
Kemal Oral Yariz

University of Miami, koyariz@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation

https://scholarlyrepository.miami.edu/oa_dissertations/115

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.

By

Kemal O. Yariz

A DISSERTATION

Submitted to the Faculty of the University of Miami in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Coral Gables, Florida

June 2008

Kemal O.Yariz

Approved:

Dr. Richard S. Myers
Lecturer
Department of Biochemistry and Molecular Biology

Dr. Terri A. Scandura
Dean of the Graduate School

Dr. Karoline Briegel
Assistant Professor
Department of Biochemistry and Molecular Biology

Dr. Frans Huijing
Professor
Department of Biochemistry and Molecular Biology

Dr. Louis J. Elsas
Professor Emeritus
Department of Pediatrics

Dr. Lisa L. Baumbach-Reardon
Associate Research Professor
Department of Pediatrics, Neurology and Biochemistry

Dr. Stephan Zuchner
Associate Professor in the Department of Human Genetics and Genomic Medicine
Spinal Muscular Atrophy (SMA) is an autosomal recessive motor neuron disease. SMA is associated with homozygous mutations in the Survival of Motor Neuron gene I (SMN1). SMN protein does not appear to exist in cells in isolation but associates with several proteins to form a large multi-protein complex. The functions of SMN complex include assembly, metabolism and transport of diverse classes of ribonucleoproteins.

X-Linked Spinal Muscular Atrophy is a rare congenital disorder characterized by multiple joint contractures. It is associated with hypotonia, areflexia, chest deformities and congenital joint contractures. A candidate interval was defined for XL-SMA in Xp11.3-Xq11.2 in 1995.

The purpose of this study was to refine the XL-SMA gene region and discover the XL-SMA gene. In addition to that, the gene product was investigated to delineate the genotype-phenotype correlation.

My studies were focused on single nucleotide polymorphism (SNP) analysis. The candidate gene interval was refined by studying 14 SNPs in the three largest families. This analysis revealed a recombination event which
allowed elimination of the NDP gene. Significantly positive LOD scores were obtained from these SNP studies.

The exons and exon-intron boundaries of 12 genes were screened. No mutations were found in these genes in affected male samples. In late 2006, \textit{UBE1} (Ubiquitin activating enzyme 1) was discovered as the XL-SMA gene. UBE1 protein is responsible for the first step of ubiquitination of proteins in a cell. To investigate a possible common molecular mechanism between SMA and XL-SMA, proteins in the SMN Complex in XL-SMA patient cell lines were studied. SMN and Gemin3 protein levels were found to be consistently lower in XL-SMA patient cell lines (lymphoblasts) compared to healthy cell line. These results imply that there may be a common disease mechanism. To understand if the SMN and Gemin3 RNA levels decrease, RNA expression studies were performed. These studies confirmed that there is no difference of RNA expression of SMN and Gemin3 in XL-SMA cell lines when compared to healthy cell lines.

As for UBE1, the same experimental procedure for SMN Complex proteins were repeated with antibodies to UBE1 to determine if there is any decline of UBE1 protein levels in XL-SMA patient cell lines compared to a healthy cell line. There was a decline in protein levels of UBE1 in XL-SMA patients.

Two possible models are proposed for a molecular mechanism in XL-SMA: 1) UBE1 involves in degradation of a protein which downregulates SMN Complex (or a protein which stabilizes SMN Complex). When \textit{UBE1} is mutated, the protein in question is not degraded and this results in excess downregulation
of SMN Complex (maybe via a pathway involving SMN-Gemin3 interaction). 2) UBE1 and UBA6 interact with the proteins of SMN Complex as they monoubiquinate them for different cellular processes. When \textit{UBE1} is mutated, UBA6 cannot compensate the deficiency of UBE1, which in turn disrupts normal cellular RNA metabolism required for motor neuron development and survival.
Acknowledgements

"Know what is in front of your face, and what is hidden from you will be disclosed to you. For there is nothing hidden that will not be revealed. "
The Gospel of Thomas, Verse 5

I have to thank everybody who made me who I am today. Nobody will read this part (if not any part of this thesis) anyway and this is basically intellectual masturbation, but you gotta do what you gotta do. Here’s the list:

My family and friends, you know who you are.

Some of my teachers like Klaus Wirth and Manfred Liedtke, who taught me that there are more important things than a classroom or lab (but have I followed those ideas? No, of course not, so I am writing this).

All the musicians who created the soundtrack of my life. From Prince to Boney James, from Megadeth to Queensryche, from Army of Lovers to Dany Brilliant, from Teddy Riley to Victor Wooten, you guys rock!

Mr. Hitchcock, Mr. Kubrick, Mr. Scorsese, Mr. Allen and Mr. Scott. Thank you for all the life changing experiences.

Evan Stone and David Stanley for being so down to earth. ADT for the coolest party I have ever attended to. Devinn, Janine and Asia for their beauty and honesty.

UCLA Extension and WGA for giving me hope for the future.

Larry David, Jon Stewart & Stephen Colbert for making me laugh every single day. You guys are geniuses!

Hef, you are the man!

And finally, First Cause*, I hope you forgive all of us someday.

*Every finite and contingent being has a cause.
Nothing finite and dependent (contingent) can cause itself.
A causal chain cannot be of infinite length.
Therefore, there must be a first cause; or, there must be something which is not an effect.
(Cosmological Argument)

Love is the law, love under will.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
</tbody>
</table>

### Chapter

1 GENERAL INTRODUCTION
   Spinal Muscular Atrophy (SMA) ...................................................... 1
   The Types of SMA .............................................................................. 2
   Other Variants of SMA ................................................................. 3
   Genetic Basis of SMA ..................................................................... 3
   SMN Protein .................................................................................... 8
   SMN Complex ................................................................................... 9
   Gemin2 and Gemin3 .......................................................................... 11
   X-Linked Spinal Muscular Atrophy (XL-SMA) ................................. 12

2 XL-SMA DNA LINKAGE STUDIES & CANDIDATE GENE ANALYSIS...
   Background ........................................................................................ 14
   Materials and Methods ...................................................................... 17
   Results ................................................................................................ 26
   Discussion ......................................................................................... 35

3 SMN COMPLEX & UBE1 STUDIES ...................................................... 36
   SMN Complex & UBE1 ...................................................................... 36
   Materials and Methods ...................................................................... 40
   Results ................................................................................................ 45
   Discussion ......................................................................................... 53

4 DISCUSSION ..................................................................................... 56
   Conclusions ....................................................................................... 59
   Future Studies .................................................................................. 62

REFERENCES ......................................................................................... 64
List of Figures

Figure 1.1. Genomic organization of the SMA locus in a healthy individual.............. 7
Figure 1.2. Genomic organization of the SMA locus in type I and type III patients................................................................. 7
Figure 1.3. Schematic diagram of SMN showing coding exons and relative localization of selected domains with known functions......... 9
Figure 1.4. SMN Complex .......................................................................................................................... 11
Figure 2.1. Design strategy for allelic discrimination assay with fluorogenic probes in the 5’ nuclease assay............................................. 17
Figure 2.2. Taqman PCR reaction................................................................................................. 18
Figure 2.3. Representative sample of possible results of SNP assays.............. 20
Figure 2.4. Selected SNPs in XL-SMA region........................................................................... 27
Figure 2.5. SNP analysis of Family 2............................................................................ 29
Figure 2.6. SNP analysis of Family 5............................................................................ 30
Figure 2.7. Representative sequencing data........................................................................ 33
Figure 3.1. The ubiquitin pathway......................................................................................... 39
Figure 3.2. Western blot analysis of protein content of SMN, Gemin2 and 3 in lymphoblastoid cell cultures of XL-SMA, SMA type 1 and healthy individuals........................................................................ 46
Figure 3.3. SMN Complex Protein Levels........................................................................... 48
Figure 3.4. Raw data from RNA Expression reaction ................................................. 50
Figure 3.5. Western blot analysis of UBE1 protein content in lymphoblastoid cultures of XL-SMA, SMA type 1 and healthy individuals........................................................ 53
Figure 4.1. Degradation model of XL-SMA...................................................................... 61
Figure 4.2. UBA6 Model of XL-SMA................................................................................. 62
List of Tables

Table 1.1. The different variants of Spinal Muscular Atrophy……………………. 4
Table 2.1. Overview of XL-SMA families............................................................. 16
Table 2.2. SNPs and their correlating cumulative LOD scores in
    Families 2, 3 and 5........................................................................................ 28
Table 2.3. The candidate genes screened in Miami.............................................. 34
Table 3.1. SMN Complex Protein Studies............................................................. 47
Table 3.2. Comparison of RNA Expression levels of SMN gene in XL-SMA
    and SMA Type I cell lines to healthy control cell line....................................51
Table 3.3. Comparison of RNA Expression levels of Gemin3 gene in
    XL-SMA and SMA Type I cell lines to healthy control cell line ............... 52
CHAPTER I

General Introduction

Spinal Muscular Atrophy (SMA)

Motor neuron diseases represent a heterogeneous group of disorders with respect to clinical presentation, disease course, genetic identity, underlying mutations, and etiologies. They are generally characterized by weakness due to muscle atrophy and/or spastic paralysis reflecting the selective involvement of lower and/or upper motor neurons (Ramser J. et al. 2008). The most common childhood motor neuron diseases are the autosomal recessive proximal spinal muscular atrophies associated with deterioration and destruction of anterior horn cells. These disorders are often fatal, leading to progressive symmetrical limb and trunk paralysis and severe muscle atrophy. Considerable clinical heterogeneity has been reported; genetic heterogeneity has also been documented.

Spinal muscular atrophy (SMA), the number one genetic killer of children under the age of two, is a group of inherited and often fatal diseases that destroys the nerves controlling voluntary muscle movement, which affects crawling, walking, head and neck control, and even swallowing. SMA affects approximately one in 6000 live births and has a carrier frequency of approximately one in 35 (Monani S 2005). Of children diagnosed before age two, 50 percent will die before their second birthday. The child of two carriers has a one in four chance of developing SMA.
The Types of SMA (SMN related Spinal Muscular Atrophy)

The classification of spinal muscular atrophy mapped to chromosome 5q13 (Melki J et al. 1990) is based on clinical criteria (Dubowitz V 1991)(see Table 1.1). Patients are classified into five types based on developmental milestones achieved at onset of SMA. Very severe spinal muscular atrophy with respiratory distress at birth is classified as spinal muscular atrophy type 0 and associated with limited life expectancy (Russman BS 2007). Type I and II are the most common forms of the disease.

Type I, or Werdnig-Hoffmann Disease, is the severe form of SMA. Werdnig and Hoffman are credited with first describing infantile spinal muscular atrophy (the severe form) in papers published in 1891 and 1892 (Werdnig G 1891, Hoffman J 1892). Children affected with Type I cannot sit without support. Fifty percent of Type I babies will die before their 2nd birthday. Spinal muscular atrophy type II is usually symptomatic between ages 6 and 18 months but may manifest earlier (Russman BS 2008). Type II patients may be able to sit unaided or even stand with support. They are at increased risk for complications from respiratory infections. Type III, also known as Kugelberg-Welander Disease, is the least deadly form of childhood-onset SMA. Description of the mild form arose as a result of Kugelberg and Welander’s efforts to separate and define spinal muscular atrophy from muscular dystrophy (Kugelberg E, Welander L. 1956). Type III patients are able to walk, but weakness is prevalent. Most patients eventually need to use a wheelchair. Spinal muscular atrophy type IV is usually
associated with onset older than 10 years and is associated with normal life expectancy (Russman BS 2008).

**Other Variants of Spinal Muscular Atrophy**

Other forms of spinal muscular atrophy are caused by mutation of other genes, some known and some not yet defined. All forms of SMA have in common weakness caused by denervation, that is, the muscle atrophies because it has lost the signal to contract due to loss of the innervating nerve. Spinal muscular atrophy only affects motor nerves. The term spinal muscular atrophy thus refers to atrophy of muscles due to loss of motor neurons within the spinal cord. Other forms of spinal muscular atrophy share certain characteristics with proximal SMA; however, they are genetically distinct and often affect different subsets of neurons and muscle. They include autosomal dominant forms of the disease, X-linked forms, recessive forms that affect the distal muscles, and a severe form of SMA (SMARD) with respiratory distress (Grohmann K et al. 2001) (Table 1).

**Genetic Basis of Spinal Muscular Atrophy**

Identifying the gene involved in SMA was complicated by the highly complex and unstable nature of the genome where it localizes and by phenotypes which range from very severe to very mild (Pearn J 1980). However, the genetics of the disease ended up being relatively simple.
<table>
<thead>
<tr>
<th>SMA Type</th>
<th>Mode of Inheritance</th>
<th>Gene/Chromosome Location</th>
<th>Phenotype</th>
<th>Onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMA Type 0</td>
<td>Autosomal Recessive</td>
<td>SMN1; 5q13</td>
<td>Very severe proximal muscle weakness with respiratory distress</td>
<td>Fatal at birth without respiratory support</td>
</tr>
<tr>
<td>SMA Type I</td>
<td>Autosomal Recessive</td>
<td>SMN1; 5q13</td>
<td>Proximal muscle weakness, patients never sit unaided, life span &lt; 2 years</td>
<td>&lt; 6 months</td>
</tr>
<tr>
<td>SMA Type II</td>
<td>Autosomal Recessive</td>
<td>SMN1; 5q13</td>
<td>Proximal muscle weakness, patients sit unaided but become wheelchair bound, develop scoliosis of spine</td>
<td>6-18 months</td>
</tr>
<tr>
<td>SMA Type III</td>
<td>Autosomal Recessive</td>
<td>SMN1; 5q13</td>
<td>Proximal muscle weakness, patients walk unaided, normal life span</td>
<td>&gt;18 months</td>
</tr>
<tr>
<td>SMA Type IV</td>
<td>Autosomal Recessive</td>
<td>SMN1; 5q13</td>
<td>Proximal muscle weakness</td>
<td>&gt;5 years</td>
</tr>
<tr>
<td>Distal SMA</td>
<td>Autosomal Recessive</td>
<td>11q3</td>
<td>Distal muscle weakness, diaphragmatic involvement</td>
<td>2 months-20 years</td>
</tr>
<tr>
<td>SMARD</td>
<td>Autosomal Recessive</td>
<td>IGHMBP2; 11q13.2</td>
<td>Distal lower limb weakness, diaphragmatic weakness, sensory, autonomic neurons also affected</td>
<td>1-6 months</td>
</tr>
<tr>
<td>XL-SMA</td>
<td>X-Linked</td>
<td>UBE1; Xp11.3-q11.2</td>
<td>Arthrogryposis, respiratory insufficiency, chest deformities, loss of anterior horn cells</td>
<td>At birth</td>
</tr>
<tr>
<td>SBMA</td>
<td>X-Linked</td>
<td>AR; Xq11.2-12</td>
<td>Proximal muscle weakness, lower motor neuron loss, bulbar involvement</td>
<td>30-50 years</td>
</tr>
<tr>
<td>Distal SMA V</td>
<td>Autosomal Dominant</td>
<td>GARS; 7p15</td>
<td>Distal muscles affected, bilateral weakness in hands</td>
<td>12-36 years</td>
</tr>
</tbody>
</table>

Table 1.1. The different variants of Spinal Muscular Atrophy
The characterization of the SMA locus on chromosome 5q13 revealed a chromosomal region characterized by an inverted duplication, each element (~500 kb) containing several genes (Melki J et al. 1994). Subsequently, SMA was associated with homozygous mutations in the Survival of Motor Neuron gene 1 (SMN1, OMIM number 600354) (Lefevbre S et al. 1995). The smallest deletions involving Survival of motor neuron gene 1 (SMN; renamed SMN1) and the presence of intragenic mutations of SMN1 in patients- including missense, nonsense or splice site mutations- have pinpointed SMN1 as the gene mutated in SMA (Melki J et al. 1994, Lefebvre S et al. 1995). SMN1 and its centromeric homolog SMN2 (OMIM number 601627), lie within the telomeric and centromeric halves, respectively, of a large inverted repeat in chromosome region 5q13 (Lefevbre S et al. 1995). SMN1 consists of 9 exons (exons 1, 2a, 2b and 3-8), with the stop codon near the end of the exon 7 (Burglen L et al. 1996).

Regardless of the severity of the disease, 95% of all the patients have a deletion of exon 7. The remaining 5% carry small mutations of the gene.

In contrast to SMN1, SMN2 is dispensable since 5-10% of normal individuals lack both of the copies (Gerard B et al. 2000). Only five nucleotides distinguish SMN1 from SMN2 gene without any effect on the amino acid sequence (Lefebvre S et al. 1995). One of these nucleotides is located in exon 7 (840 C>T). It is a translationally silent single nucleotide transition that profoundly influences splicing which is responsible for the difference in expression between two genes (Lorson CL et al. 1999). More recent evidence suggests that the 840 C>T transition in SMN2 activates an exonic splicing silencer, which functions as
a binding site for the known repressor protein hnRNP A1 (Kashima T, Manley JL 2003) (see Figure 1.1).

Full-length transcripts are almost exclusively produced by SMN1, whereas the predominant form encoded by SMN2 is lacking exon 7 (Melki J et al. 1994). The truncated transcript lacking exon 7 (SMNΔ7) is unstable both in vitro and in vivo, which explains why SMNΔ7 protein does not protect from SMN1 defect (Lorson CL et al. 2000). SMN2 does produce full length SMN protein albeit in very low levels (Figure 1.1.). Low levels of SMN protein are clearly insufficient for the survival of motor neurons and results in disease phenotype. Mutations of SMN1 and SMN2 on both chromosomes have not been reported. Such a genotype would likely be responsible for either an extremely severe form of SMA or a non viable fetus.

There is an inverted correlation between the amount of the protein encoded by the SMN2 gene and the clinical severity of human SMA disease (Lefevbre S et al. 1997). The protein level could thus depend on SMN2 copy number in patients, identifying SMN2 as a modifying gene in SMA (Frugier T et al. 2002). The greater the number of SMN2 genes, the more the protein is expressed, and the milder the disease phenotype (Figure 1.2.) (Mc Andrew PE et al. 1997, Feldkotter M et al. 2002).
**Figure 1.1. Genomic organization of the SMA locus in a healthy individual.**

*SMN1* is transcribed into full-length form (shown as 1…6-8) whereas *SMN2* is transcribed into both full-length or truncated form lacking exon7 (indicated as 1…6-8).

**Figure 1.2 Genomic Organization of the SMA locus in type I and type III patients.** In type I, *SMN1* gene is deleted and *SMN2* gene remains present. In type III, the *SMN1* gene is converted into *SMN2*, leading to an increased number of *SMN2* genes on each chromosome and thus, a higher amount of SMN protein.
SMN Protein

The SMN protein is transcribed and translated from two almost identical copies of *SMN* [centromeric (*SMN2*) and telomeric (*SMN1*)] within the 5q gene region (Lefebvre *et al.* 1995). The SMN protein consists of 294 amino acids and does not exhibit homology to any previously identified protein (Figure 1.3.). It has a molecular weight of 38 kDa. It is expressed in all metazoans and in all cell types of vertebrate organisms (Kolb SJ *et al.* 2007).

SMN does not appear to exist in cells in isolation but associates with several proteins to form a large multi-protein complex. The SMN complex is found both in the cytoplasm and in the nucleus where it is concentrated in a structure known as “gems” (for “Gemini of coiled bodies”) most often associated with or identical to Cajal bodies (coiled bodies) depending on the cell type or tissue type analyzed (Liu Q *et al.* 1996). Cajal bodies (CBs) are spherical sub-organelles found in the nucleus of proliferative cells like tumor cells, or metabolically active cells like neurons. Cajal bodies are sites of assembly or modification of the transcription machinery of the nucleus. (Cremer T *et al.* 2000).
Figure 1.3. Schematic diagram of SMN showing coding exons and relative localization of selected domains with known functions. The numbers above the boxes represent the corresponding amino acid localization. In SMN2, exon 7 is spliced out resulting in a truncated protein.

SMN Complex

The SMN complex is composed of the SMN protein and 7 additional proteins, Gemin 2-8 (Baccon J et al. 2002, Charroux B et al. 1999, Charroux B et al. 2000, Gubitz AK et al. 2002, Liu Q et al. 1996, Liu Q et al. 1997, Pellizoni L et al. 2002, Carissimi C et al. 2006) (Figure 1.4.). The Gemins bind to and co-localize with SMN in the cytoplasm and in discrete nuclear bodies called gems (Liu Q et al. 1997). The SMN protein is an essential protein in all cell types. Complete knockout of SMN results in embryonic lethality in divergent organisms, including human, mouse, chicken, C.elegans and S. pombe (Kolb SJ et al. 2007). SMN expression appears necessary for cell survival regardless of cell type; therefore, a major mystery of the pathogenesis in spinal muscular atrophy is the
basis of motor neuron-specific cell loss with relative depletion of SMN throughout the body.

The SMN complex is large, bound to several proteins, including the spliceosomal Sm proteins (Pellizoni L et al. 2002, Yong J et al. 2002). The functions of SMN complex include but not necessarily limited to: spliceosomal small nuclear ribonucleoprotein (snRNP) assembly; assembly, metabolism and transport of diverse classes of ribonucleoproteins, including small nucleolar ribonucleoproteins, telomerase ribonucleoprotein, micrornucleoproteins, and machineries that carry out transcription and premessenger RNA splicing (Friesen WJ et al. 2000, Buhler D et al. 1999, Jones KW et al. 2001, Meister G et al. 2000, Mourelatos et al. 2001, Mourelatos Z et al. 2002, Pellizoni L et al. 2001). Major spliceosomal snSNP’s are composed of one U snRNA molecule (U1, U2, U4/U6 or U5), a common core of seven Sm proteins and additional proteins specific for each snRNP (Will CL et al. 2001). From the eight gemins, three of them directly interact with SMN protein, while the remaining five gemins interact indirectly (Otter S et al. 2007) (Figure 1.4.).

The role of SMN in snRNP biogenesis and pre-mRNA splicing has been most extensively documented. However, numerous other SMN binding partners have been identified, and SMN has been found in various cellular compartments. It is therefore likely that SMN has other functions. But, based on our current understanding, it is reasonable to conclude that pathologic consequence of low SMN levels in spinal muscular atrophy is the disruption of normal RNA
metabolism required for normal motor development and survival (Kolb SJ et al. 2007).

From all the gemins, only 2, 3 and 8 directly interacts with SMN protein (Figure 1). From these three gemins, Gemin2 and Gemin3 had commercially available antibodies, and they were included in the protein studies.

**Figure 1.4. SMN Complex** (not drawn to scale)

**Gemin2 and Gemin3**

Gemin2 is expressed as a 1.3-kb transcript that encodes a 279-amino acid protein with a calculated molecular mass of 32 kD (Liu Q et al. 1997). SMN and Gemin2 interact tightly *in vivo* and *in vitro* and co-localize in gems in the nucleus as well as in the cytoplasm (Liu Q et al. 1997). Gemin2 is a core protein that is essential for the formation of the SMN complex, although the mechanism by
which it drives formation is unclear. Gemin2 also plays an important role in snRNP assembly through the stabilization of the SMN oligomer/complex via novel self-interaction (Ogawa C et al. 2007).

Gemin3 encodes a deduced 824-amino acid protein containing all of the common DEAD box motifs (Charroux B et al. 1999). It has an apparent molecular mass of 103 kD. Immunoprecipitation experiments detected ATPase activity (Grundhoff et al. 1999). Gemin3 interacts directly with SMN, as well as with SmB, SmD2, and SmD3. Gemin3 co-localizes with SMN in gems (Charroux B et al. 1999) Gemin3 binds SMN via its unique COOH-terminal domain, and SMN mutations found in some SMA patients strongly reduce this interaction. The presence of a DEAD box motif in Gemin3 suggests that it may provide the catalytic activity that plays a critical role in the function of the SMN complex on RNPs (Charroux B et al. 1999). In experimental systems, a reduction of Gemins also decreases smallnucleoprotein assembly. (Feng W et al. 2005).

X-Linked Spinal Muscular Atrophy (XL-SMA)

X- linked Infantile Spinal Muscular Atrophy (OMIM number 301830) is a rare congenital disorder characterized by multiple joint contractures. This disease phenotype overlaps with arthrogypsis and SMA. Arthrogryposis, also known as Arthrogryposis Multiplex Congenita, is a rare congenital disorder that causes multiple joint contractures and is characterized by muscle weakness and fibrosis. It is a non-progressive disease. The disease derives its name from Greek, literally meaning 'curved or hooked joints' (www.wikipedia.com).
There are three different types of X-linked spinal muscular atrophy (Hall JG et al. 1982): 1) a severe (lethal) type, characterized by severe contractures, scoliosis, chest deformities, hypotonia and death due to respiratory insufficiency within three months of birth; 2) a moderately severe type, characterized by severe contractures, ptosis, microphallus, cryptorchidism, inguinal hernias, and normal intelligence; and 3) a resolving type, characterized by mild to moderate contractures at birth that improve with time. Progressive loss of anterior horn cells has been associated with the severe lethal form of X-linked spinal muscular atrophy (Hall J et al. 1982). Under the label “X-linked spinal muscular atrophy” a similar disorder to Spinal Muscular Atrophy Type I (SMA1) was described as X-linked recessive and was associated with hypotonia, areflexia, chest deformities, facial dysmorphic features and congenital joint contractures (Greenberg F et al. 1988). The findings of electromyography and muscle biopsy were consistent with autosomal recessive infantile spinal muscular atrophy (Werndig-Hoffman, OMIM 253300).

Our laboratory studied the family originally reported by Greenberg (Greenberg et al. 1988). In 1995, a pericentromeric candidate interval was defined for XL-SMA in Xp11.3-Xq11.2 with positive LOD-scores between markers MAOB and DXS991 (Kobayashi H et al. 1995). This thesis chronicles what has been done by me since that discovery.
CHAPTER 2

XL-SMA DNA Linkage Studies and Candidate Gene Analysis

Background

X-linked lethal infantile spinal muscular atrophy is a rare disease with scarce number of families available to work with for further studies. A substantial amount of effort was placed on patient and family identification, ascertainment, recruitment, and collection of blood and other tissues specimens for molecular studies. All human subjects were recruited and studies performed under institutional human subjects’ approval. Subsequently, meiotic breakpoint mapping (concordance analysis) and whole chromosome multipoint linkage analysis to the X-linked spinal muscular family reported by Greenberg et al. was described by Dr. Baumbach’s group (Baumbach, L et al. 1993; Baumbach, L et al. 1993(b); Baumbach, L et al. 1996; Baumbach, L et al. 1996(b); Kobayashi H et al. 1995).

In brief, XL-SMA displays clinical features most similar to a severe form of X-linked arthrogryposis, while the disease course is most similar to Type I SMA, resulting in death in the majority of patients within the first two years of life. Arthrogryposis (ARGY) is a complex clinical phenotype involving multiple congenital contractures and limited movement of multiple body areas, more often distal than proximal (Banker BQ 1986). Several forms of X-linked ARGY have been reported (Hall JG 1985). The disease phenotype we have been studying is referred to as ‘XL-SMA’ presents with the clinical characteristics of hypotonia,
areflexia, and multiple congenital contractures (arthrogryposis) associated with
loss of anterior horn cells and death in infancy. Muscle biopsy findings and/or
electromyogram (EMG) studies have been consistent with and indicative of
neurogenic atrophy, and autopsy examinations have revealed loss of anterior
horn cells. Carrier females appear to have no clinical manifestations of this
disorder. The list of the defining features and laboratory findings associated with
XL-SMA can be described as follows: congenital hypotonia; arthrogryposis +
bone fractures; dysmorphic features, including myopathic facies and digital
contractures; death of at least one affected within one year of birth, due to
respiratory distress; family history of miscarriages / spontaneous abortions;
genital abnormality (undescended testes); muscle biopsy confirmation of
neurogenic atrophy; electromyogram (EMG) indicative of denervation and
autopsy showing anterior horn cell loss.

The initial gene linkage studies (1993-1995) were performed on XL-SMA
family 2; a large multigenerational family which was previously reported
(Greenberg F et al. 1988). Cell lines were established from many family
members including affected males. This enabled DNA extraction for further
linkage studies. These studies localized the XL-SMA disease gene interval to a
12 cM pericentromeric region between Xp11.3-q11.2, which excluded the great
remainder of the X chromosome (Kobayashi H et al. 1995). The approach of
concordance analysis was used as it scans for regions of a chromosome whether
there is complete segregation of particular alleles of a DNA marker (in this case,
dinucleotide repeats) with the disease phenotype. This approach allowed for
initial identification of the 12cM region. Multipoint and 2-point linkage analysis was performed, and the region defined by markers DXS1003 and DXS991 showed the highest positive LOD scores (explained in Materials and Methods), with a max LOD score of 2.3 at \( \theta \) of 0.0 (Kobayashi H et al. 1995). Subsequently, using framework markers (microsatellite markers and dinucleotide repeats) in the region, additional families were found to map as the same region as family 2 (Dressman D et al. 2007). To date, 15 families with XL-SMA were identified and specimens were collected. Of 15, 14 are multigenerational, which are from North America and Europe (Table 2.1.).

<table>
<thead>
<tr>
<th>Family</th>
<th>Family Origin</th>
<th>Generations</th>
<th>Affected Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>South Carolina</td>
<td>4</td>
<td>7(0)</td>
</tr>
<tr>
<td>2</td>
<td>Texas/Indiana</td>
<td>4</td>
<td>4(3)</td>
</tr>
<tr>
<td>3</td>
<td>Spain</td>
<td>4</td>
<td>7(2)</td>
</tr>
<tr>
<td>4</td>
<td>California Utah Colorado</td>
<td>5</td>
<td>8(1)</td>
</tr>
<tr>
<td>5</td>
<td>California (L.A.)</td>
<td>4</td>
<td>6(2)</td>
</tr>
<tr>
<td>6</td>
<td>Virginia</td>
<td>3</td>
<td>4(0)</td>
</tr>
<tr>
<td>7</td>
<td>Illinois</td>
<td>3</td>
<td>3(2)</td>
</tr>
<tr>
<td>8</td>
<td>United States</td>
<td>3</td>
<td>1(0)</td>
</tr>
<tr>
<td>9</td>
<td>New York/Mexico</td>
<td>3</td>
<td>5(0)</td>
</tr>
<tr>
<td>10</td>
<td>California</td>
<td>3</td>
<td>2(0)</td>
</tr>
<tr>
<td>11</td>
<td>New York</td>
<td>2</td>
<td>2(0)</td>
</tr>
<tr>
<td>12</td>
<td>Germany</td>
<td>1</td>
<td>2(0)</td>
</tr>
<tr>
<td>13</td>
<td>Penn.</td>
<td>2</td>
<td>2(0)</td>
</tr>
<tr>
<td>14</td>
<td>Indiana</td>
<td>2</td>
<td>2(0)</td>
</tr>
<tr>
<td>15</td>
<td>Thailand</td>
<td>3</td>
<td>10(1)</td>
</tr>
</tbody>
</table>

Table 2.1. Overview of XL-SMA families: (a) The numbers in parentheses indicate the number of affected males available for DNA collection.
The next chapter describes my portion of the additional linkage studies subsequent to the 1995 paper which eventually helped finding the XL-SMA gene. My studies were focused on single nucleotide polymorphisms (SNP) analysis for linkage studies. A single nucleotide polymorphism is a DNA sequence variation occurring when a single nucleotide - A, T, C, or G - in the genome differs between members of a species (or between paired chromosomes in an individual). This chapter consists of results of SNP analysis of different families and explains how this data contributed to new LOD scores. It also discusses how the candidate gene screening was performed in Miami.

The data presented in this chapter have been published as peer reviewed manuscripts in Genetics in Medicine (Dressman D. et al. 2007) and American Journal of Human Genetics (Ramser J. et al. 2008). I was a co-author in both of these papers.

Materials and Methods

TaqMan® Technology for SNPs

TaqMan® technology is used in allele specific detection of SNPs in a PCR based reaction. The TaqMan® approach (developed by Applied Biosystems) uses an allele specific probe for each SNP. These probes have a reporter dye at the 5’ end of the probe and a non-fluorescent quencher dye at the 3’ end. When both dyes are attached to the probe, reporter dye emission is quenched via Förster type energy transfer (FRET) (Förster T, 1948). The FRET principle relies on the non-radioactive transfer of energy from an excited donor fluorophore to an
acceptor fluorophore by means of intermolecular long-range dipole-dipole coupling. Förster showed that the efficiency of this energy transfer (E) depends on the inverse sixth-distance between donor and acceptor. 

\[ E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \]

where \( R_0 \) is the distance where half the energy is transferred. Typically \( R_0 \) is between 10 – 70 Å, therefore the distance between the two fluorophores and their relative orientation is crucial. During the PCR reaction, the Taq polymerase cleaves the probe which separates the reporter from quencher resulting in fluorescence (Figure 2.2.). Accumulation of the PCR products is detected directly by monitoring the increase in fluorescence dye. There are two probes in each reaction specifically binding to two different alleles of the SNP of the same locus. Both of the probes have different reporter dyes; hence the equipment recognizes two different fluorescence emissions (two different colors) (Figure 2.1.).

**Figure 2.1. Design strategy for allelic discrimination assay with fluorogenic probes in the 5’ nuclease assay.** NFQ: Non-fluorescent quencher, MGB: Minor groove binder (DNA probes with conjugated minor groove binder (MGB) groups form extremely stable duplexes with single-stranded DNA targets, allowing shorter probes to be used for hybridization based assays (Kutyavin IV et al. 2000)
Figure 2.2. Taqman PCR reaction. Taq polymerase cleaves the probe between the reporter and the quencher (with its 5’ nuclease activity) only if the probe hybridizes to the target. The probe fragments are then displaced from the target, and polymerization of the strand continues.
SNP Analysis

Genomic DNA was extracted from human blood samples and/or cheek swabs with the GenePure Kit from Gentra according to the manufacturer's instructions. DNA concentrations and purity were measured by using standard protocols with an Eppendorf spectrophotometer. Allele detection of SNPs was based on TaqMan® technology from Applied Biosystems. The 14 SNPs used in this chapter were from the Applied Biosystems catalogue of Assay-on-Demand SNPs (SNPs from the genes NDP, FLJ22843 (2 SNPs)), UTX, CXOrf36, FLJ20344, CHST7, SLC9A7, PHF16, ZNF41, GRIPAP1, FLJ31204, ARHGEF9, VSIG4). These SNPs were chosen as they were either at the upper or the lower boundaries of the XL-SMA region or they were close to the region with the highest LOD score. All of the SNP detection reactions were performed and analyzed using an ABI Prism 7000 Sequence Detection System. The SNP reactions were set up as follows (for 25 µl reaction: 12.5 µl of 2X TaqMan Universal Master Mix, 1.25 µl of 20X SNP Genotyping Assay Mix and 11.25 µl mix of water and DNA (DNA was 10 ng for each reaction)). Each SNP Genotyping assay consisted of a single tube containing two primers for amplifying the sequence of interest and two TaqMan probes (of different colors) for detecting alleles. Every SNP reaction for each DNA sample is performed in pairs. Every reaction set had at least three Non Template Controls (NTCs) as negative controls (water controls). The PCR reaction is as follows: 10 min 95°C hold and 40X of 15 sec 92°C Denature and 1 min 60°C Anneal/Extend cycles. After the PCR reaction, there is a one minute analysis for the SNPs as allele data.
is expressed as fluorescent output. After analysis, there is an output with three
distinct clusters of samples (for two different alleles of the SNP and the
SNP1/SNP2 heterozygote) (Figure 2.3.).

**Figure 2.3. Representative sample of possible results of SNP assays.** The
figure displays 2 SNP alleles with two distinct colors, while heterozygotes have
their own color and on the diagonal. No template controls (negative controls)
cluster on the bottom left of the output.

**LOD Score**

LOD (logarithm (base 10) of odds) score is a statistical estimate of
whether two loci are likely to lie near each other on a chromosome and are
therefore likely to be inherited together. Newton E. Morton developed the LOD
score method (Morton NE 1955). It is an iterative approach where a series of LOD scores are calculated from a number of proposed linkage distances. The following is the formula for the LOD score:

\[ Z(\theta) = \log \frac{(\theta^k)(1-\theta)^{n-k}}{(0.5)^n} \]

where \( \theta \) denotes recombination the frequency, \( n \) denotes the total number of people examined, and \( k \) denotes the number of people with recombinant genotypes. The reason 0.5 is used in the denominator is that any alleles that are completely unlinked (e.g. alleles on separate chromosomes) have a 50% chance of recombination, due to independent assortment. The value between \( 0 < \theta < 0.5 \) resulting in the highest LOD score is called the maximum likelihood estimate (MLE) \( (\theta) \) of the recombination fraction. A series of these LOD scores are obtained using different linkage distances, and the linkage distance giving the highest LOD score is considered the estimate of the linkage distance. Linkage is considered likely if a LOD score is greater than 3.0 (for the X-chromosome this number is 2.0). Today, linkage software programs (Allegro, Merlin etc.) are available to researchers which calculate the LOD scores. The LOD scores presented in this thesis were calculated by M. Barmada (with Allegro software) of the University of Pittsburgh who is our collaborator and an expert in statistical genetics.
Candidate Gene Screening

This part of the project relied heavily on following the literature, having the latest build of the emerging human genome sequence and using bioinformatics. Every new build of the human genome sequence was carefully analyzed for new candidate genes. The relevant literature was scanned every week for new information. Four major groups of genes were considered as these were related to either SMA or the disease phenotype: 1) genes involved in RNA processing/expression, 2) genes involved in cell-to-cell communication, cell cycle and/or apoptosis, 3) genes involved in neuronal and muscle development and 4) genes expressed in fetal muscle or the central nervous system.

The basic experimental strategy involved: 1) choosing the genes to be screened, 2) designing primers 3) using PCR and 4) DNA sequencing. The structures of the genes were checked both in NCBI (www.ncbi.nlm.nih.gov) and Ensemble (www.ensemble.org) websites. Primers were designed at Primer3 website (frodo.wi.mit.edu). PCR was performed with reagents from Promega (Taq polymerase, PCR Buffer, MgCl₂) and Perkin-Elmer (dNTPs) in a Perkin-Elmer 9600 thermal cycler. All the optimizations were done by me by varying magnesium concentrations and changing primers where necessary. The sequencing included affected male samples from two different XL-SMA families (Families 2 and 5), obligate carrier samples and control samples. The PCR reactions were set up as follows (for each 25 µl reaction: 18 µl of water and magnesium chloride mix, 2.5 µl of 10X PCR Buffer, 1 µl of dNTP mix (10mM), 1 µl of forward primer (0.5 µM), 1 µl of reverse primer (0.5 µM), 1 µl of DNA (100
ng) and 0.5 µl of Taq Polymerase). The PCR reaction was as follows: 5 min 94°C hold and 30 cycles of 30 sec 94°C Denature, 30 sec 60°C Anneal and 30 sec 72°C Extend cycles and a 10 min 72°C hold. After the PCR reaction, the PCR products were run on a 2% agarose gel and the bands were visualized by Ethidium Bromide under UV light. The cleanest products were purified by Qiagen’s QIAquick PCR Purification Kit. The purified PCR products were then mixed with their corresponding forward and reverse primers (each tube had 0.25 µM Primer, 10 to 20 ng DNA and ddH2O to constitute a total volume of 14 µl) and sent to the Rosenstiel Sequencing Facility to be sequenced. Results were analyzed with Genemapper 3.0 software by Applied Biosystems by me. The primers for specific genes were:

<table>
<thead>
<tr>
<th>Gene/Exon/Direction</th>
<th>Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUSP21/F</td>
<td>5'-gggagtgtgtgagaggagaca-3'</td>
</tr>
<tr>
<td>DUSP21/R</td>
<td>5'-ctgacagcatttgccca-3'</td>
</tr>
<tr>
<td>I-4/F</td>
<td>5'-ctgatgtcttaactagggcg-3'</td>
</tr>
<tr>
<td>I-4/R</td>
<td>5'-gcatgcgctacagaaaaac-3'</td>
</tr>
<tr>
<td>SPIN2/1/F</td>
<td>5'-tgcaaaaaaagggcagagac-3'</td>
</tr>
<tr>
<td>SPIN2/1/R</td>
<td>5'-catccccctctccattctg-3'</td>
</tr>
<tr>
<td>SPIN2/2A/F</td>
<td>5'-ggcagactacatcagaca-3'</td>
</tr>
<tr>
<td>SPIN2/2A/R</td>
<td>5'-gagctaagaccatccccctc-3'</td>
</tr>
<tr>
<td>SPIN2/2B/F</td>
<td>5'-acaggtggcatcatca-3'</td>
</tr>
<tr>
<td>SPIN2/2B/R</td>
<td>5'-cagctgtcggtcccttcca-3'</td>
</tr>
<tr>
<td>GSPT2/1/F</td>
<td>5'-cgctcacacacaagaggaga-3'</td>
</tr>
<tr>
<td>GSPT2/1/R</td>
<td>5'-tgtaacggctgaaagttggaa-3'</td>
</tr>
<tr>
<td>GSPT2/2/F</td>
<td>5'-ggcgaagctcaacgtcaac-3'</td>
</tr>
<tr>
<td>GSPT2/2/R</td>
<td>5'-aggtccagacgagacagc-3'</td>
</tr>
<tr>
<td>GSPT2/3/F</td>
<td>5'-agtcacatgacgacgacg-3'</td>
</tr>
<tr>
<td>GSPT2/3/R</td>
<td>5'-ccacagtgcctccatttc-3'</td>
</tr>
<tr>
<td>GSPT2/4/F</td>
<td>5'-ggactgagcggagcaaat-3'</td>
</tr>
<tr>
<td>GSPT2/4/R</td>
<td>5'-gatgtctgctctgcttct-3'</td>
</tr>
<tr>
<td>GSPT2/5/F</td>
<td>5'-acagctgaactctgtgttgc-3'</td>
</tr>
<tr>
<td>GSPT2/5/R</td>
<td>5'-actgtcagacacactgtgttgc-3'</td>
</tr>
<tr>
<td>LOC286512/1/F</td>
<td>5'-tctttgggtgtcagattttc-3'</td>
</tr>
<tr>
<td></td>
<td>Primer Sequence</td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>LOC286512/1/R</td>
<td>5’-agcctgcagtttttgcattt-3’</td>
</tr>
<tr>
<td>LOC286512/2/F</td>
<td>5’-caggctgaagcaccacatatt-3’</td>
</tr>
<tr>
<td>LOC286512/2/R</td>
<td>5’-tcagggcactccttttactg-3’</td>
</tr>
<tr>
<td>ASB12/1/F</td>
<td>5’-gcctaaacattgcccccttg-3’</td>
</tr>
<tr>
<td>ASB12/1/R</td>
<td>5’-gaaacagcctacaacagggaa-3’</td>
</tr>
<tr>
<td>ASB12/2A/F</td>
<td>5’-acagttggattggagtgag-3’</td>
</tr>
<tr>
<td>ASB12/2A/R</td>
<td>5’-tccaagctgtaacatcagc-3’</td>
</tr>
<tr>
<td>ASB12/2B/F</td>
<td>5’-gctgttagttgacagcttgga-3’</td>
</tr>
<tr>
<td>ASB12/2B/R</td>
<td>5’-agggcagatttcaagaggt-3’</td>
</tr>
<tr>
<td>ASB12/3/F</td>
<td>5’-ggcttcgcttgcgctaatga-3’</td>
</tr>
<tr>
<td>ASB12/3/R</td>
<td>5’-ggggacaaaaactaggttaatgg-3’</td>
</tr>
<tr>
<td>ZNF157/1/R</td>
<td>5’-ggggacaaaaactaggttaatgg-3’</td>
</tr>
<tr>
<td>ZNF157/2-3/F</td>
<td>5’-aatcttcaacagggctccatt-3’</td>
</tr>
<tr>
<td>ZNF157/2-3/R</td>
<td>5’-aatcttcaacagggctccatt-3’</td>
</tr>
<tr>
<td>ZNF157/4A/F</td>
<td>5’-cagaaagccagttggtgat-3’</td>
</tr>
<tr>
<td>ZNF157/4A/R</td>
<td>5’-ctcccctgtgtgagttctctg-3’</td>
</tr>
<tr>
<td>ZNF157/4B/F</td>
<td>5’-cagaaagccagttggtgat-3’</td>
</tr>
<tr>
<td>ZNF157/4B/R</td>
<td>5’-ttcgttagttgacagcttgga-3’</td>
</tr>
<tr>
<td>ZNF157/4C/F</td>
<td>5’-cagaaagccagttggtgat-3’</td>
</tr>
<tr>
<td>ZNF157/4C/R</td>
<td>5’-cagaaagccagttggtgat-3’</td>
</tr>
<tr>
<td>DT1P1A10/1-2/F</td>
<td>5’-ctcccctgtgtgagttctctg-3’</td>
</tr>
<tr>
<td>DT1P1A10/1-2/F</td>
<td>5’-ctcccctgtgtgagttctctg-3’</td>
</tr>
<tr>
<td>DT1P1A10/3-4/F</td>
<td>5’-ctcccctgtgtgagttctctg-3’</td>
</tr>
<tr>
<td>DT1P1A10/3-4/F</td>
<td>5’-ctcccctgtgtgagttctctg-3’</td>
</tr>
<tr>
<td>DT1P1A10/5/F</td>
<td>5’-ctcccctgtgtgagttctctg-3’</td>
</tr>
<tr>
<td>DT1P1A10/5/F</td>
<td>5’-ctcccctgtgtgagttctctg-3’</td>
</tr>
<tr>
<td>UXT/1-2/F</td>
<td>5’-tgcccacttaacctgcacc-3’</td>
</tr>
<tr>
<td>UXT/1-2/R</td>
<td>5’-tgcccacttaacctgcacc-3’</td>
</tr>
<tr>
<td>UXT/3-4/F</td>
<td>5’-ctcccctgtgtgagttctctg-3’</td>
</tr>
<tr>
<td>UXT/3-4/R</td>
<td>5’-ctcccctgtgtgagttctctg-3’</td>
</tr>
<tr>
<td>UXT/5/F</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>BMP15/1/R</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>BMP15/1/R</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>BMP15/2/F</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>BMP15/2/F</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>INE1/F</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>INE1/R</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>LOC392453/F</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>LOC392453/R</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>BMP15/1/F</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>BMP15/1/R</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>BMP15/2/F</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
<tr>
<td>BMP15/2/R</td>
<td>5’-gggtggagacaacactataaa-3’</td>
</tr>
</tbody>
</table>
Results

SNP Analysis

14 SNPs from nine different genes were chosen according to their respective location and heterozygosity (Figure 2.4). The investigated locations were clustered in three areas: 1) the upper boundary of the XL-SMA region that is near the *NDP* gene, 2) the region around DXS1003 (the area of highest LOD score), and 3) the Xq part of the region. Thus, using both microsatellites and SNPs, relatively dense genotyping was completed in our three largest XL-SMA families (Families 2, 3, and 5). Haplotypes were constructed and two-point and multipoint LOD score analyses completed. These studies slightly narrowed the XL-SMA region, and failed to reveal genetic heterogeneity among the families analyzed, strongly suggesting that XL-SMA in these families originate from genetic defects in a single disease locus (Dressman D *et al.* 2007).

The candidate gene interval was refined by studying 14 SNPs in the three largest families. This analysis revealed a recombination event in an affected male in Family 2 which allowed elimination of the *NDP* gene (Figure 2.5). In Family 5, there are similar recombination events in two unaffected siblings who have a non-carrier mother (Figure 2.6). Additionally, with the data from these 14 SNPs, the calculation of new multipoint LOD scores was possible. These data showed significantly positive LOD scores in the region between *FLJ22843* SNP and *ARHGEF9* SNP (Table 2.2).
Figure 2.4. Selected SNPs in XL-SMA region (in green). The region between two plus signs is the XL-SMA region. Microsatellite markers were used as the other leg of this linkage studies (Dressman D et al. 2007).
<table>
<thead>
<tr>
<th>SNP</th>
<th>Location (cM)</th>
<th>LOD Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDP</td>
<td>38.154</td>
<td>-∞</td>
</tr>
<tr>
<td>FLJ22843</td>
<td>38.507</td>
<td>3.5996</td>
</tr>
<tr>
<td>FLJ22843_2</td>
<td>38.507</td>
<td>3.6122</td>
</tr>
<tr>
<td>UTX</td>
<td>39.16</td>
<td>4.069</td>
</tr>
<tr>
<td>CXOrf36</td>
<td>39.361</td>
<td>4.0687</td>
</tr>
<tr>
<td>FLJ20344</td>
<td>40.657</td>
<td>3.9427</td>
</tr>
<tr>
<td>CHST7</td>
<td>40.778</td>
<td>3.9213</td>
</tr>
<tr>
<td>SLC9A7</td>
<td>40.916</td>
<td>3.8875</td>
</tr>
<tr>
<td>PHF16</td>
<td>41.165</td>
<td>3.8616</td>
</tr>
<tr>
<td>ZNF41</td>
<td>41.673</td>
<td>3.8727</td>
</tr>
<tr>
<td>GRIPAP1</td>
<td>43.197</td>
<td>3.9021</td>
</tr>
<tr>
<td>FLJ31204</td>
<td>51.811</td>
<td>2.7287</td>
</tr>
<tr>
<td>ARHGEF9</td>
<td>57.232</td>
<td>1.5526</td>
</tr>
<tr>
<td>VSIG4</td>
<td>59.612</td>
<td>-0.0857</td>
</tr>
</tbody>
</table>

Table 2.2. SNPs and their correlating cumulative LOD scores in Families 2, 3 and 5.
Figure 2.5. SNP analysis of Family 2. SNP analysis in this family shows a recombination in individual IV:2, an affected male, between NDP and FLJ22843, which allows elimination of the NDP gene. There is a second recombination in a noncarrier female, individual III:15, which is not informative for XL-SMA.
Figure 2.6. SNP analysis of Family 5. SNP analysis in this family shows a recombination in individuals III:8 and III:9 between GRIPAP1 (Xp11.23) FLJ31204 (Xp11.21); however this does not help to narrow the region as the recombination has occurred in a portion of the family without affected individuals.
Candidate Gene Screening

In total, I have screened the exons and exon-intron boundaries of 12 genes in genomic DNA (Table 2.3, Figure 2.7.) : **BMP15** (Bone morphogenetic protein 15), **I-4** (protein phosphatase inhibitor 4), **DUSP21** (Dual specificity phosphatase 21), **SPIN-2** (Spindlin like protein 2), **LOC286512** (similar to translation initiation factor eIF4A-II), **ASB12** (ankyrin repeat and SOCS box-containing 12), **DT1P1A10** (hypothetical protein DT1P1A10), **ZNF157** (Zinc finger protein 157), **GSPT2** (G1 to S phase transition 2), **LOC392453** (similar to arginine-serine-rich splicing factor 6; splicing factor, arginine-serine-rich, 55 kDa; pre-mRNA splicing factor SRP55), **INE1** (Inactivation escape 1), **UXT** (Ubiquitously-expressed transcript). None of these genes had mutations in the affected males in their exons or intron boundaries.

In parallel, all annotated genes in the critical region were screened at the cDNA level in two unrelated XL-SMA males using denaturing high-performance liquid chromatography (DHPLC) followed by DNA sequencing. These studies were conducted in the laboratory of Dr. Alfons Meindl (Germany) whose laboratory is internationally known for their efforts in X chromosome disease gene identification, particularly in this region (Thiselton DL et al. 2002, Lenski C et al. 2004, Ramser J et al. 2004, Ramser J et al. 2005).

In late 2006, the Meindl Lab discovered that mutations in **UBE1** (Ubiquitin activating enzyme 1) was the cause of XL-SMA. **UBE1** catalyzes the ubiquitin-proteasome system (UPS) as the first step in ubiquitin conjugation to mark
cellular proteins for degradation. Two novel missense mutations (c.1617 G>T, pM539I; c.1639 A>G, pS547G) in two families (families # 5, # 7) and a novel synonymous C>T substitution (c.1731 C>T, pN577N) in the index patients of another three families (families # 2, # 4, # 15) were discovered (Ramser J et al. 2008).
Figure 2.7. Representative sequencing data. The raw data was obtained from the sequencing center and then analyzed with Sequencher 3.0 software. The results were compared to the control sequence from Ensembl webpage (www.ensembl.org). In this example, three forward and reverse pairs (two affected individuals and one normal individual) were examined (for ASB12 exon1).
<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUSP21</td>
<td>Xp11.4-p11.23</td>
<td>Dual specificity phosphatase 21</td>
</tr>
<tr>
<td>GSPT2</td>
<td>Xp11.23-p11.21</td>
<td>G1 to S phase transition 2</td>
</tr>
<tr>
<td>I-4</td>
<td>Xp11.4-p11.3</td>
<td>Type 1 protein phosphatase inhibitor (inhibits activity of myosin-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>associated phosphates)</td>
</tr>
<tr>
<td>INE1</td>
<td>Xp11.4-p11.3</td>
<td>Inactivation escape 1</td>
</tr>
<tr>
<td>LOC286512</td>
<td>Xp11.23</td>
<td>Similar to translation initiation factor eIF4a-II</td>
</tr>
<tr>
<td>LOC392453</td>
<td>Xp11.3</td>
<td>similar to arginine/serine-rich splicing factor 6; splicing factor,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>arginine/serine-rich, 55 kDa; pre-mRNA splicing factor SRP55</td>
</tr>
<tr>
<td>BMP15</td>
<td>Xp11.2</td>
<td>Bone morphogenetic protein 15</td>
</tr>
<tr>
<td>SPIN2</td>
<td>Xp11.23-p21</td>
<td>Spindlin like protein 2 / nuclear anti-apoptotic protein</td>
</tr>
<tr>
<td>ASB12</td>
<td>Xq11.1</td>
<td>Ankyrin repeat, SOCS box-containing 12/ protein degradation targeting</td>
</tr>
<tr>
<td>ZNF157</td>
<td>Xp11.2</td>
<td>Zinc finger protein 157 (HZF22)</td>
</tr>
<tr>
<td>DT1P1A10</td>
<td>Xp11.22</td>
<td>Hypothetical protein DT1P1A10</td>
</tr>
<tr>
<td>UXT</td>
<td>Xp11.23-p11.22</td>
<td>Ubiquitously-expressed transcript</td>
</tr>
</tbody>
</table>

Table 2.3. The candidate genes screened in Miami.
Discussion

These results strongly supported the existence of a disease locus for XL-SMA within the Xp11.3-Xq11.2 region. Analysis of linkage data from only families 2, 3 and 5 using both microsatellite markers and SNPs generated a maximum multipoint LOD score of 4.0818 and 4.0766 at UTX and DXS8026, respectively, suggesting a slightly smaller region of peak LOD. The overall congruence of positive LOD score in essentially the same region of the X chromosome with data generated by different laboratory methods and separate by LOD score analysis lends further evidence that the disease-causing gene for X-linked SMA lies within this region (Dressman D et al. 2007).

With the strategy of candidate gene screening, the genes in this region were sequenced in two different labs with two different approaches. In summary, more than hundred genes were screened, 12 of which were screened in Miami by me. In the end, the XL-SMA disease causing mutations were found in the UBE1 gene. I offer an additional explanation for a protein structure change in the case of the substitution (in Ramser J et al. 2008, it is explained as a CpG island disruption by bisulfate sequencing). As recently reported, single nucleotide polymorphisms that are considered "silent" can affect protein folding (Kimchi-Sarfaty C et al. 2007). The presence of a rare codon, marked by the synonymous polymorphism, affects the timing of cotranslational folding, thereby altering the structure of substrate and inhibitor interaction sites. This can be a possibility in this UBE1 synonymous substitution, because it causes a change from the major Asn to minor Asn codon (AAC to AAT) (Kotlar D and Lavner Y 2006).
SMN Complex and UBE1 Studies

SMN Complex and UBE1

XL-SMA and SMA share many phenotypic similarities. Thus, it is only logical to hypothesize there may be a similar underlying disease mechanism. Therefore, I studied the SMN Complex (both at the protein and RNA levels) in cell cultures from XL-SMA patients. Gemin2 and Gemin3 proteins were investigated with SMN as they were the only commercially available antibodies. UBE1 protein was subsequently added to these studies as mutations in UBE1 gene were found to be the cause of XL-SMA.

The SMN Complex

The SMN protein is a 294-amino acid polypeptide that is expressed in all metazoans and in all cell types of vertebrates (Kolb SJ et al. 2007). Biochemically, SMN protein does not appear to exist within cells in isolation but instead forms a large protein complex, the SMN complex. The SMN Complex is composed of the SMN protein and 7 additional proteins, Gemins 2-8. (Bacco et al. 2002, Charroux B et al. 1999, Charroux B et al. 2000, Gubitz AK et al. 2002, Liu Q et al. 1996, Liu Q et al. 1997, Pellizoni L et al. 2002, Carissimi C et al. 2006). The gemins bind to and colocalize with SMN in the cytoplasm and in discrete nuclear bodies called gems (Liu Q and Dreyfuss G 1996). SMN protein is an essential protein in all cell types studied thus far. Complete knockout of
SMN results in embryonic lethality in diverse organisms (Yong J et al. 2004). SMN expression appears necessary for cell survival regardless of cell type.

SMN Complex is essential for the biogenesis of spliceosomal small nuclear ribonucleoproteins and likely functions in the assembly, metabolism and transport of diverse number of other ribonucleoproteins (Gubitz AK et al. 2004). SMN also accumulates in the axons of motor neurons during nervous system development and thus may also have functions that are unique to neurons (Giavazzi A et al. 2006, Carrel TL et al. 2006, and Zhang H et al. 2006). Given the large number of binding partners of SMN and SMN Complex, it is likely that the cellular functions attributed to the SMN Complex will be likely to increase, however based on the current knowledge, it is reasonable to postulate that the pathologic consequence of low SMN levels in spinal muscular atrophy is the disruption of normal cellular RNA metabolism required for motor neuron development and survival (Kolb SJ et al. 2007).

**UBE1**

Ubiquitin is best known for its function in targeting proteins for degradation by the proteasome. In all tissues, the majority of intracellular proteins are degraded by the ubiquitin proteasome pathway (UPP) (Rock KL et al. 1994). The UPP consists of the concerted actions of enzymes that link chains of the polypeptide co-factor, Ubiquitin (Ub), onto proteins to mark them for degradation. This tagging process leads to their recognition by the 26S proteasome, a very large multicatalytic protease complex that degrades ubiquitinated proteins to
small peptides (Baumeister W et al. 1998) (Figure 3.1.). Three enzymatic components are required to link chains of Ub onto proteins that are destined for degradation. E1 ubiquitin-activating enzymes catalyze the first step in the ubiquitylation cycle, the ATP-dependent formation of a thioester bond between the C-terminal glycine residue of ubiquitin and the active site cysteine of the E1 (Haas and Siepmann 1997). This is followed by the transfer of ubiquitin from the E1 to a ubiquitin-conjugating enzyme (E2). The final step is the conjugation of ubiquitin to target proteins mediated by ubiquitin ligases (E3) (Ciechanover A & Schwartz AL 2004). The specificity of the ubiquitylation process is conferred by specific E3 ligases, because they recognize a specific protein substrate and catalyze the transfer of activated Ub to their respective protein substrate. There are at least two E1, over 30 E2 and approximately 1000 E3 enzymes in humans. Avram Hershenko, Aaron Ciechanover and Irwin Rose were awarded the Nobel Prize in Chemistry in 2004 for their discovery of Ub and the biochemistry of its conjugation to substrate proteins.

Recent studies have revealed several new functions of ubiquitin that are independent of proteasomal degradation. These functions include the novel signaling roles of ubiquitin in DNA repair, growth, apoptosis and the activation of protein kinases (Figure 3.1.) (Sun L, Chen ZJ 2004).
Figure 3.1. The ubiquitin pathway. Free ubiquitin (Ub) is activated by the ubiquitin-activating enzyme (E1), which forms a complex with ubiquitin. Subsequently, ubiquitin is transferred to one of many distinct ubiquitin-conjugating enzymes (E2s). Some E2s can directly ubiquitylate substrates, whereas others require the help of ubiquitin ligases (E3s). Some E3s function catalytically, whereas other E3s support ubiquitylation by recruiting substrates to the ubiquitylating enzymes. Multiubiquitylation serves mainly to label the substrate for degradation, whereas monoubiquitylation regulates several processes, such as DNA repair and transcriptional regulation.
E1 is the common first step in ubiquitylation, whether or not the modified protein is ultimately degraded by the proteasome (Yang Y et al. 2007). Until recently, it was presumed that a single E1 (UBE1, ubiquitin activating enzyme E1) activated ubiquitin. A second E1 was identified in 2007, UBA6 (also known as UBE1L2 (ubiquitin-activating enzyme E1-like 2) (Jin J et al. 2007). Interestingly, some E2’s were charged by UBA6 and UBE1 with equal efficiencies, the two E1’s showed different specificities. Almost half of tested E2’s were charged by UBE1 only, whereas a previously uncharacterized E2 was identified as the first UBA6-specific E2 (Jin J et al. 2007). These findings show that there is a variation to the first step of the ubiquitylation process and UBE1 is not the only E1 enzyme as it was considered before.

Materials and Methods

Western Blots

Transformed cell lines (lymphoblasts) were established from patients who come from different families. I had 2 lymphoblast cell lines from 2 affected boys of two different families (Family 2 and 5). For experimental controls, I ordered one SMA Type I and one age/gender matched healthy control cell line from the Coriell Cell Repository. All of these cell lines were kept in Dr. Fan’s lab in Cytogenetics and in Dr. Lampidis’ Lab in Cell Biology where I personally took care of the cell lines.

I performed total protein extraction from all cell lines when they were confluent. Lysis buffer included 0.1 mM Tris-HCl (pH: 7.4), 1% Sodium Dodecyl
Sulfate (SDS), ddH₂O, protease and phosphatase inhibitiior cocktail from Sigma. Protein concentrations were measured by the Pierce Micro BCA Protein Assay Kit.

I used precast gels purchased from BioRad (4%-10,12 or 15%). Whole cell protein lysates were loaded into the gel lanes (concentrations of protein samples were taken into account so that every lane had similar protein concentration). The running buffer consisted of 25mM Tris, 192 mM Glycine and 1% Sodium Dodecyl Sulfate (SDS). After the electrophoresis at 100V for 2.5 hours at room temperature is finished (all the dye had run off the gel), proteins were transferred to a nitrocellulose membrane at 100 V for 2 hours (transfer buffer : 2.6 mg Tris, 11.7 mg Glycine, 180 ml Methanol and 620 ml ddH₂O). The blot was incubated with a blocking buffer (5% non-fat dry milk in Tris-Buffered Saline (TBS)) 2 hours at room temperature. Following that, the blot was incubated with a 1:1000 dilution of primary antibody in Tris-Buffered Saline Tween-20 (TBS-T) overnight at 4°C (All the primary antibodies were obtained from Abcam.). Next, blots were washed 3 times with TBS-T (each for 10 min) is performed. After that, the blot was incubated with secondary antibody (Pierce ImmunoPure anti-mouse IgG-HRP (Horse Radish Peroxidase)) conjugate, 1:10000 dilution) for 1 hour at room temperature. Following the 3 X 10 minute washes with TBS-T, Pierce SuperSignal West Femto Maximum Sensitivity Substrate was added. Finally, an X-Ray exposure was done for different time periods depending on the experiment (1 sec-1min). Western Blot results were quantified by ImageQuant TL (GE Health Sciences).
RNA Expression

RNA was extracted from cell lines using Gentra Systems’ Versagene RNA Extraction Kit. Concentration and quality of RNA were measured with spectrophotometer (Eppendorf). Applied Biosystems’ GeneAmp RNA PCR Kit was used for Reverse Transcriptase-PCR. After this PCR step, the PCR products (cDNA) were used in ABI’s Taqman® RNA Expression Assays. For internal control, 18S RNA assay was used. The comparative $c_t$ method was used to analyze the data obtained.

Comparative $C_T$ method

(modified from http://www.uic.edu/depts/rrc/cgf/realtime/data.html)

The equation that describes the exponential amplification of PCR is:

$$X_n = X_0 (1 + E_X)^n \quad (1)$$

where $X_n$ is the number of template copies at cycle $n$, $X_0$ is the initial number of template copies, $E_X$ is the efficiency of target amplification, and $n$ is the number of cycles.

The threshold cycle ($C_T$) indicates the fractional cycle number at which the amount of amplified copies reaches a fixed threshold.

$$X_{T} = X_0 (1 + E_X)^{C_{T,X}} \quad (2)$$

where $X_T$ is the threshold number of copies and $C_{T,X}$ is the threshold cycle.
A similar equation for the endogenous reference (housekeeping gene like GAPDH, β-actin, or rRNA) is given by equation 3.

\[ R_T = R_0 (1 + E_R)^{C_{T,R}} \] (3)

where \( R \) is denoted as the quantities corresponding to the endogenous reference gene.

For normalization of the gene of interest to the endogenous reference, we can derive the following ratio giving the ratio of copy number of the gene of interest to the endogenous reference gene at the threshold cycle (\( K \)).

\[ K = \frac{X_T}{R_T} = \frac{X_0 (1 + E_X)^{C_{T,X}}}{R_0 (1 + E_R)^{C_{T,R}}} \] (4)

Assuming the efficiencies of the gene of interest amplification and the endogenous reference are equal (\( E = E_X = E_R \)) we can further derive equation 4 to

\[ X = \frac{X_0}{R_0} \left( \frac{X_T}{R_T} \right) (1 + E)^{-\Delta C_T} \] (6)

\[ X = K (1 + E)^{-\Delta C_T} \] (7)

where \( \Delta C_T \) is equal to \( C_{T,X} - C_{T,R} \)

To compare the relative expression of our gene of interest in our test sample to a control sample, we use the following equation
Further rearranging equation 7, we have

\[
\frac{X_{\text{test}}}{X_{\text{control}}} = \frac{X_{0,\text{test}}R_{0,\text{control}}}{R_{0,\text{test}}X_{0,\text{control}}} = \frac{K(1 + E)^{-\Delta C_{T,\text{test}}}}{K(1 + E)^{-\Delta C_{T,\text{control}}}}
\]  

(7)

Where \( \Delta C_T \) is the difference in threshold cycles for the target and control samples. For amplicons designed and optimized according to ABI (amplicon size < 150 bp), the efficiency is close to one (\( E = 1 \)). Therefore the amount of target normalized to an endogenous reference and relative to a control sample is given by

\[
\frac{X_{\text{test}}}{X_{\text{control}}} = (1 + E)^{\Delta C_{T,\text{control}} - \Delta C_{T,\text{test}}}
\]  

(8)

Where \( \Delta \Delta C_T \) is the difference in threshold cycles for the target and control samples. For amplicons designed and optimized according to ABI (amplicon size < 150 bp), the efficiency is close to one (\( E = 1 \)). Therefore the amount of target normalized to an endogenous reference and relative to a control sample is given by

\[
\frac{X_{\text{test}}}{X_{\text{control}}} = 2^{\Delta \Delta C_T} = 2^{\Delta C_{T,\text{control}} - \Delta C_{T,\text{test}}}
\]  

(9)

\[
\frac{X_{\text{test}}}{X_{\text{control}}} = 2^{\Delta \Delta C_T} = 2^{(C_{T,X} - C_{T,R})_{\text{control}} - (C_{T,X} - C_{T,R})_{\text{test}}}
\]  

(10)

Where \( C_{T,X} \) is the threshold cycle of the gene of interest and \( C_{T,R} \) is the threshold cycle of the endogenous reference gene (e.g. 18S RNA).
Results

SMN Complex (SMN, Gemin2 and Gemin3) Western Blots

Western Blot results for lymphoblasts show that XL-SMA may have a similar pathway deficiency as in autosomal SMA. Significant decreases of SMN and Gemin3 protein levels can be seen for both of the XL-SMA patients (they are from different families), but there is no difference for Gemin2 protein. For SMN, XL-SMA patients have a severe decrease of their protein level but not as high as SMA Type I patient, whereas for Gemin3 XL-SMA patients have similar level of decrease as SMA Type I patient (Figure 3.2., Figure 3.3, Table 3.1). For Gemin2, XL-SMA patients have no difference compared to healthy control, however the SMA Type I patient has a decrease of protein level of Gemin2 as well (Figure 3.2., Figure 3.3., Table 3.1). It should be also noted that both of the cell lines of XL-SMA patients, who are from different families, have similar behavior through all experiments regarding their levels of specific proteins. This observation is consistent with the notion that both of these patients have the same disease.
Figure 3.2. Western blot analysis of protein content of SMN, Gemin2 and Gemin3 in lymphoblastoid cell cultures of XL-SMA, SMA type 1 and healthy individuals. Gemin3 and SMN1 protein levels are lower (50% and 62% of healthy cell, respectively) in SMA1 and XL-SMA patients where XL-SMA patients have more SMN1 & Gemin3 than protein levels of SMA patient, but less than the healthy individual. For Gemin2, however, XL-SMA patients seem to have same amount of this protein in the lymphoblastoid cell lines as a healthy individual.
Table 3.1. SMN Complex Protein Studies. Every experiment was repeated three times and averages of relative intensities compared to wildtype protein levels were calculated with standard deviations. SMN and GEMIN3 protein levels in XL-SMA patients are lower compared to healthy cell lines. GEMIN2 protein levels in XL-SMA patients are same as healthy cell line. SMA type I cell line protein levels are decreased for all three SMN Complex proteins.

<table>
<thead>
<tr>
<th></th>
<th>XL-SMA_1</th>
<th>XL-SMA_2</th>
<th>SMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMN</td>
<td>64,70,61</td>
<td>60,55,62</td>
<td>25,19,34</td>
</tr>
<tr>
<td>Avg</td>
<td>65</td>
<td>59</td>
<td>26</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>3.7</td>
<td>2.9</td>
<td>6.2</td>
</tr>
<tr>
<td>GEMIN2</td>
<td>98,92,88</td>
<td>101,98,89</td>
<td>65,55,71</td>
</tr>
<tr>
<td>Avg</td>
<td>93</td>
<td>96</td>
<td>64</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>4.1</td>
<td>5.1</td>
<td>6.6</td>
</tr>
<tr>
<td>GEMIN3</td>
<td>53,60,44</td>
<td>47,39,55</td>
<td>33,51,29</td>
</tr>
<tr>
<td>Avg</td>
<td>52</td>
<td>47</td>
<td>37</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>6.6</td>
<td>6.5</td>
<td>9.6</td>
</tr>
</tbody>
</table>
Figure 3.3. SMN Complex Protein Levels. The healthy cell line protein levels were considered as 100.

Results indicate that XL-SMA patients have a decrease in their SMN and Gemin 3 protein levels compared to a healthy control. Additionally, this decrease is similar to what is seen with SMA Type I cell lines. This can lead us to believe that XL-SMA may have a deficiency or a defect related to SMN Complex. This defect can be transportation, stabilization or a cofactor defect. It is pure speculation at this point to explain what is wrong in XL-SMA; however these results are significant enough to consider a similar disease mechanism. Table 1 displays the normalized quantitative measurements (in reference to wild type) of protein levels of the blot shown in Fig 1. The two XL-SMA cell lines have
approximately (in average) 62% and 50% of SMN and Gemin3, respectively, in comparison to the healthy cell line (WT). In contrast, Gemin2 levels are approximately the same in these three cell lines. As expected based on published literature, all three of the proteins are reduced in the SMA Type 1 cell line (Helmken C et al. 2003).

**SMN and Gemin3 RNA Expression**

I performed RNA expression studies for the two XL-SMA lymphocyte cell lines from Family 2 and Family 5, respectively. The normalization was performed using 18S RNA as an internal control (Figure 3.4.). These studies were performed for SMN and Gemin3, and not for Gemin2, because the Gemin2 protein level was not significantly different in XL-SMA compared to a healthy control (Figure 3.2, Figure 3.3). Tables 3.2 and 3.3 display the results from three independent RT-PCR experiments for SMN and Gemin3, respectively. Standard deviations of these experiments are also shown. According to these results, there is no significant difference in gene expression of SMN and Gemin3 in XL-SMA and SMA cell lines compared to the healthy control. Moreover, the similar expression pattern of XL-SMA patient cell lines, which is a basically healthy RNA expression levels of SMN and Gemin 3, also indicate that a posttranscriptional mechanism is involved in controlling (decreasing) SMN and Gemin3 levels in these cell lines.
**Figure 3.4. Raw data from RNA Expression reaction.** RNA expression reaction is based on Taqman technology. This plot shows the increase in fluorescence per reaction cycle. As seen, there are three main groups, the internal control (18S RNA), the RNA in question (SMN) and negative controls (water control). Cycle threshold values are given as an Excel report, and RNA expression comparison is made through the cycle threshold values. Every sample is run in triplicate and average values are calculated for every sample.
<table>
<thead>
<tr>
<th></th>
<th>XSMA_1</th>
<th>XSMA_2</th>
<th>SMA</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔCt_1</td>
<td>12.23</td>
<td>11.84</td>
<td>11.69</td>
<td>9.93</td>
</tr>
<tr>
<td>ΔCt_2</td>
<td>12.01</td>
<td>10.73</td>
<td>11.07</td>
<td>10.57</td>
</tr>
<tr>
<td>ΔCt_3</td>
<td>12.47</td>
<td>13.5</td>
<td>14.44</td>
<td>15.38</td>
</tr>
<tr>
<td>Av. ΔCt</td>
<td>12.24</td>
<td>12.02</td>
<td>12.4</td>
<td>11.96</td>
</tr>
<tr>
<td>St. Dev</td>
<td>0.19</td>
<td>1.14</td>
<td>1.46</td>
<td>2.43</td>
</tr>
<tr>
<td>ΔΔCt</td>
<td>0.28</td>
<td>0.06</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>2^{ΔΔCt}</td>
<td>1.21</td>
<td>1.04</td>
<td>1.36</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2. Comparison of RNA Expression levels of SMN gene in XL-SMA and SMA Type I cell lines to healthy control cell line. All of the reactions were performed in triplicate and their averages were calculated. ΔCt value = sample average Ct for the SMN RNA minus the sample average Ct for 18S RNA for all cell lines separately; ΔΔCt value = the difference between two ΔCt values, e.g. ΔCt of SMN in XL-SMA cell line 1 minus ΔCt of SMN in healthy control cell line.; 2^{ΔΔCt} = the fold difference between samples.
<table>
<thead>
<tr>
<th></th>
<th>XSMA_1</th>
<th>XSMA_2</th>
<th>SMA</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔCt_1</td>
<td>7.27</td>
<td>6.41</td>
<td>6.96</td>
<td>6.29</td>
</tr>
<tr>
<td>ΔCt_2</td>
<td>6.1</td>
<td>6.44</td>
<td>6.82</td>
<td>8.55</td>
</tr>
<tr>
<td>ΔCt_3</td>
<td>5.85</td>
<td>6.58</td>
<td>6.4</td>
<td>7.99</td>
</tr>
<tr>
<td>Av. ΔCt</td>
<td>6.41</td>
<td>6.48</td>
<td>6.73</td>
<td>7.61</td>
</tr>
<tr>
<td>St. Dev</td>
<td>0.62</td>
<td>0.07</td>
<td>0.24</td>
<td>0.96</td>
</tr>
<tr>
<td>ΔΔCt</td>
<td>-1.20</td>
<td>-1.13</td>
<td>-0.88</td>
<td></td>
</tr>
<tr>
<td>2^ΔΔCt</td>
<td>0.44</td>
<td>0.46</td>
<td>0.54</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3. Comparison of RNA Expression levels of Gemin3 gene in XL-SMA and SMA Type I cell lines to healthy control cell line. All of the reactions were performed in triplicate and their averages were calculated.

**UBE1 Western Blots**

Western blots for UBE1 protein was performed with protein extracted from cell lines of two affected boys from Family 2 and Family 5. Although the initial experiment did not show any difference in protein levels of UBE1 in these cell
lines compared to a healthy cell line, subsequent experiments showed an average of 20% decrease in affected cell lines. Moreover, these results are confirmed by an independent laboratory (Barrington B (NIH), personal communication). Therefore, it can be concluded that UBE1 protein is decreased in cells of affected individuals. As the only caveat, it is unfortunately impossible to confirm this in neuronal cells due to the fact that we don’t have neuron cells of affected individuals.

![Western blot analysis](image)

**Figure 3.5. Western blot analysis of UBE1 protein content in lymphoblastoid cell cultures of XL-SMA, SMA type 1 and healthy individuals (including an obligate control).** XL-SMA cell lines have lower UBE1 protein levels compared to healthy cell lines. Here, SMA and obligate control cell lines also have similar protein levels as the healthy cell lines.

**Discussion**

SMN Complex holds the key to understand the molecular basis of the pathogenesis of Spinal Muscular Atrophy. To this date, there have been many speculative theories, but the consensus is that low SMN levels in spinal muscular atrophy disruptions normal cellular RNA metabolism required for motor neuron development and survival. Therefore, it was only logical to investigate SMN and
two foundational proteins of SMN Complex in XL-SMA patient cell lines as XL-SMA pathogenesis closely resembles SMA pathogenesis.

First, Western blotting was used as a means to compare the protein levels of SMN, Gemin2 and Gemin3 of XL-SMA patient lymphoblast cell lines to levels of aforementioned proteins in healthy and SMA Type I cell lines. As a result, SMN and Gemin3 protein levels were decreased in XL-SMA patient cell lines compared to healthy cell line, albeit not as low as SMN Type I cell line. Gemin2 protein levels were similar to Gemin2 protein level of healthy cell line. These results suggest that there may be a common disease mechanism involving SMN Complex. The difference between Gemin2 results and Gemin3 results can be explained as Gemin2 and Gemin3 interact with SMN protein via different motifs and they are not interacting with each other. This suggests that it is possible that Gemin 2 and Gemin 3 can have different expression patterns without influencing each other. They also may have different turnover mechanisms. These results also may suggest that UBE1 can be involved in the interaction of SMN and Gemin3.

Second, as suggested by the results of the protein studies, I measured RNA expression levels of SMN and Gemin3 in the same cell lines used in Western blotting studies. The results of RNA studies show that the RNA expression levels of SMN and Gemin3 are similar in comparison to healthy cell lines. These results suggest that a post-transcriptional, translational or post-translational mechanism is involved in the decrease of the protein levels of SMN and Gemin3 in XL-SMA patient cell lines.
Third, UBE1 protein levels were measured in the aforementioned cell lines. A measurable, albeit low, decrease in UBE1 protein levels were observed in patient cell lines compared to a healthy cell lines. This result suggests that the mutations in the *UBE1* gene of these patients cause a problem which leads to a decrease in UBE1 protein.

In summary, XL-SMA patient cell lines have decreased SMN, Gemin3 and UBE1 protein levels to varying degrees. Not much is known of the interaction of UBE1 and SMN Complex. It is known that SMN is ubiquitylated (Chang HC et al. 2004), but there is no report about the ubiquitin pathway and Gemin3. There are two models I can suggest: 1) UBE1 is involved in degradation of a protein which downregulates SMN Complex (or a protein which stabilizes SMN Complex). When *UBE1* is mutated, the protein in question is not degraded and this results in excess downregulation of SMN Complex (maybe via a pathway involving SMN-Gemin3 interaction). 2) UBE1 and UBA6 interact with the proteins of SMN Complex as they monoubiquinate them for different cellular processes. When *UBE1* is mutated, UBA6 cannot compensate the deficiency of UBE1, which in turn disrupts normal cellular RNA metabolism required for motor neuron development and survival.
Chapter 4 – Discussion

The motor neuron diseases are a group of progressive neurological disorders that destroy cells that control essential muscle activity such as speaking, walking, breathing, and swallowing. The most common childhood motor neuron diseases are the autosomal recessive proximal spinal muscular atrophies associated with deterioration and destruction of anterior horn cells. X-linked Infantile Spinal Muscular Atrophy is a rare congenital disorder characterized by multiple joint contractures. The findings of electromyography and muscle biopsy in XL-SMA patients were consistent with autosomal recessive infantile spinal muscular atrophy.

The purpose of my dissertation was to further define the X-Linked Spinal Muscular Atrophy gene region, to refine it, and discover the XL-SMA gene. In addition to that, the gene product was to be investigated further to delineate the genotype-phenotype correlation, if possible. Most of these goals were achieved (albeit not all by me), and there is still ample room to continue with the XL-SMA studies.

The first linkage paper of the X-Linked SMA gene region described a 12 cM pericentromeric region between Xp11.3-q11.2, excluding a majority of the X chromosome (Kobayashi H et al. 1995). My first step was to be involved in refining this region with further SNP studies. These studies only helped to take the NDP gene (Norrie Disease Protein) off the list containing approximately 250 genes. Moreover, these studies further cemented Xp11.3-q11.2 region as the XL-
SMA region. There was a consensus that XL-SMA gene resided in this region of X-chromosome. These results were published in Genetics in Medicine (Dressman D et al. 2007).

While working on linkage studies, I was also involved in candidate gene screening. This was a two-way effort, as I continued to screen genes in Miami; our collaborators in Germany also screened genes. I scanned all the literature necessary for this study and I followed every update of Human Genome Project. These updates were crucial as every new build discarded genes from the region and added genes to the region. After carefully choosing the genes to screen, my procedure included exons, intron-exon boundaries, and if possible, 5’ and 3’ UTR regions. In some cases, I included whole introns between two exons when those introns were manageable to screen. In summary, I screened 12 genes by PCR. I designed (and re-designed if necessary) all the primers, optimized the reactions, analyzed the sequencing results. At the end of all the efforts, our collaborative team has found the XL-SMA gene, UBE1.

As described in Chapter 3, UBE1 is responsible for the first step of ubiquitylation of proteins in a cell. Until recently, UBE1 was believed to be the only E1 (the starter) enzyme in the ubiquitin cascade. However, that was proven not to be true (Yang Y et al. 2007). UBE1 is involved with ubiquitylation of many proteins and it is still a crucial protein.

Analogies can be drawn between SMN and UBE1: a) UBE1 is also ubiquitously expressed, b) the gene product is a part of a general cellular
mechanism which is required in every tissue, c) it is extremely important in development, d) mutations of this gene cause a neuronal defect in patients, e) it has a ‘body double’ which may not work as efficiently with some proteins. When we add up these complications, it should be obvious that we don’t have a linear genotype-phenotype correlation, but a convoluted one. It is still not known exactly why mutations of SMN cause SMA, so we may have a long road ahead to decipher UBE1’s involvement with XL-SMA.

To investigate a possible common biological mechanism between SMA and XL-SMA, I worked on determining levels of foundational proteins in SMN Complex in XL-SMA patient cell lines. In these experiments, I also used SMA cell lines as negative controls. I determined that SMN and Gemin3 protein levels were consistently lower in XL-SMA patient cell lines (lymphoblasts) compared to healthy cell line. At the same time, SMN and Gemin3 protein levels were higher in XL-SMA cell lines when compared to SMA type I cell line. These results imply that there may be a common disease mechanism. To understand if the SMN and Gemin3 levels decrease in RNA level, I performed RNA expression studies. These studies confirmed that there is no difference in expression in XL-SMA cell lines when compared to healthy cell lines. So, the postulation of a decrease of RNA levels of SMN and Gemin3 in XL-SMA patients was ruled out.

As for UBE1, I repeated the same experimental procedure for SMN Complex proteins with antibodies to UBE1 to determine if there is any decline of UBE1 protein levels in XL-SMA patient cell lines compared to a healthy cell line.
There was a decline in protein levels of UBE1. This result was also independently confirmed as explained in Chapter 3.

All in all, these studies determined that there may be a commonality of the disease pathways between SMA and XL-SMA. XL-SMA patient cell lines also showed a decline in UBE1 protein levels as expected.

**Conclusions**

In the last 5 years, it is determined that a) XL-SMA is caused by mutations in the ubiquitin activating enzyme (UBE1) gene, b) the cell lines from XL-SMA patients show decline in SMN and XL-SMA protein levels with no RNA expression perturbation, and c) the cell lines from XL-SMA patients show decline in UBE1 protein levels.

These findings are important but the whole picture does not emerge as clearly as I expected. First of all, the identification of UBE1 as XL-SMA disease gene was a total surprise as neither I nor our collaborators predicted this gene to be the disease gene. UBE1 protein serves a very general purpose in cells and it may be hard for us to determine the exact molecular mechanism of XL-SMA pathogenesis. On the other hand, now that we know that SMN Complex may be involved in XL-SMA, there is room for further investigation of the proteins interacting with SMN Complex and their relation to UBE1.

As I explained in Chapter 3, I propose two possible scenarios for a molecular mechanism in XL-SMA: 1) UBE1 involves in degradation of a protein which downregulates SMN Complex. When UBE1 is mutated, the protein in
question is not degraded and this results in downregulation of SMN Complex (maybe via a pathway involving SMN-Gemin3 interaction). (Figure 4.1) 2) UBE1 and UBA6 interact with the proteins of SMN Complex as they monoubiquinate them for different cellular processes. When UBE1 is mutated, UBA6 cannot compensate the deficiency of UBE1 (similar to the relation of SMN1 and SMN2), which in turn results in disruption of normal cellular RNA metabolism required for motor neuron development and survival (Figure 4.2.)

Although the first hypothesis is hard to test as there is no candidate for such a protein which downregulates SMN Complex, the second hypothesis may be tested as the experimental protocol in the Science paper can be followed (Jin J et al. 2007).
Figure 4.1. Degradation model of XL-SMA. In this model, a protein (or a group of proteins) is involved in degradation of SMN Complex (or some of its members). The same protein is also degraded by ubiquitination. The turnover of this protein is in equilibrium under normal conditions. When UBE1 protein is not stable, degradation of this protein stops. This, in turn, causes an excessive degradation of SMN Complex.
Figure 4.2. UBA6 Model of XL-SMA. UBE1 and UBA6 interact with the proteins of SMN Complex as they monoubiquinate them for different cellular processes. When UBE1 takes a hit, UBA6 cannot compensate the deficiency of UBE1.

Future Studies

In addition to what is being proposed to test the two models offered in this thesis, I am more intrigued with another aspect of UBE1 which can be tested in vivo and can be a definitive in vivo test for UBE1 involvement in XL-SMA if results come out as expected. Recently a cell-permeable inhibitor of UBE1 was discovered (Yang Y et al. 2007). This inhibitor, called PYR-41 is pitched as a new cancer therapeutic. UBE1 inhibitors provide proof of principle for the capacity to differentially kill transformed cells (Yang Y et al. 2007). This report is exclusively based on cell cultures. My proposition is to use this cell permeable inhibitor in animals and see if this inhibitor causes systemic reactions which resemble XL-SMA like symptoms. This would be a breakthrough as we will prove in vivo that
defects in UBE1 are the necessary and sufficient cause of XL-SMA. Additionally, this experimental setup may help us to solve the molecular correlation between the genotype and phenotype.

Moreover, other proteins of SMN Complex, which could not be investigated in this report (due to lack of commercial antibodies), should be included in the future studies. Not only those proteins in the complex, but also the proteins which interact with the whole complex should be taken into account. The only caveat here is that we don’t have neuron cell lines from the patients, but only lymphoblasts. However, the lymphoblast studies in SMA have been giving extremely valuable insight to molecular basis of SMA. In the light of this reality, I believe lymphoblasts will continue to help us with XL-SMA.
References


Baumbach L., Iannaconne S., Best B. X-Linked lethal infantile spinal muscular atrophy: From clinical description to molecular mapping. 1996. Neuromusc.Disorders: 1st International World Congress of Neuromuscular Disorders (Special Issue), vol. 6(5).


Golembe TJ, Yong J, Dreyfuss G. Specific sequence features, recognized by the SMN complex, identify snRNAs and determine their fate as snRNPs. Mol Cell Biol. 2005 Dec;25(24):10989-1004.


Yong J, Pellizzoni L, Dreyfuss G. Sequence-specific interaction of U1 snRNA with the SMN complex. EMBO J. 2002 Mar 1;21(5):1188-96.
Yong J, Golemb VJ, Battle DJ, Pellizzoni L, Dreyfuss G. snRNAs contain specific SMN-binding domains that are essential for snRNP assembly. Mol Cell Biol. 2004 Apr;24(7):2747-56.