrane traffic. *Neuron* **17(3),** 379–388 (1996).

124. Sutton RB, *et al*. Structure of the first C2 domain of synaptotagmin I: A novel Ca2+/phospholipid-binding fold. *Cell* **80(6),** 929–938 (1995).

125. Middel V, *et al.* Dysferlin-mediated phosphatidylserine sorting engages macrophages in sarcolemma repair. *Nat Commun.* **7,** 12875 (2016).

126. Therrien C, *et al*. Characterization of lipid binding specificities of dysferlin C2 domains reveals novel interactions with phosphoinositides. *Biochemistry* **48(11),** 2377–2384 (2009).

127. Davis DB, *et al*. Calcium-sensitive phospholipid binding properties of normal and mutant ferlin C2 domains. *J Biol Chem.* **277(25),** 22883–22888 (2002).

128. Ampong BN, *et al*. Intracellular localization of dysferlin and its association with the dihydropyridine receptor. *Acta Myol.* **24(2),** 134–144 (2005).

129. Klinge L, *et al.* From T-tubule to sarcolemma: damage-induced dysferlin translocation in early myogenesis. *FASEB J.* **21(8),** 1768–1776 (2007).

130. Huang Y, *et al.* AHNAK, a novel component of the dysferlin protein complex, redistributes to the cytoplasm with dysferlin during skeletal muscle regeneration. *FASEB J.* **21(3),** 732–42 (2007).

131. Hofhuis J, *et al.* Dysferlin mediates membrane tubulation and links T-tubule biogenesis to muscular dystrophy. *J Cell Sci.* **130(5),** 841–852 (2017).

132. Lennon NJ, *et al.* Dysferlin Interacts with Annexins A1 and A2 and Mediates Sarcolemmal

Wound-healing. *J Biol Chem.* **278(50),** 50466–50473 (2003).

133. Bansal D & Campbell KP. Dysferlin and the plasma membrane repair in muscular dystrophy. *Trends Cell Biol.* **14(4),** 206–213 (2004).

134. Azakir BA, *et al*. Dysferlin interacts with tubulin and microtubules in mouse skeletal muscle. *PLoS One* **5(4),** e10122 (2010).

135. Sharma A, *et al.* A new role for the muscle repair protein dysferlin in endothelial cell adhesion and angiogenesis. *Arterioscler Thromb Vasc Biol.* **30(11),** 2196–2204 (2010).

136. Di Fulvio S, *et al*. Dysferlin interacts with histone deacetylase 6 and increases alpha-tubulin acetylation. *PLoS One* **6(12),** e28563 (2011).

137. de Morree A, *et al.* Dysferlin regulates cell adhesion in human monocytes. *J Biol Chem.* **288(20),** 14147–14157 (2013).

138. Matsuda C, *et al*. The sarcolemmal proteins dysferlin and caveolin-3 interact in skeletal muscle. *Hum Mol Genet.* **10(17),** 1761–1766 (2001).

139. Matsuda C, *et al*. Dysferlin interacts with affixin (beta-parvin) at the sarcolemma. *J Neuropathol Exp Neurol* **64(4),** 334–340 (2005).

140. Anderson LV, *et al*. Secondary reduction in calpain 3 expression in patients with limb girdle muscular dystrophy type 2B and Miyoshi myopathy (primary dysferlinopathies). *Neuromuscul Disord.* **10(8),** 553–559 (2000).

141. De Luna N, *et al.* Absence of dysferlin alters myogenin expression and delays human muscle differentiation ‘in vitro’. *J Biol. Chem.* **281(25),** 17092–17098 (2006).

142. Cooper ST & McNeil PL. Membrane Repair: Mechanisms and Pathophysiology. *Physiol Rev.* **95(4),** 1205–1240 (2015).

143. Terasaki M, Miyake K, McNeil PL. Large plasma membrane disruptions are rapidly resealed by Ca2+- dependent vesicle-vesicle fusion events. *J Cell Biol.* **139(1),** 63–74 (1997).

144. Cai C, *et al.* MG53 nucleates assembly of cell membrane repair machinery Chuanxi. **11(1),** 56–64 (2009).

145. Cai C, *et al.* Membrane repair defects in muscular dystrophy are linked to altered interaction between MG53, caveolin-3, and dysferlin. *J Biol Chem.* **284(23),** 15894–15902 (2009).

146. Steinhardt RA, Bi G, Alderton JM. Cell membrane resealing by a vesicular mechanism similar to neurotransmitter release. *Science (80-.).* **263(5145),** 390–393 (1994).

147. Miyake K & Mcneil PL. Vesicle Accumulation and Exocytosis at Sites of Plasma Membrane Disruption. *J Cell Biol.* **131,** 1737–1745 (1995).

148. Matsuda C, *et al*. The C2A domain in dysferlin is important for association with MG53 (TRIM72). *PLoS Curr.* **4,** doi: 10.1371/5035add8caff4 (2012).

149. Lek A, *et al.* Calpains, Cleaved Mini-DysferlinC72, and L-Type Channels Underpin

Calcium-Dependent Muscle Membrane Repair. *J Neurosci.* **33(12),** 5085–5094 (2013).

150. Redpath GM, *et al.* Calpain cleavage within dysferlin exon 40a releases a synaptotagmin-like module for membrane repair. *Mol Biol Cell* **25(19),** 3037–3048 (2014).

151. Sáenz A, *et al.* LGMD2A: Genotype-phenotype correlations based on a large mutational survey on the calpain 3 gene. *Brain* **128(4),** 732–742 (2005).

152. Fanin M, *et al.* Molecular diagnosis in LGMD2A: Mutation analysis of protein testing? *Hum. Mutat.* **24(1),** 52–62 (2004).

153. Pénisson-Besnier I, *et al*. Pseudometabolic expression and phenotypic variability of calpain deficiency in two siblings. *Muscle Nerve* **21(8),** 1078–1080 (1998).

154. Chae J, *et al*. Calpain 3 gene mutations : genetic and clinico-pathologic findings in limb-girdle muscular dystrophy. *Neuromuscul Disord.* **11(6-7),** 547–555 (2001).

155. Weiler T, *et al*. Identical mutation in patients with limb girdle muscular dystrophy type 2B or Miyoshi myopathy suggests a role for modifier gene(s). *Hum Mol Genet.* **8(5),** 871–7 (1999).

156. Illarioshkin SN, *et al*. Identical dysferlin mutation in limb-girdle muscular dystrophy type 2B and distal myopathy. *Neurology* **55(12),** 1931–1933 (2000).

157. Klinge L, *et al.* New aspects on patients affected by dysferlin deficient muscular dystrophy. *J Neurol Neurosurg Psychiatry* **81(9),** 946–953 (2010).

158. Paradas C, *et al*. Redefining dysferlinopathy phenotypes based on clinical findings and muscle imaging studies. *Neurology* **75(4),** 316–323 (2010).

159. Mercuri E, *et al.* Muscle MRI in Inherited Neuromuscular Disorders : Past , Present , and Future. *J Magn Reson Imaging.* **25(2),** 433–440 (2007).

160. Norwood F, *et al*. EFNS guideline on diagnosis and management of limb girdle muscular dystrophies. *Eur J Neurol.* **14(12),** 1305–1312 (2007).

161. Mitsuhashi S & Kang PB. Update on the Genetics of Limb Girdle Muscular Dystrophy. *Semin Pediatr Neurol.* **19(4),** 211–218 (2012).

162. Inashkina I, *et al.* Robust genotyping tool for autosomal recessive type of limb-girdle muscular dystrophies. *BMC Musculoskelet Disord.* **17,** 200 (2016).

163. Flanigan KM. The Muscular Dystrophies. *Semin Neurol* **32(3),** 255–263 (2012).

164. Nigro V & Savarese M. Next-generation sequencing approaches for the diagnosis of skeletal muscle disorders. *Curr Opin Neurol.* **29(5),** 621–627 (2016).

165. Urtasun M, *et al*. Limb-girdle muscular dystrophy in Guipúzcoa (Basque Country, Spain). *Brain* **121(9),** 1735–1747 (1998).

166. Vainzof M, *et al*. The effect of calpain 3 deficiency on the pattern of muscle degeneration in the earliest stages of LGMD2A. *J Clin Pathol.* **56(8),** 624–626 (2003).

167. Hermanova M, *et al*. Analysis of histopathologic and molecular pathologic findings in Czech LGMD2A patients. *Muscle Nerve* **33(3),** 424–432 (2006).

168. Keira Y, *et al.* Characterization of lobulated fibers in limb girdle muscular dystrophy type 2A by gene expression profiling. *Neurosci Res.* **57(4),** 513–521 (2007).

169. Haravuori H, *et al.* Secondary calpain3 deficiency in 2q-linked muscular dystrophy Titin is the candidate gene. *Neurology* **56(7),** 869–877 (2001).

170. Fanin M, *et al.* Loss of calpain-3 autocatalytic activity in LGMD2A patients with normal protein expression. *Am J Pathol.* **163(5),** 1929–36 (2003).

171. Groen EJ, *et al.* Analysis of the UK diagnostic strategy for limb girdle muscular dystrophy 2A. *Brain* **130(12),** 3237–3249 (2007).

172. Vissing J, *et al.* A heterozygous 21-bp deletion in CAPN3 causes dominantly inherited limb girdle muscular dystrophy. *Brain* **139(8),** 2154–2163 (2016).

173. Sáenz A & De Munain AL. Dominant LGMD2A: Alternative diagnosis or hidden digenism? *Brain* **140(2),** doi: 10.1093/brain/aww281 (2017).

174. Jaka O, *et al*. Entire CAPN3 gene deletion in a patient with limb-girdle muscular dystrophy type 2A. *Muscle and Nerve* **50(3),** 448–453 (2014).

175. Piccolo F, *et al*. Intracellular accumulation and reduced sarcolemmal expression of dysferlin in limb-girdle muscular dystrophies. *Ann Neurol.* **48(6),** 902–912 (2000).

176. Gallardo E, *et al.* Comparison of dysferlin expression in human skeletal muscle with that in monocytes for the diagnosis of dysferlin myopathy. *PLoS One* **6(12),** e29061 (2011).

177. Ho M, *et al*. A novel, blood-based diagnostic assay for limb girdle muscular dystrophy 2B and Miyoshi myopathy. *Ann Neurol.* **51(1),** 129–133 (2002).

178. Fanin M & Angelini C. Muscle pathology in dysferlin deficiency. *Neuropathol Appl Neurobiol.* **28(6),** 461–470 (2002).

179. Cagliani R, *et al.* Mutation finding in patients with dysferlin deficiency and role of the dysferlin interacting proteins annexin A1 and A2 in muscular dystrophies. *Hum Mutat.* **26(3),** 283 (2005).

180. Therrien C, *et al*. Mutation impact on dysferlin inferred from database analysis and computer-based structural predictions. *J Neurol Sci.* **250(1-2),** 71–78 (2006).

181. Vincent AE, *et al.* Dysferlin mutations and mitochondrial dysfunction. *Neuromuscul Disord.* **26(11),** 782–788 (2016).

182. Matsuda C, *et al.* The sarcolemmal proteins dysferlin and caveolin-3 interact in skeletal muscle. *Hum Mol Genet.* **10(17),** 1761–1766 (2001).

183. Piluso G, *et al.* Extensive scanning of the calpain-3 gene broadens the spectrum of LGMD2A phenotypes. *J Med Genet.* **42(9),** 686–693 (2005).

184. Nascimbeni AC, *et al*. Transcriptional and translational effects of intronic CAPN3 gene mutations. *Hum Mutat.* **31(9),** 1658–1669 (2010).

185. Todorova A, *et al.* A large deletion and novel point mutations in the calpain 3 gene (CAPN3) in Bulgarian LGMD2A patients. *Neurogenetics.* **8(3),** 225–229 (2007).