# GLOBAL GENE PROFILING UTILIZING A CELL MODEL OF SPINAL MUSCULAR ATROPHY

by

Miho Maeda

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Biological Sciences

Fall 2012

**GLOBAL GENE PROFILING UTILIZING A CELL MODEL**

**OF SPINAL MUSCULAR ATROPHY**

by

Miho Maeda

Approved: _____

John H. McDonald, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Matthew E.R. Butchbach, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Randall L. Duncan, Ph.D.
Chair of the Department of Biological Sciences

Approved: _____

George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____

Charles G. Riordan, Ph.D.
Vice Provost for Graduate and Professional Education

## ACKNOWLEDGMENTS

I have been privileged to have had three wonderful advisors at different points of my study. First and foremost this thesis would not have been possible without my current mentor Dr. Matthew Butchbach, who although I was grandfathered into his lab, gave me the same attention and dedication as he showed to others. This project would never have progressed this far and be as developed if it wasn't for him. I would also like to thank Dr. Robert Mason, who was not only a temporary advisor for our SMA lab, but has continued to assist me as part of my committee. I am indebted to the countless advice and support he has given me during my time here.

I am very grateful to my committee members, Dr. Jia Song and Dr. John McDonald who has through their guidance and support that has shaped my project to where it is today.

I am thankful for my lab members both past and present for both their advice and support. Dr. Chia-Yen Wu has taught me many techniques including mES cell differentiation which was a big part of my project. Dr. Sharyn Rossi has assisted me in immunostaining of cells. Casey Lumpkin, Ashlee Harris, and Kevin Lozo have also been helpful in various aspects of my project.

I would like to thank Dr. Lee Rubin from Harvard University for providing the ES cells. I would also like to thank Bruce Kingham who processed the samples for RNA-Seq. I would like to convey a thank you to Dr. Suzanne McCahan who has helped in the analysis of the RNA-Seq data. I am also very grateful for Dr. Melinda

Duncan whose advice and assistance was much appreciated when I was stuck with my project.

A special thank you goes out to the SNERPS, particularly David Wu, Hunter Stitik, Sona Balasubramaniam, and Pratima Patil who have given me numerous advice and much needed assistance throughout my project. I also need to give thanks to many research assistants and post-docs in Rockland I at Nemours AI Dupont Hosptial for Children. Everyone has not only been supportive, but has given me guidance for many techniques as well as advices to carry with me through life. Furthermore, my time at both UD and Nemours would not have been as smooth if it wasn't for the help of Betty Cowgill and Kathy Sewell, and for that I am thankful.

A big thank you goes out to my friends and family without whom I would not have been able to get this far.

With this I would like to dedicate this thesis to two very important people in my life. The first is Dr. Wenlan Wang who was my first mentor when I came to UD. This project was her conception, and although she could not see it come to fruition, there would be no project without her. She was a great mentor who has taught me many things, not just about science, but about life as well. Although it is true that I would not be in this present lab with this project without Dr. Wang, I can also say that I would not be here in graduate school pursuing this field without the love of my life, Daniel Hess. Because he loved science, I too entered the field, and because he pursued research with such passion, I could not help, but love it too. He has been my source of inspiration and support throughout my education and career. And although he and his nephew Mason have recently moved on along with Dr. Wang, I hope that the three of them will continue to watch me as I move forward with my career.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| SMA | Spinal muscular atrophy |
| MN | Motor neurons |
| SMN | Survival motor neuron |
| snRNP | Small nuclear ribonucleoprotein |
| RNP | Ribonucleoprotein |
| UsnRNP | Uridine-rich small nuclear ribonucleoproteins |
| Sm protein | Smith protein |
| snRNA | Small nuclear ribonucleic acid |
| UNRIP | Unr-interacting protein |
| eGFP | Enhanced green fluorescent protein |
| hnRNP | Heterogeneous nuclear ribonucleoprotein |
| UTR | Untranslated region |
| P | Post natal day |
| ES cell | Embryonic stem cell |
| mES cell | Murine embryonic stem cell |
| RA | Retinoic acid |
| Shh | Sonic hedgehog |
| FACS | Fluorescence activated cell sorting |
| EB | Embryoid bodies |
| TBS | Tris buffered saline |
| PBS | Phosphate buffered saline |

| | |
|---|---|
| HRP | Horse radish peroxidase |
| Tuj1 | β-III tubulin |
| NeuN | Neuronal nuclei |
| CNS | Central nervous system |
| FPKM | Fragments per kilobase of transcript per million mapped reads |
| FGF | Fibroblast growth factor |

## ABSTRACT

Spinal muscular atrophy (SMA) is an autosomal recessive neurodegenerative disorder caused by mutations in the survival motor neuron gene (SMN). Despite understanding the genetic basis behind the diseases, questions still remain about the specificity of the disease; why are motor neurons selectively affected? Using a mouse embryonic stem (mES) cell model for severe SMA, our present study had two aims. The first aim was to differentiate ES cells into motor neurons (MNs) and to characterize the differentiated cells to test their identity. The second aim was to isolate RNA from the ES cell derived MNs and to study their gene expression pattern through next generation sequencing technology (RNA-Seq). Our results have found that the ES cells from a severe SMA mouse model can be induced to generate MNs. When the transcriptome of SMA cells were compared to control cells, we found distinct gene expression patterns. Pluripotency and cell proliferation markers were increased significantly in SMA cells, whereas control cells had higher expression in neuronal development markers. Taken together, our study suggests an overall reduction in MN differentiation in this specific mES cell model of SMA. These findings suggest that SMN reduction in mES cells affects processes critical for normal development and maintenance in motor neurons.

# Chapter 1

# INTRODUCTION

## 1.1 Spinal Muscular Atrophy

Spinal muscular atrophy (SMA) is an autosomal recessive neurodegenerative disorder, characterized by the degeneration of α-motor neurons (MN) in the anterior horn of the spinal cord, which leads to progressive muscle weakness and atrophy. SMA is the leading genetic cause of infant death; 1 in 5000 babies are born with the disease (Cusco et al., 2002) and between 1 in 25-50 people are genetic carriers of the disease (Ben-Shachar, Orr-Urteger, Bardugo, Shomrat & Yaron, 2011; Lyahyai, Sbiti, Barkat, Ratbi & Sefiani, 2012; Sugarman et al., 2012; Su et al., 2011). SMA is classified into three clinical childhood types based on severity and age of onset (Russman, 2007). Type I SMA (Werdnig-Hoffmann disease) is an acute form of the disease, which accounts for about 50% of patients with SMA (D'Amico, Mercuri, Tiziano & Bertini, 2011). The onset of the disease is before 6 months of age. Infants with SMA type I exhibit the most severe symptoms. It is characterized by an inability to sit without support, and the life expectancy of these infants is less than 2 years due to respiratory complications (Morrison, 1996). Type II SMA is also known as intermediate SMA. The onset of the disease is usually between 6 and 18 months of age. Patients with SMA type II have the ability to sit without support, but usually cannot walk without assistance (Morrison, 1996). Patients with type III SMA (Kugelberg-Welander disease) develop symptoms after 18 months of age (Morrison,

1996). Children diagnosed with this form gain the ability to walk independently, although some may lose this ability later in life.

## 1.2 Survival Motor Neuron Gene

In 1995, the SMA disease causing gene, the survival motor neuron (*SMN*) gene was mapped to chromosome 5q13 (Lefebvre et al., 1995). In humans, two nearly identical copies of the SMN gene exist on chromosome 5 as an inverted duplication. The *SMN* genes, the telomeric or *SMN1* gene and the centromeric or *SMN2* gene span nine exons and code for an identical protein (Lefebvre et al., 1995).

The two genes differ by only five nucleotides, but only one is functionally relevant. A single C-> T transition in an exonic splice enhancer in exon 7 of the *SMN2* gene is responsible for an alternative splicing, which excludes exon 7 (Lorson, Hahnen, Androphy & Wirth, 1999; Monani et al., 1999). This alternative splicing event results in a truncated mRNA and protein, may be ubiquitinated and rapidly degraded by the ubiquitin proteasome system (Burnett et al., 2009). *SMN2* produces approximately 10% of the functional protein that *SMN1* produces (DiDonato et al., 2001).

Loss of the *SMN2* gene can occur with no consequences in healthy individuals (Burghes, 1997). However, 95% of patients with SMA have a homozygous deletion or gene conversion of the *SMN1* gene, and therefore must rely on the *SMN2* genes present (Cobben et al., 1995; Rodrigues et al., 1995). All SMA patients have multiple copies of the *SMN2* gene (2-4 copies), and copy numbers are correlated with disease severity. Type I patients have two copies of the *SMN2* gene, while type II patients have three copies, and type III patients have three or four copies (Prior, 2010). To date, no patients have been identified with a complete lack of the *SMN* gene, which

suggests that *SMN* has essential housekeeping functions (Eggert, Chari, Laggerbauer & Fischer, 2006). While, the genetics behind SMA is understood, the mechanism that leads to SMA has yet to be tested.

### 1.2.1 Cellular Functions of SMN

The discovery of *SMN* brought up questions of its function. As the disease causing gene in SMA, it was predicted to perform neuronal functions. However, it was discovered that SMN is a 38 kDa protein that is ubiquitously expressed, with particularly high expression in motor neurons. It is found in the cytoplasm and nucleus of all cells. In 1996, SMN was found to be colocalized in the nucleus of HeLa cells in structures called gems (Gemini of coiled bodies) (Liu & Dreyfuss, 1996). Gems are not unique to HeLa cells, but were found to be common in eukaryotic cells and are associated with Cajal bodies. Many nuclear factors that are involved in transcriptional activity and assembly colocalize to Cajal bodies including small nuclear ribonucleoproteins (snRNPs) (Ogg & Lamond, 2002).

RNAs are found in nucleoproteins called ribonucleoproteins (RNPs). RNPs play roles in biogenesis and maintaining stability in RNAs. There exist many different types of RNPs which are involved in different functions in cells. Among RNPs is the spliceosomal uridine-rich small nuclear ribonucleoproteins (UsnRNPs), which play an essential role in pre-mRNA processing (Bleicheret & Baserga, 2010). The UsnRNP complex are composed of a heptameric ring core structure of Sm proteins (B/B', D1, D2, D3, E, F, and G), which bind several short non-coding transcripts called small nuclear ribonucleic acids (snRNAs) (Mattaj & De Robertis, 1985; Kambach et al., 1999). The biogenesis of snRNPs takes place in both the nucleus and the cytoplasm,

and the SMN complex has been implicated in facilitating the assembly (Matera, R. Terns & M Terns, 2007).

The SMN complex is composed of SMN, Gemins 2-8 and Unr-interacting protein (UNRIP). SMN contains a highly conserved Tudor domain that binds to Smn proteins, and a point mutation within the domain has been shown to prevent SMN interaction with Sm proteins by affecting the charge distribution within the binding site (Buhler, Raker, Luhrmann & Fischer, 1999; Selenko et al., 2001). Furthermore, the SMN complex has been shown to bind to UsnRNAs. Therefore the SMN complex functions to ensure that the correct Sm proteins are assembled on specific UsnRNAs (Cauchi, 2010). However, this function of SMN in UsnRNP assembly is required by all cells, and therefore still does not answer the question as to why motor neurons are specifically affected in SMA.

Studies have shown that SMN may have a function unique to motor neurons. Pagliardini and colleagues have found that in rat spinal cords SMN protein to associate with cytoskeletal elements in spinal dendrites and axons (Pagliardini et al., 2000). In support of these findings, neuronal cells transfected with an enhanced green fluorescent protein (eGFP)-SMN showed rapid, bidirectional movements that were cytoskeletal-based (Zhang et al., 2003). Furthermore, later studies by Zhang and his colleagues have found the presence of Gemin and SMN particles that do not colocalize in neuronal cells, as well as SMN-Gemin complexes that lack Sm proteins (H. Zhang et al., 2006). Taken together, it is suggested that SMN may play a unique neuron specific role independent of UsnRNP biogenesis.

Several studies have suggested that SMN may play a role in axonal trafficking. Transport of β-actin mRNA and protein into growth cones has been shown to be

essential in developing neurites (Bassell et al., 1998). SMN has been shown to interact with β-actin by binding to heterogeneous nuclear ribonucleoprotein-R (hnRNP protein R) (Mourelatos et al., 2001). hnRNP-R binds to the 3'-UTR (untranslated region) of β-actin mRNA and knockdown of either hnRNP-R or Smn in zebrafish lead to distrubances in axonal outgrowth (Glinka et al., 2010; McWhorter, Monani, Burghes & Beattie, 2003). In support of these findings, overexpressing either Smn or hnRNP-R increased axonal outgrowth in PC12 cells (Rossoll et al., 2003). Furthermore reduced β-actin and hnRNP-R mRNA and protein levels were seen in the distal axons and growth cones in motor neurons from a SMA mouse model (Rossoll et al., 2003). Although neuronal specific roles of SMN have been proposed, further studies are needed to determine the exact link between SMN and its role in the pathogenesis of SMA.

### 1.3 Mouse Model of SMA

Different animal models have been generated to study the function of SMN; among them the mouse models have been important in understanding SMA. *SMN* duplication is unique to humans, and therefore mice have only one *Smn* gene equivalent to the human *SMN1*. A homozygous disruption in the gene has been shown to be embryonic lethal (Schrank et al., 1997). Several mouse models have been generated to overcome this lethality (Hsieh-Li et al., 2000; Frugier et al., 2000; Le et al., 2005). One severe mouse model in particular phenotypically resembles type I SMA in humans (Monani et al., 2000). These mice were generated by introducing human *SMN2* onto the *Smn-/-* genetic background, thus producing mice that are *Smn* knockout with 2 copies of human *SMN2* (Monani et al., 2000). Pathological severity in mice were found to be correlated with SMN2 protein amount available (Hsieh-Li et al.,

2000; Monani et al., 2000). Mice with high copy SMN2 (eight copies) were able to rescue the SMA phenotype; whereas mice with low copy SMN2 (two copies) exhibited SMA like phenotypes (Monani et al., 2000). The low copy mice at birth appeared the same as the other littermates, but at post natal day 3 (P3), the pups exhibited signs of deterioration and death occurred by between P4-6. The short lifespan as well as the low frequency of pups that survive past P1 limits their use for mechanistic studies, and therefore an *in vitro* system was developed to overcome these challenges.

### 1.4 Mouse Embryonic Stem Cell Model of SMA

Embryonic stem (ES) cells are cells derived from the inner cell mass of blastocysts and are both pluripotent and immortal (O'Shea, 1999). Murine ES (mES) cells were found to be able to differentiate into first spinal progenitor cells, and ultimately into motor neurons through exposure to retinoic acid (RA) and Sonic hedgehog (Shh). This was shown through the activation of several neural progenitor and MN markers during the course of differentiation (Figure 1.1) (Wichterle, Lieberam, Porter, & Jessell, 2002). Motor neurons differentiated from mES cells were found to generate action potentials, and when co-cultured with muscle cells developed axons and synapses (Miles et al., 2004). However, exposure to RA and Shh generates a heterogenous pool of cells, and hence a method was needed to identify MNs.

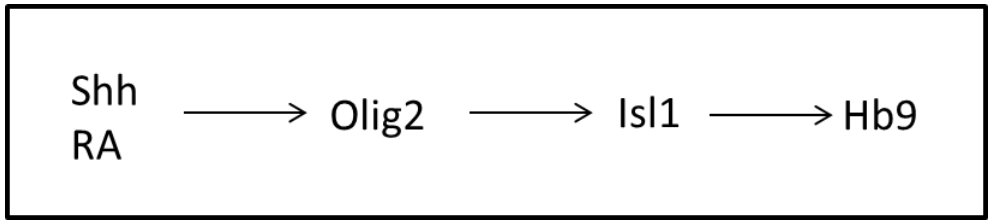


**Figure 1.1** Shh and RA induced pathway to generate motor neuron. Shh and RA activates the generation of neural progenitor markers, which leads to motor neuron differentiation. Modified from original (Wichterle, Lieberam, Porter, & Jessell, 2002)

The homeobox gene Hb9 (Mnx1 in mice) is a marker of MNs in developing spinal cords, and is necessary in the consolidation of the identity of MNs (Arber et al., 1999). Wichterle and his colleagues generated a transgenic mouse which expressed eGFP under the control of the mouse Hb9 promoter (Wichterle, Lieberam, Porter, & Jessell, 2002). When ES cells obtained from these mice differentiate into MNs, cells can be identified under a fluorescence microscope (Figure 1.2) and MN populations can be isolated by fluorescence activated cell sorting (FACS) (Wichterle, Lieberam, Porter, & Jessell, 2002). Thus MNs derived from mES cells of SMA mice can potentially provide insight into studying the pathogenesis of SMA.

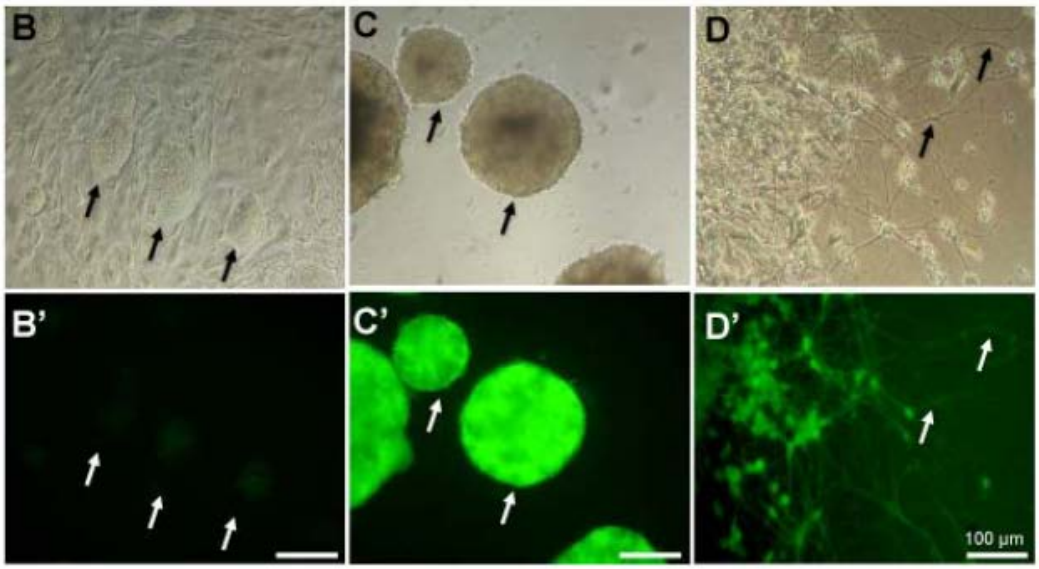


**Figure 1.2** mES cells expressing eGFP under Hb9 promoter. eGFP is expressed as mES cells differentiate into MNs. Modified from original (Wu et al., 2011)

## 1.5 Next-Generation Sequencing: A Novel Technology for Transcriptomics

RNA-Seq is a recently developed transcriptome profiling approach using deep-sequencing technology. It works by converting RNA into a cDNA library, attaching adapter sequencings to the ends, and then running the fragmented sequences through a high-thoroughput sequencing machine. Despite higher costs than microarrays, RNA-Seq has several advantages. The greatest advance of RNA-Seq is that it is not limited by existing knowledge of genomics as microarrays are, and therefore can be used to discover novel genes (Forrest & Carninci, 2009). Another advantage of RNA-Seq is that it directly reads the cDNA sequence. Microarrays in contrast rely upon an indirect method of measuring hybrization intensity, and therefore could result in high background noise, which are problematic when comparing samples (Morozova, Hirst & Marra, 2009).

The basic steps in RNA-Seq data analysis for determining differentially expressed genes are to first map the reads, then to assemble the transcriptome with the mapped reads, and to quantify the expression and calculate differential expression (Garber, Grabherr, Guttman & Trapnell). Many programs have been developed based upon different computational approaches for RNA-Seq data analysis. The free, open-source "Tuxedo Suite" comprising of the programs Bowtie, TopHat, and Cufflinks provide a pipeline to perform the necessary analyses. TopHat is a program built on the ultrafast short read aligner that can also discover transcript splice sites (Trapnell, Pachter & Salzberg, 2009). Cufflinks which have several built in subprograms can assemble the reads and calculate differentially expressed genes (Trapnell et al., 2010). Overview of the "Tuxedo Suite" pipeline is shown in Figure 1.3.

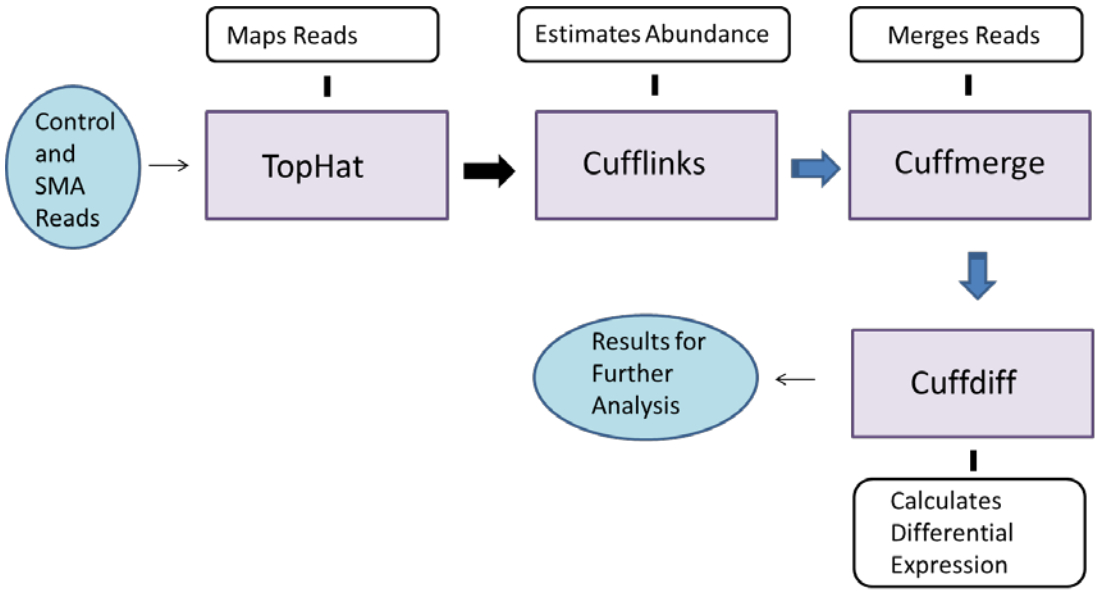


**Figure 1.3** Overview of Tuxedo Suite pipeline. Modified from original (Trapnell et al., 2012).

## 1.6 Hypothesis and Specific Aims

The objective of this thesis is to answer the key question: Why do reduced levels of a ubiquitously expressed protein, SMN result in selective motor neuron death and muscular atrophy? From previous studies conducted, we hypothesize that motor neurons have a unique gene expression pattern that leads to SMN deficiency susceptibility. In the present study, we differentiate, isolate, and characterize MNs generated from a mES model for SMA. We then run a global gene profiling to determine changes in gene expression patterns between wild-type and SMA MNs.

# Chapter 2

# MATERIALS & METHODS

## 2.1 ES Cell Culture

Hb9 (wild-type) and SMA-A2 murine ES cells were generously gifted to us from Dr. Lee Rubin of Harvard University. Murine ES cells were derived from a HB9::GFP (*mHB9-Gfp1b*) transgenic mouse crossed with either a wild-type (SMN2; Smn+/+) or SMA (SMA; Smn-/-) mouse (FVB.Cg-Tg(SMN2)89Ahmb*Smn*$^{tm1Msd}$/J) as previously reported (Wichterle, Lieberam, Porter, & Jessell, 2002). ES cells were grown as previously described (Wu et al., 2011). Briefly, mES cells were grown on a primary mouse embryonic fibroblast feeder layer (Millipore, Billerica, MA, USA) in 100mm tissue culture dishes. Cells were cultured with medium containing DMEm supplemented with 15% fetal bovine serum (Stem Cell Technology, Vancouver, BC, Canada), 1% GlutaMax-I supplement (Invitrogen, Carlsbad, CA, USA), 1% MEM non-essential amino acids, 1% nucleosides, 0.1 mM β-mercaptoethanol, 1% penicillin/streptomycin, and 1 ul/ml murine leukemia inhibitory factor (Millipore).

## 2.2 ES Cell Differentiation into MNs

Embryonic stem cells were differentiated as previously described (Wu et al., 2011). Briefly, ES cells were dissociated with 0.25% trypsin/EDTA and placed into a T75 cell culture flask coated with 0.1% gelatin (Stem Cell Technology). Fibroblasts were allowed to re-attach to the flask. The floating mES cells were collected and plated into a 100mm petri dish containing 10 ml of neural differentiation medium

containing DMEM supplemented with 15% FBS, 1% MEM nonessential amino acids, 1% GlutaMax-I supplement, 1% penicillin/streptomycin, 1 mM monothio-glycerol, 50ng/ml Noggin (Invitrogen), 20 ng/ml bFGF (Invitrogen), and 20 ng/ml FGF-8 (Invitrogen). Media was replaced daily. After two days, embryoid bodies (EB) were re-suspended in MN differentiation medium (NITSf) containing basal medium A (Stem Cell Technology) supplemented with 10% knockout serum replacement (Invitrogen), 1% N-2 supplement (Invitrogen), 1% ITS-B supplement (Stem Cell Technologies), 1% ascorbic acid (Stem Cell Technologies), 1% penicillin/streptomycin, 0.1% mM β-mercaptoethanol, 0.5% GlutaMax-I supplement, 30% D-glucose, 20 µg/ml heparin (Sigma, St. Louis, MO, USA) and 50 µg/ml fibronectin (Stem Cell Technologies) in the presence of 1µM RA (Sigma) and 1µM SAG (Calbiochem, Gibbstown, NJ, USA). Media was replaced daily for 5 days when EBs were collected, washed with phosphate-buffered saline (PBS), and dissociated in Accumax (Millipore). Cell aggregates were filtered through a BD cell strainer, and single cells were then suspended in NITSf medium supplemented with 10 ng/ml of each BDNF, GDNF, CNTF, and NT-3 and plated at a concentration of 1x $10^5$ cells per well in a 24 well plate that contained poly-DL-ornithine hydrobromide (Sigma)/laminin (Millipore)/matrigel (BD Bioscience, San Jose, CA, USA) coated coverslips.

### 2.3 Immunocytochemistry

Cells grown on coverslips were washed with PBS. Cells were fixed with 4% paraformaldehyde (Sigma) in PBS for 20 min. Cells were rinsed with PBS+ solution containing PBS, 0.1% Saponin (Sigma), 0.02% $NaN_3$ (Sigma), and blocked with PBS+BSA containing 100 mg bovine serum albumin (Sigma)/5 ml PBS+ for 30 min.

Cells were incubated overnight at 4ºC with mouse anti-Hb9 (1:10, Developmental Studies Hybridoma Bank , Iowa City, IA, USA), mouse anti-Tuj-1 (1:1000, Covance, Princeton, NJ, USA), mouse anti-Islet-1 (1:50, Developmental Studies Hybridoma Bank), mouse anti-Nestin (1:500, Millipore), or mouse anti-NeuN (1:200, Millipore) in PBS+BSA supplemented with rabbit anti-GFP (1:2000, Rockland Immunochemicals, Gilbertsville, PA, USA). After three washes with PBS+, cells were incubated for 1 hr at room temperature with Alexa Fluor 594 goat anti-mouse IgG (Invitrogen) and Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen). Cells were washed three times with PBS+ and incubated with Hoechst 33342 (Invitrogen) in $ddH_2O$ for 10 min and rinsed with $ddH_2O$ before mounting in Immu-Mount (GeneTex, Irvine, CA, USA). Images were obtained using a Leica TCS SP5 confocal microscope.

### 2.4 Western Blot Analysis

Cells were pelleted and lysed in a lysis buffer containing 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 1 mM EGTA, 1 mM sodium glycerolphosphate, 2.5 mM sodium pyrophosphate, 100 mM NaF, 10% glycerol, 1 mM sodium orthovanadate, 1 mM PMSF, 5µg/ml of aprotinin, 2 µg/ml of leupeptin . The lysates were sonicated and centrifuged at 13,200-rpm for 15 min. Protein concentration of the supernatants were analyzed with Pierce BCA Protein Assay Kit (Thermo Scientific, Rockford, IL, USA). 50µg of protein per well was separated in a 10% SDS-PAGE gel and transferred onto a Amersham Hybond-P PVDF membrane (GE Healthcare, Piscataway, NJ, USA) activated with methanol. The blots were blocked in 5% nonfat dry milk in Tris buffered saline with 0.1% Tween-20 (TBS-T) for 1 hr. The blots were then incubated overnight at 4ºC with mouse anti-SMN antibody (1:5000, BD Bioscience) diluted in 3% BSA in TBS-T. The following day,

the blots were washed three times with TBS-T and then incubated for 1 hr with horseradish peroxidase (HRP) conjugated sheep anti-mouse antibody (1:2500, GE Healthcare) diluted in 3% BSA in TBS-T. The protein on the membrane was detected by a ECL containing 250 mM luminol (Sigma), 90 mM p-coumaric acid (Sigma), 1M Tris-HCl, and pH 8.5, 0.0183% $H_2O_2$ (Sigma). The membrane was then stripped according to manufacturer's protocol with Re-Blot Plus Strong Solution (Millipore), and the procedure described above was followed using rabbit anti-GAPDH (1:20000, Sigma) as primary antibody and HRP goat anti-rabbit (1:2500, Rockland Immunochemicals) as secondary antibody.

### 2.5 Cell Sorting and RNA Isolation

Dissociated ES cells were filtered through a cell strainer, suspended in DMEM with 20µg/ml of propidium iodide and transported on ice to the Kimmel Cancer Center in Thomas Jefferson University to be sorted by a Coulter MoFlo cell sorter equipped with a 488nm laser. Cells were passed through a 100µm nozzle tip at a speed of approximately 12,500 events per second. Embryonic stem cells from wild-type mice described previously (Wu et al., 2011) were used as a negative control to set the cutoff for background fluorescence. Cells that were GFP positive, PI negative were collected in a 5 ml round bottom test tube (BD, Franklin Lakes, NJ, USA) with PBS supplemented with 60% FBS. Approximately a million cells were collected per sample. Sorted cells were centrifuged, PBS was removed, and pellets were snap frozen for RNA isolation. Total RNA was isolated from cells using the RNeasy mini kit (Qiagen, Germantown, MD, USA) with an additional DNase I (Qiagen) digestion step to remove any genomic DNA contamination.

## 2.6 RNA-Seq

RNA integrity was assessed by the Agilent Technologies 2100 Bioanalyzer. Concentration of isolated RNA was determined by a ND-1000 NanoDrop the spectrophotometer (NanoDrop Technologies, Wilmington, DE). 1 µg of RNA from each sample with three biological replicates was collected and sent to the Sequencing and Genotyping Center at the Delaware Biotechnology Institute in the University of Delaware for RNA-Seq library preparation and sequencing. The cDNA library was prepared according to the manufacturer using the TruSeq RNA Sample Prep Kit. The samples were then clustered and sequenced on an Illumina HiSeq 2000. The sequencing was performed on triplicates for each cell line (total six samples) for a 50 cycle single end run.

## 2.7 RNA-Seq Data Analysis

RNA-Seq reads were assessed for quality control by the FastQC software (version 0.10.1; Babraham Bioinformatics, Cambridge, UK). Adapter sequences and poly-A tails were removed from sequencing reads with the CutAdapt software (Martin, 2011; MIT, Cambridge, MA), and was confirmed by FastQC. Reads were mapped to a reference mouse transcriptome (GTF file) and genome (NCBIM37) using TopHat (version 2.04; Trapnell et al., 2009). The reference genome and annotation was obtained from Ensembl (http://www.ensembl.org). Transcripts were assembled and transcript abundances were measured as fragments per kilobase of exon per million fragments mapped (FPKM) by Cufflinks (Trapnell et al., 2010). The program Cuffdiff was then used for differential expression.

Differential expressed genes were plotted using the R package CummeRbund. The differentially expressed gene list was also submitted to the Ingenuity Pathway

Analysis (Ingenuity Systems, Inc. Redwood City, CA). Genes with a 4 fold or greater change in expression, p-value<0.01, and FPKM greater than 30 were loaded into the application.

### 2.8 cDNA Synthesis and qRT-PCR Validation

cDNA was prepared prior to setting up the qPCR reaction by iScript™ cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA) as per manufacturer's instructions. The cDNA was amplified via real-time polymerase chain reaction using the SYBR Green PCR Master Mix (Qiagen). Primers used are shown Table 2.1. Quantitative PCR was performed in a 384 well plate on a 7900HT Fast Real-Time PCR system (Applied Biosystems, Foster City, CA). Samples were assayed in triplicates and RNA levels were calculated by relative quantification (RQ) normalizing the samples to the endogenous control Cyclophilin A. Graphs represent the average of the relative mean expression level (RQ value) of three different experiments. The error bars represent the standard error of the RQ value (GraphPad Software, La Jolla, CA).

**Table 2.1** List of primers used in this work

| Gene Name | Forward Primer (5'->3') | Reverse Primer (5'->3') | Ampicon size (bp) |
|---|---|---|---|
| Crabp1 | CTTCAAGGTCGGAGAGG | GGAACAAGCTGGCCACC | 238 |
| Crabp2 | GGAGATTAACTTCAAGATCGGGGA | GCTAGTTTGTAAGATGGACGTGGG | 321 |
| Fabp7 | GGGTAAGACCCGAGTTCCTC | ATCACCACTTTGCCACCTTC | 213 |
| Gdf3 | ACCTTTCCAAGATGGCTCCT | CCTGAACCACAGACAGAGCA | 164 |
| Isl1 | CGGAGAGACATGATGGTGGTT | GGCTGATCTATGTCGCTTTGC | 109 |
| Klf2 | ACAGACTGCTATTTATT | CAGAACTGGTGGCAGAG | 78 |

| | | | |
|---|---|---|---|
| | GGACCTTAG | TCATTT | |
| Mybl2 | GTGAGGCAGTTTGGACAGCAA | GGATTCAAAACCCTCAGCCA | 101 |
| Nanog | AGATGCGGACTGTGTTCTC | TGCGTTCACCAGATAGCC | 281 |
| Nkx2.2 | CCTCCCCGAGTGGCAGAT | GAGTTCTATCCTCTCCAAAAGTTCAAA | 74 |
| Oct4/Pou5f1 | GCTCACCCTGGGCGTTCTC | GGCCGCAGCTTACACATGTTC | 102 |
| Olig2 | CAAATCTAATTCACATTCGGAAGGTTG | GACGATGGGCGACTAGACACC | 111 |
| Pla2g1b | CAGGCGCTGCTGCACACAG | GTCTAAGTCGTCCACTGGGGTGC | 158 |
| Rbp1 | CCAAAAATGCCTGTGGACTT | TGCACGATCTCTTTGTCTGG | 140 |
| Rtn1 | CCTGGTCCAGGACCTGGTG | AGCCACATGAGGACTGCAAAT | 51 |
| Smn1 | TGCTCCGTGGACCTCATTTCTT | TGGCTTTCCTGGTCCTAATCCTGA | 70 |
| Vim | GAATGACCGCTTTGCCAACTACAT | GCTTCCTCTCTCTGGAGCATCTCCT | 267 |
| Zic3 | CGAAGGCTGTGACAGACGGT | CATGTGCTTCTTGCGGTCG | 60 |

# Chapter 3

# RESULTS

## 3.1 mES Cells Differentiate into Motor Neurons *In Vitro*

In order to generate a supply of motor neurons, control and SMA embryoid bodies were cultured in the presence of RA and SHH for five days. At the end of the fifth day, cells were plated in the presence of neurotrophic factors to induce axonal growth and arborization. One noticeable difference after plating the cells was the number of GFP+ neuron-like cells present on the coverslips. Whereas the control lines had an abundance of GFP+ cells, the SMA cell line produced very few neuron-like GFP+ cells. The GFP+ control cells (Figure 3.1) and SMA cells (Figure 3.2) were characterized using various neuronal markers to test their motor neuron identity. As expected differentiated control and SMA cell lines expressed motor neuron specific markers Hb9 (Figures 3.1 A and 3.2 A) and Islet 1 (Figures 3.1B and 3.2 B) (Arber et al., 1999; Pfaff, Mendelsohn, Stewart, Edlund & Jessell, 1996). The presence of β-III tubulin (Tuj-1), a neuron specific tubulin (Figures 3.1 C and 3.2 C) as well as post-mitotic neuronal marker NeuN (neuronal nuclear antigen) (Figures 3.1 D and 3.2 D) confirmed that the GFP+ cells are indeed neurons. Nestin is a neural progenitor marker not expressed in post-mitotic MNs (Park et al., 2010). Staining with the nestin (Figures 3.1 E and 3.2 E) showed that the GFP+ cells do not express nestin, demonstrating that mES cell lines can be differentiated into GFP+ post-mitotic MNs.

SMN protein levels were tested through western blot analysis (Figure 3.3), reaffirming that the SMA cell line is knocked down for SMN. Thus we established

that the SMA cell line has reduced levels of SMN, and furthermore both the control and SMA mES cell can be induced to generate GFP+ MNs.

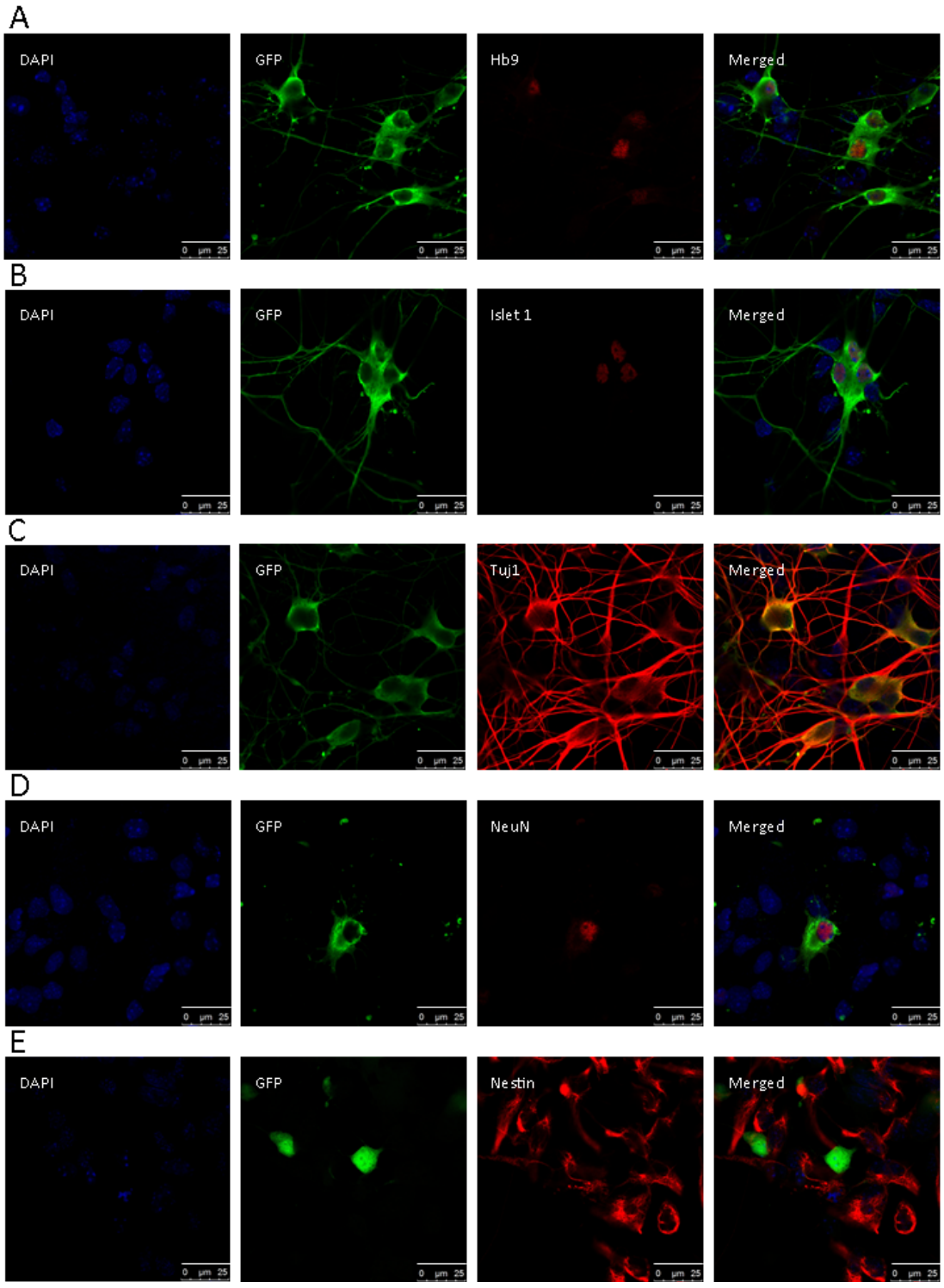
A
DAPI
GFP
Hb9
Merged
B
DAPI
GFP
Islet 1
Merged
C
DAPI
GFP
Tuj1
Merged
D
DAPI
GFP
NeuN
Merged
E
DAPI
GFP
Nestin
Merged

**Figure 3.1** Characterization of mES cell derived MNs. Differentiated control mES cells were dissociated, plated, and probed through indirect immunofluorescence for neuronal markers. (A and B) Expressions of motor neuron specific markers Hb9 and Islet1 were detected after 5 days of differentiation. (C and D) Panneuronal markers Tuj1 and NeuN were expressed. (E) Neural progenitor marker Nestin was not expressed in mature neurons.

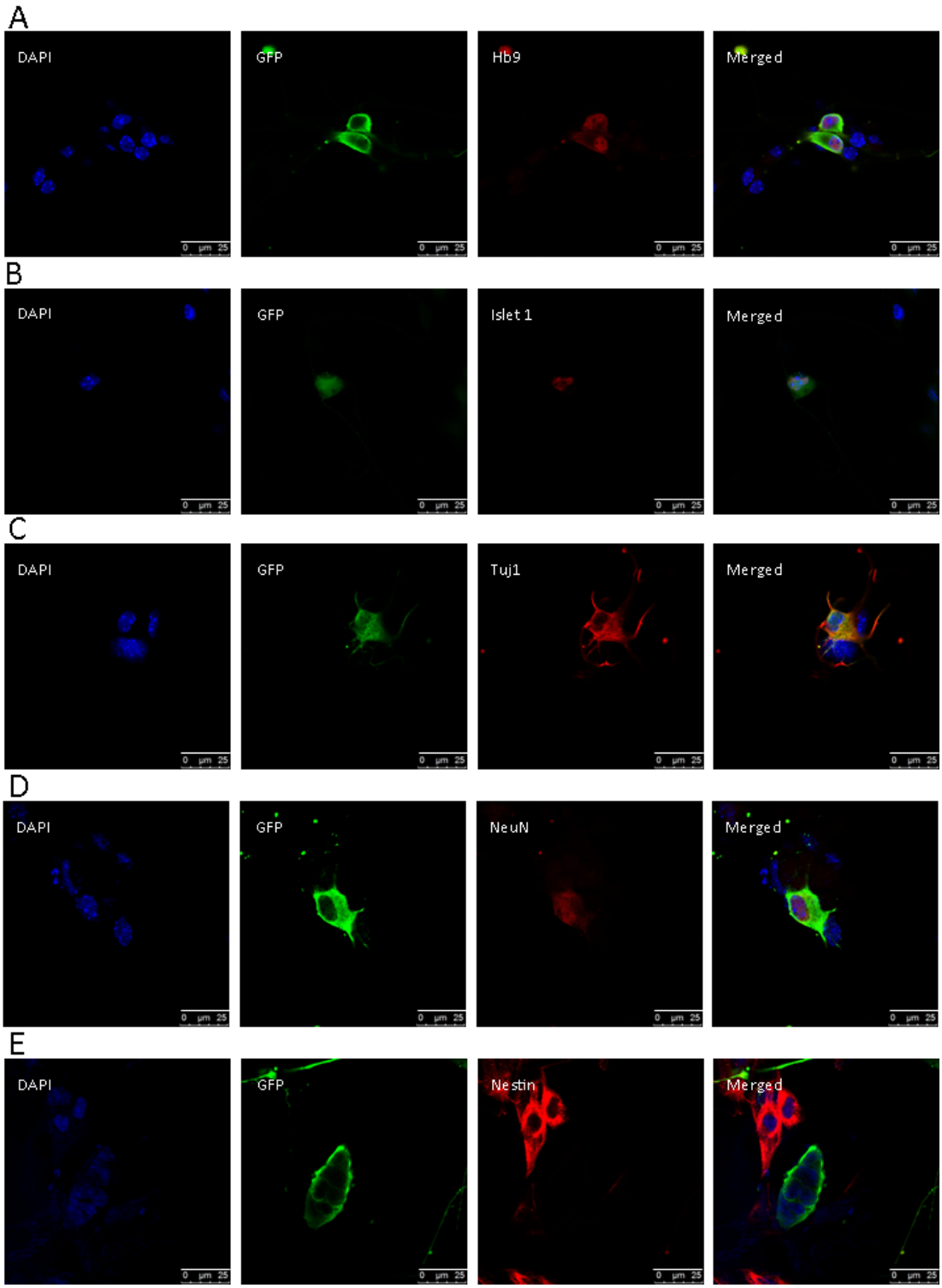
A
DAPI
GFP
Hb9
Merged
B
DAPI
GFP
Islet 1
Merged
C
DAPI
GFP
Tuj1
Merged
D
DAPI
GFP
NeuN
Merged
E
DAPI
GFP
Nestin
Merged

**Figure 3.2** Characterization of mES cell derived SMA MNs. Differentiated SMA mES cells were dissociated, plated, and probed through indirect immunofluorescence for neuronal markers. (A and B) Expressions of motor neuron specific markers Hb9 and Islet1 were detected after 5 days of differentiation. (C and D) Panneuronal markers Tuj1 and NeuN were expressed. (E) Neural progenitor marker Nestin was not expressed in mature neurons.

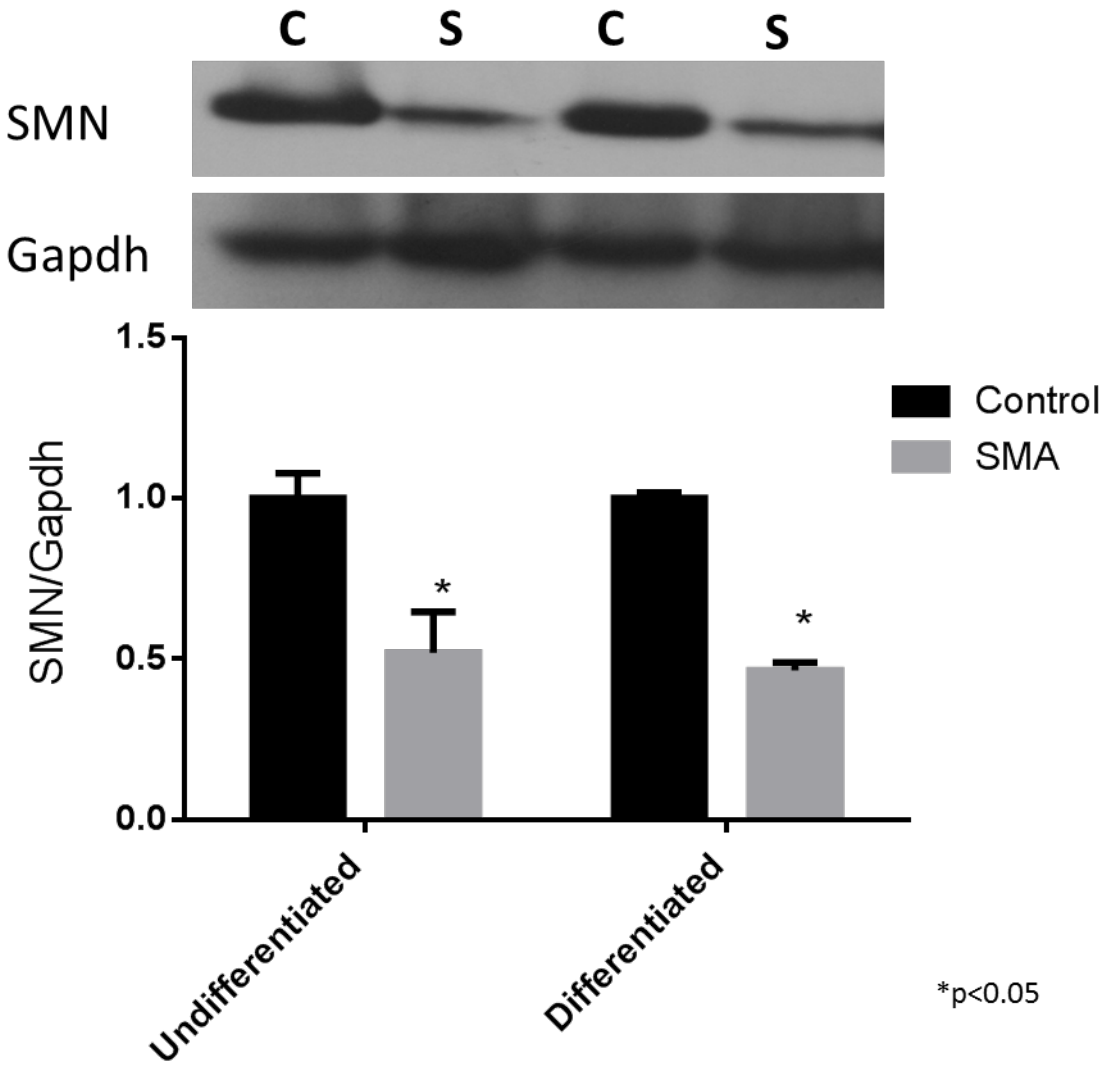


**Figure 3.3** Western blot analysis of differentiated and undifferentiated mES cell lines. Western blot analysis of both differentiated and undifferentiated mES cells revealed a significantly reduced protein expression of SMN in SMA cell lines compared to control. Triplicates of each sample was blotted and probed for SMN and Gapdh as a loading control. Blots were scanned and band intensity was analyzed by ImageJ. SMN levels were normalized to Gapdh. Data are expressed as means ± SD. U = undifferentiated, C = control, S = SMA. Asterisks indicate significant differences compared to controls (*p <0.05, Student's *t* test).

## 3.2 RNA-Seq Quality Analysis

The total number of reads produced from each sample after RNA-Seq was around 100 million reads per sample. Samples were then trimmed with the software Cutadapt (Martin, 2011) to remove adapters, polyA tails, and low quality ends. Quality of trimmed reads was checked through FastQC. Sequence quality scores as well as base quality scores were determined individually and sequence quality was found to be high with Phred quality scores around 39 (Figure 3.4 A and B). Average base quality scores were also high with Phred quality scores above 32 (Figure 3.4 C and D).

The trimmed reads were mapped to the mouse genome with TopHat. After trimming, the total number of reads per sample input into the program ranged between 29,763,880 and 44,570,352 (Table 3.1). Between 69% and 76% of reads were uniquely mapped to the reference genome. The difference in the average number of reads between control and SMA samples was not statistically significant (Student's t-test, $p = 0.22$).

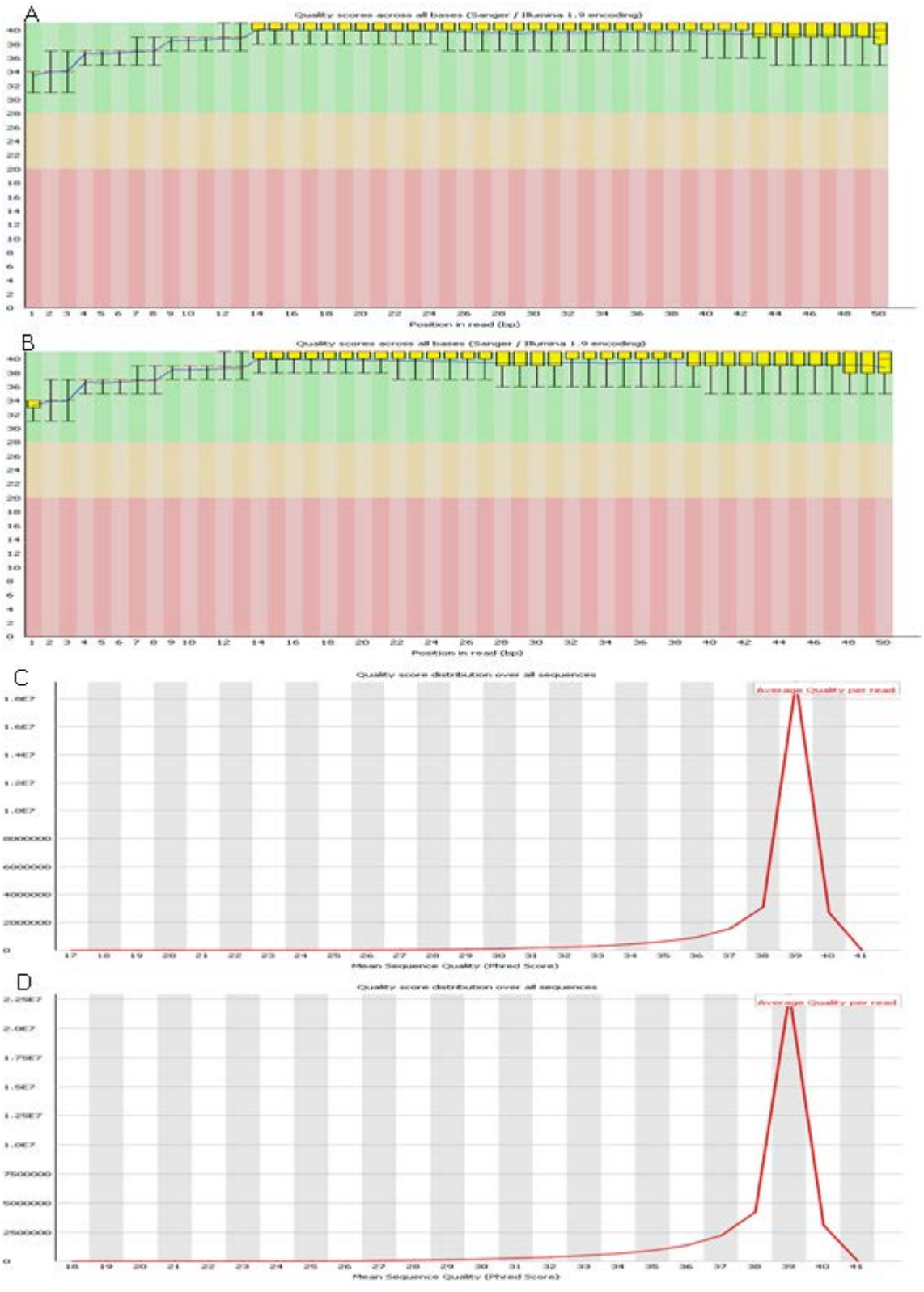
A
Quality scores across all bases (Sanger / Illumina 1.9 encoding)
Position in read (bp)
B
Quality scores across all bases (Sanger / Illumina 1.9 encoding)
Position in read (bp)
C
Quality score distribution over all sequences
Average Quality per read
Mean Sequence Quality (Phred Score)
D
Quality score distribution over all sequences
Average Quality per read
Mean Sequence Quality (Phred Score)

**Figure 3.4** Quality scores determined for all base positions and overall sequence quality using FastQC. Every control sample and SMA sample showed very similar trends. (A) Control sample quality scores across all base positions (B) SMA sample quality scores across all base positions (C) Control sample overall sequence quality (D) SMA sample overall sequence quality. According to Phred quality scores higher scores correspond to higher quality. A Phred quality score of 30 corresponds to a base call accuracy of 99.9%.

**Table 3.1** RNA-Seq reads mapped to NCBI mouse genome build 37 using TopHat

| | Control | | | SMA | | |
|---|---|---|---|---|---|---|
| | Sample 1 | Sample 2 | Sample 3 | Sample 1 | Sample 2 | Sample 3 |
| Total Reads | 29,763,880 | 14,863,496 | 36,877,629 | 37,203,080 | 32,327,505 | 44,570,352 |
| Reads Removed | 23.86% | 22.74% | 23.65% | 29.20% | 28.81% | 28.16% |
| Reads Mapped | 74.49% | 75.50% | 74.88% | 69.33% | 69.67% | 70.60% |

## 3.3 Differentially Expressed Genes

After reads were mapped, transcripts were assembled and transcript abundance measured in FPKM with Cufflinks. The expression levels were then input into Cuffdiff to test for differential expression. Overall, there were 41,945 expressed genes of which 10,058 genes were found to be statistically significant ($p<0.05$). Out of the 10,058 statistically significant differentially expressed genes, 3094 genes were up-regulated in SMA samples compared to control, and 6964 genes were down-regulated.

To reduce the number of genes from this list the threshold was set to filter data under 4 fold-change compared to the control, and for p-values above 0.01. However, even under these conditions, 286 genes were found to be up-regulated, and 814 genes were found to be down-regulated. Therefore, a density graph of all statistically significant genes was plotted using the CummeRbund package in the programming

language R to set a FPKM cut-off in order to reduce the number of significant genes (Figure 3.5).

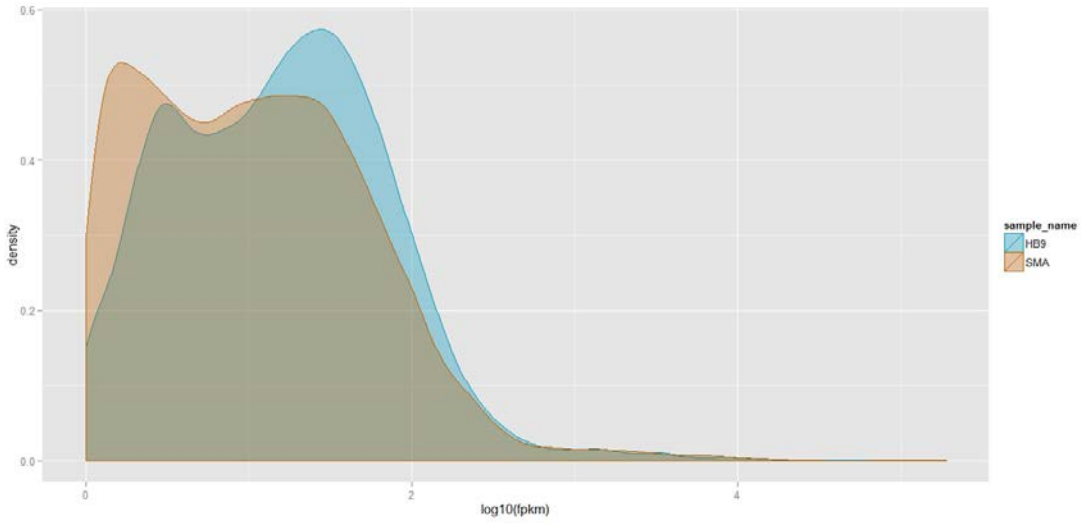


**Figure 3.5** Distribution of FPKM values of significant genes across samples. The density plot was made using the CummeRbund package in R. HB9 represents control samples in blue and SMA samples are represented in orange.

From the density plot, we had initially set the cut-off to 2 (FPKM of 100), however, we discovered that under these conditions only 5 genes were shown to be up-regulated in SMA versus control that had a minimum of 4 fold-change and p-value <0.01. Therefore, to reduce stringency, we arbitrarily set the cut-off to 1.5 (FPKM of 30). In the up-regulated gene list the SMA samples with FPKM lower than 30 were filtered out, and in the down-regulated gene list, control samples with FPKM lower than 30 were filtered out. Our final list has 27 genes that are up-regulated and 220 genes that are down-regulated.

Top 10 up- and down- regulated genes are presented in tables 3.2 and 3.3 respectively. The top 3 up-regulated genes were Lefty1, Nr0b1, and Zic3. All three genes coincidentally play a role in development. Lefty1 plays several roles in embryonic development, including determining left-right asymmetry with the help of Shh and RA (Tsukui et al., 1999). Nr0b1 encodes for a protein called Dax1. Dax1 is thought to play a role in the development of the hypothalamic-pituitary-adrenal-gonadal axis (Guo, Burris, & McCabe, 1995; Ikeda et al., 2001). Zic3, similar to Lefty1, also is implicated in several roles throughout development, among which the most widely studied is its role in left-right asymmetry (Purandare et al., 2002).

The top 3 down-regulated genes Hes5, Glra1, and Fabp7 are all expressed in the central nervous system (CNS). Hes5 is activated by Notch during development and helps regulate mammalian neuronal differentiation (Ohtsuka et al., 1999). Glra1 is the gene responsible for coding the alpha 1 subunit of the glycine receptor protein. The glycine receptor is an inhibitory receptor found in the spinal cord and brain stem. Finally Fabp7 also known as brain lipid binding protein is expressed in radial glia in the developing CNS (Feng & Heintz, 1995).

**Table 3.3** Top 10 up-regulated genes in SMA samples as compared to the control samples. The differentially expressed genes that fit the fold change, FPKM, and p-value criteria were ranked based on fold-change. The 10 highest fold changes are shown below with the gene name, symbol, FPKM values of both control and SMA, and fold change (log2).

| Entrez Gene Name | Gene Symbol | Control Intensity/FPKM | SMA Intensity/FPKM | Fold Change |
|---|---|---|---|---|
| left-right determination factor 1 | Lefty1 | 2.220 | 41.606 | 4.228 |
| nuclear receptor subfamily 0, group B, member 1 | Nr0b1 | 3.172 | 51.639 | 4.025 |
| Zic family member 3 | Zic3 | 8.862 | 76.782 | 3.115 |
| T-cell leukemia/lymphoma 1A | Tcl1 | 3.726 | 30.744 | 3.045 |
| phospholipase A2, group IB (pancreas) | Pla2g1b | 8.643 | 70.212 | 3.022 |
| POU class 5 homeobox 1 | Pou5f1 | 98.252 | 775.225 | 2.980 |
| growth differentiation factor 3 | Gdf3 | 9.742 | 70.405 | 2.853 |
| zinc finger protein 42 homolog (mouse) | Zfp42 | 11.504 | 79.924 | 2.796 |
| RAS guanyl releasing protein 2 (calcium and DAG-regulated) | Rasgrp2 | 5.330 | 33.840 | 2.666 |
| zinc finger and SCAN domain containing 10 | Zscan10 | 8.671 | 49.713 | 2.519 |

**Table 3.4** Top 10 down-regulated genes in SMA samples as compared to the control samples. The differentially expressed genes that fit the fold change, FPKM, and p-value criteria were ranked based on fold-change. The 10 lowest fold changes are shown below with the gene name, symbol, FPKM values of both control and SMA, and fold change (log2).

| Entrez Gene Name | Gene Symbol | Control Intensity/FPKM | SMA Intensity/FPKM | Fold Change |
|---|---|---|---|---|
| hairy and enhancer of split 5 (Drosophila) | Hes5 | 65.258 | 0.349 | -7.547 |
| glycine receptor, alpha 1 | Glra1 | 34.704 | 0.291 | -6.898 |
| fatty acid binding protein 7, brain | Fabp7 | 553.486 | 5.517 | -6.649 |
| fibroblast growth factor binding protein 3 | Fgfbp3 | 62.341 | 0.896 | -6.121 |
| microRNA 124-1 | Mir124a-2 | 36.735 | 0.653 | -5.814 |
| adenylate cyclase activating polypeptide 1 (pituitary) receptor type I | Adcyap1r1 | 94.598 | 1.892 | -5.644 |
| achaete-scute complex homolog 1 (Drosophila) | Ascl1 | 43.303 | 0.909 | -5.574 |
| oligodendrocyte lineage transcription factor 2 | Olig2 | 32.059 | 0.704 | -5.510 |
| NK2 homeobox 2 | Nkx2-2 | 52.529 | 1.333 | -5.301 |
| leucine rich repeat neuronal 1 | Lrrn1 | 114.560 | 2.997 | -5.256 |

## 3.4 Network and Pathway Analysis of Differentially Expressed Genes

The same list used to generate the top 10 up- and down- regulated genes were uploaded into Ingenuity Pathway Analysis (IPA). IPA creates networks of focus genes that are then algorithmically generated based on connectivity. Each network is assigned a score (p-value based) which represents how likely focus genes are found within a certain network by random chance. Networks with a score of 2 and higher have a 99% confidence that focus genes are not found together by random chance. Table 3.5 lists the top 3 networks affected by up- and down- regulated genes in SMA

samples. The top networks are consistent with the top 3 up- and down- regulation genes. The top networks observed to be up-regulated in SMA cells include cellular function and maintenance and cell death and survival. The top networks with genes down-regulated in SMA cells include nervous system development and function. Similarly, when the full lists of significantly up-regulated or down-regulated genes are analyzed through IPA we see similar trends. The 3094 up-regulated genes in IPA show the top network as cell cycle, embryonic development, organ development with a score of 45 and 21 genes. The 6964 down-regulated genes show the top functions as cell-to-cell signaling and interaction, nervous system development and function, gene expression with a score of 26 and 14 genes that fall into this network.

**Table 3.5** Top networks significantly influenced by up-regulated and down-regulated genes in SMA generated by IPA. Score is calculated based on p-value.

| **Up-regulated genes** | | |
|---|---|---|
| **Top Networks** | **Score** | **Genes** |
| Gene Expression, Cellular Function and Maintenance, Cell Death and Survival | 41 | 16 |
| Hereditary Disorder, Neurological Disease, Skeletal and Muscular Disorders | 30 | 5 |
| Post-Translational Modification, Cancer, Endocrine System Disorders | 3 | 1 |

| **Down-regulated genes** | | |
|---|---|---|
| **Top Networks** | **Score** | **Genes** |
| Nervous System Development and Function, Tissue Morphology, Cellular Growth and Proliferation | 58 | 29 |
| Cell Morphology, Cellular Assembly and Organization, Cellular Development | 42 | 23 |
| Nervous System Development and Function, Cell Morphology, Cellular Movement | 41 | 23 |

From the same generated lists as used in Table 3.5, IPA found canonical pathways that were significantly relevant to the data. Top canonical pathways of up- and down-regulated genes are shown in Table 3.6. Consistent with network observations, top canonical pathways affected by up-regulated genes in SMA are related to development and pluripotency. The top pathways affected by down-

regulated genes in SMA are relevant to the nervous system and play a role in neuronal development. In support of these findings, when all the significantly up- and down-regulated genes were submitted together into IPA, the top canonical pathway was found to the axonal guidance signaling pathway with 25.2% of genes falling into the pathway and with a p-value of 2.44E-10.

**Table 3.6** Top canonical pathways affected by up-regulated and down-regulated genes in SMA generated by IPA. Gene percentage is based on the ratio of focus genes present, out of total genes known to be involved within the pathway.

| **Up-regulated genes** | | |
|---|---|---|
| **Top Canonical Pathways** | **p-value** | **Genes** |
| Role of NANOG in Mammalian Embryonic Stem Cell Pluripotency | 4.07E-05 | 3.50% |
| Transcriptional Regulatory Network in Embryonic Stem Cells | 5.05E-05 | 7.50% |
| Mouse Embryonic Stem Cell Pluripotency | 6.62E+04 | 3% |
| | | |
| **Down-regulated genes** | | |
| **Top Canonical Pathways** | **p-value** | **Genes** |
| Notch Signaling | 5.37E-07 | 16.30% |
| Axonal Guidance Signaling | 3.25E-06 | 4.10% |
| Semaphorin Signaling in Neurons | 5.82E-04 | 9.60% |

## 3.5 qRT-PCR Validation

Validation of RNA-Seq results were conducted on the same 3 samples sent for RNA-Seq, which had not undergone any RNA amplification. Therefore, we would be validating the technology, but not the conditions (Fang & Cui, 2011). To validate RNA-Seq results, we planned to perform qRT-PCR on 20 significantly expressed genes from the list submitted to IPA. However, many of the top 10 up-regulated and top 10 down-regulated genes had low expression values (FPKM), and therefore another set of genes was chosen. From the list based upon a 4 fold change, p-value of

less than 0.01 and FPKM of 30 and above (either control or SMA samples), the top 10 highest expressing genes were initially chosen.

Due to the trends seen through IPA analysis, we also wanted to test whether mES cell differentiation occurred. Two genes expressed during different stages of MN differentiation were chosen (Olig2 and Islet1) along with Hb9 and Smn. However, given the limited samples, all 25 genes could not be analyzed, and therefore, two of the lowest expressing genes from the down-regulated genes list were replaced with the two MN differentiation markers. 22 genes in total were initially tested, however due to sub-optimal primer conditions, only 17 genes (Table 3.7) were able to yield measurable results. Inititally we had a list of 8 housekeeping genes, however, after comparison against RNA-Seq results, 6 of these genes were found to be significantly altered. Therefore the 17 genes were normalized against peptidylprolyl isomerase A (Cyclophilin A).

All 17 genes quantified by qRT-PCR showed significant changes in SMA samples versus control ($p<0.05$). Fold changes in expression were graphed in log2 consistent with RNA-Seq fold changes calculated in log2. A gene expression difference between RNA-Seq and qRT-PCR results was considered valid if the overall trend in the change of a gene between the two methods corresponded (Figure 3.6). Despite differences in values, overall trends within the 17 genes were in found to be in agreement.

**Table 3.7** Genes used in qRT-PCR

| Gene | Locus | FPKM Control | FPKM SMA | Log2(fold_change) | p_value | q_value |
|---|---|---|---|---|---|---|
| Crabp1 | 9:54612554-54620918 | 1134.32 | 88.2872 | -3.68347 | 0 | 0 |
| Crabp2 | 3:87752587-87757298 | 230.989 | 20.955 | -3.46245 | 0 | 0 |
| Fabp7 | 10:57504686-57508256 | 553.486 | 5.51666 | -6.64861 | 0 | 0 |
| Gdf3 | 6:122555420-122583593 | 9.74246 | 70.4049 | 2.85332 | 0 | 0 |
| Isl1 | 13:117088479-117099896 | 107.991 | 5.9504 | -4.18178 | 0 | 0 |
| Klf2 | 8:74842931-74845555 | 21.6389 | 109.051 | 2.33331 | 0 | 0 |
| Mybl2 | 2:162880292-162910433 | 58.3267 | 261.506 | 2.16462 | 0 | 0 |
| Nanog | 6:122657506-122664651 | 41.4061 | 235.718 | 2.50915 | 0 | 0 |
| Nkx2-2 | 2:147003281-147157417 | 52.5286 | 1.33272 | -5.30066 | 0 | 0 |
| Olig2 | 16:91225794-91228922 | 32.0593 | 0.703598 | -5.50985 | 0 | 0 |
| Pla2g1b | 5:115915108-115924731 | 8.64317 | 70.2122 | 3.02209 | 0 | 0 |
| Pou5f1 | 17:35642981-35647721 | 98.2525 | 775.225 | 2.98005 | 0 | 0 |
| Rbp1 | 9:98323379-98346986 | 306.808 | 50.4593 | -2.60415 | 0 | 0 |
| Rtn1 | 12:73312738-73510041 | 218.173 | 18.4291 | -3.56541 | 0 | 0 |
| Smn1 | 13:100894806-100907645 | 136.519 | 5.26433 | -4.69671 | 0 | 0 |
| Vim | 2:13495553-13504453 | 1435.96 | 177.615 | -3.01519 | 0 | 0 |
| Zic3 | X:55275876-55294913 | 8.8625 | 76.7816 | 3.11497 | 0 | 0 |

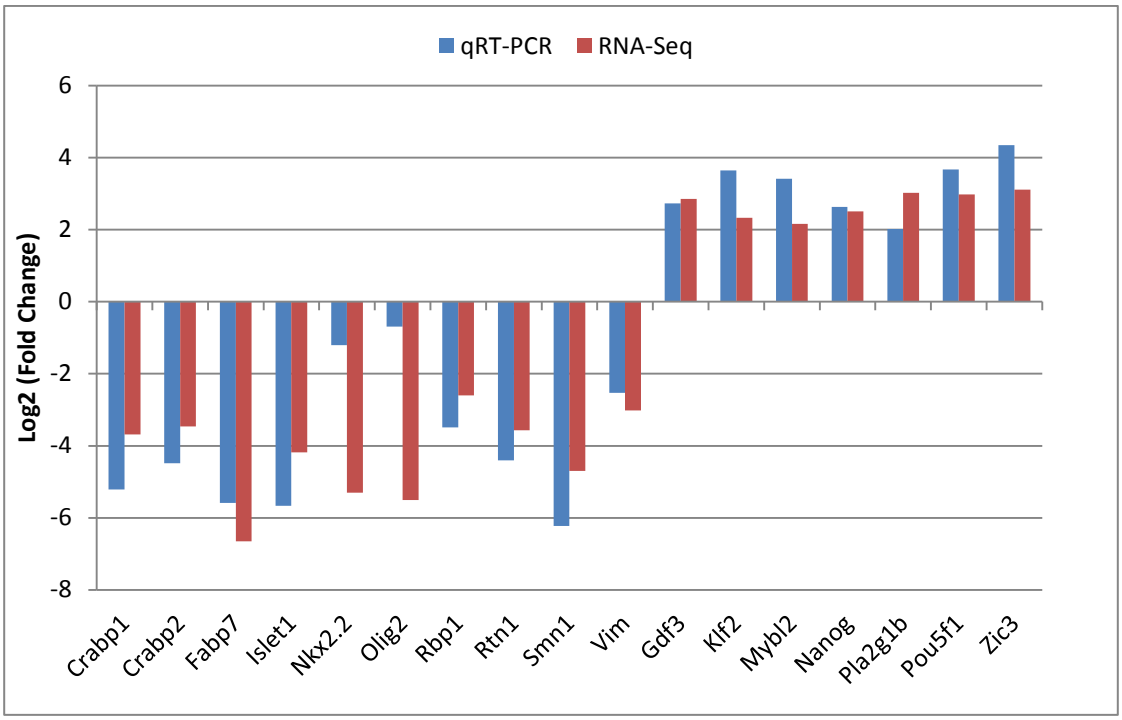


**Figure 3.6** qRT-PCR of 17 genes in total, 9 down-regulated and 7 up-regulated. Log2(fold change) of each gene is shown for qRT-PCR and RNA-Seq results. Replicates (n=3) of each sample was run and ΔCt values were averaged. Ct values were normalized against Cyclophilin A.

# Chapter 4

# DISCUSSION

Our study aimed to answer a key question in the field of SMA: Why do reduced levels of a ubiquitously expressed protein, SMN result in selective motor neuron death? We approached this question by comparing the genome wide expression pattern of ES cell derived MNs from a wild-type and a severe SMA model mouse. Through RNA-Seq we were able to identify differentially expressed genes and overall gene patterns that differed between samples.

MNs were derived from control and SMA samples, and characterized by immunostaining. We found that the ES cell derived MNs express GFP and that the GFP+ cells were an indication that the cells had differentiated into MNs. Characterizations were previously performed in our lab with an alternative mES cell model for SMA (Wu et al., 2011), which yielded similar results that mES cells were capable of producing MNs.

To study global gene expression, the Illumina HiSeq 2000 was used in a single end run with 50 bp read sequence lengths. We were able to obtain between 29 million and 44 million reads per sample. After trimming reads and FastQC analysis, we concluded that our samples were of high quality and could be used for transcriptome analysis.

When RNA-Seq analysis was conducted, a surprising pattern emerged. Significantly up-regulated genes were involved in pluripotency pathways, and down-regulated genes had roles in nervous system development. This trend was seen in the network analysis, pathway analysis as well as in the genes chosen from highest

expression and highest fold change. Among the up-regulated genes listed from both top affected genes and genes used for confirmation studies are eight genes which are either markers of pluripotency to help maintain the ES cell phenotype or have been found to be expressed in ES cells. Nanog, Pou5f1/Oct4, and Sox2 are three transcription factors considered to be hallmarks of ES cells (Boyer et al., 2005). All three factors (Sox2 data not shown) have been found to be up-regulated in SMA cells. Previous studies have shown that several genes work with these three transcription factors in their role in ES cells. Klf2 regulates the Sox2 gene (Redmond et al., 2011) whereas Zic3 is directly regulated by all three transcription factors (Lim et al., 2007). Lefty1, although it has no known function in embryogenesis, has binding sites for both Oct4 and Sox2 (Nakatake et al., 2006). Finally Zscan10/Zfp206 helps to maintain pluripotency by jointly functioning with Sox2 and Oct4 (W. Zhang et al., 2006). The other two genes, Mybl2 and Zfp42/Rex1, are also independently implicated in the maintenance of pluripotency (Papetti & Augenlicht, 2011; Masui et al., 2008). Taken together these genes suggest that the SMA cells maintain characteristics consistent with ES cell phenotypes.

In support of the idea that the SMA cells are developmentally impaired or delayed, 12 genes in the down-regulated genes list are genes expressed during neuronal development. Furthermore there are 2 other genes also down-regulated that play a role specifically in MN development. The 12 genes from the down-regulated gene list include Ascl1, Lrrn1, Fabp7, Hes5, Rbp1, Crabp1 and 2, Vimentin (Vim), Mir124a-2, Glr1, Nkx2.2 and Rtn1. To demonstrate the variety of neuronal markers, vimentin is an intermediate filament expressed by all neuronal precursors. It has been suggested to play a role in neurite outgrowth in hippocampal neurons (Boyne, Fisher

& Shea, 1996). Nkx2.2 is homeobox gene activated downstream of Shh, it has a primary role in ventral neuronal patterning, specifically in generating interneurons (Briscoe et al., 1999). Mir124a-2 is a microRNA highly expressed during mouse brain development (Smirnova et al., 2005). Finally Lrrn1 may regulate subcellular localization of signaling components in neuroepithelial cells (Andreae, Peukert, Lumsden & Gilthorpe, 2007). The variety of down-regulated genes suggests that not only are MNs not forming, but there is difficulty in neural differentiation itself. This may explain why the number of genes down-regulated in SMA is higher than up-regulated genes, as many of the genes are involved in not just MNs, but in different aspects of neuronal identity.

Islet1 and Olig2, as previously mentioned, are markers expressed during MN development (Wichterle, Lieberam, Porter, & Jessell, 2002). Surprisingly, RNA-Seq results have shown that Hb9 (MN marker) expression is not significantly affected in SMA cells (Figure S1). Hb9 is a late stage MN marker (Wichterle, Lieberam, Porter, & Jessell, 2002), and therefore it may be possible that in both control and SMA cells it is not as highly expressed as expected. However, we also need to take into consideration that the mES cell lines contain a Gfp-marker linked to the Hb9 promoter, and therefore, GFP+ cells would also express Hb9. Further works need to be done to address this issue with qRT-PCR validation as the first step in the process.

The overall gene trend is surprising taken into consideration that SMA is known as a neurodegenerative disorder, and not as a developmental one. Apoptosis and cell death markers were not among the the 4-fold or higher, $p<0.01$, and FPKM >30 lists. In concordance with our finding, a previous study has found that in the embryos of the same severe SMA mouse model, cell death was observed in the

telencephalon, but not in the ventral horn of the spinal cord (Liu, Shafey, Moores, & Kothary, 2010).

On the topic of SMA as a developmental disorder, literature sources vary. Zebrafish with a knockdown of Smn by antisense morpholinos has shown defects in motor axons suggesting early developmental defects (McWhorter, Monani, Burghes & Beattie, 2003). However, studies in the severe SMA mice ($SMN2^{+/+}$; $Smn^{-/-}$) showed no developmental defects in motor axon formation (McGovern, Gavrilina, Beattie & Burghes, 2008; Liu Shafey, Moores, & Kothary, 2010). Several studies have instead shown in both mice and Drosophila that developmental defects occur not in axonal formation, but only in the neuromuscular junction which lead to decreased denervation (McGovern, Gavrilina, Beattie & Burghes, 2008; Kariya et al., 2008; Chang et al., 2008). These studies are not consistent with our findings. When all the significantly affected genes were analyzed, the top canonical pathway found was the axonal guidance pathway. This would indicate that in our samples, developmental defects occur not just in neuromuscular junctions, but also within the axons themselves. This difference in consistency needs to be addressed in future studies.

This study was the first RNA-Seq experiment conducted on MNs from a severe SMA mouse model. No other study has examined transcriptome profiling in an ES model of SMA. A previous gene expression study of primary MN cultures from heterozygous SMA mice on microarrays did not yield any results in developmental defects (Anderson, Baban, Oliver, Potter & Davies, 2004). Another study used microarray analysis of whole spinal cords from the same severe mouse model of SMA at pre-symptomatic (P1) and post-symptomatic (P5) stages (Murray et al., 2009). Although this study by Murray et al. (2009) had concluded that SMA is a

neurodegenerative disease instead of a neurodevelopmental one, a couple findings from their microarray results show similar trends as our RNA-Seq results. The first was Dppa5a, a developmental pluripotency associated gene that was found to be significantly up-regulated in their pre-symptomatic mouse spinal cord. Although this was only one gene, and other gene expression patterns did not indicate any developmental defects, it was interesting to note that a pluripotency gene was upregulated.

The second finding by Murray et al. (2009) discovered a change in retinoic and retinoid metabolic processes in the pathway analysis of genes affected in late-symptomatic SMA spinal cords. RA is known for driving cellular differentiation, and has a neutralizing effect during development. In our study we found in the top 10 high expressing down-regulated gene list, 3 genes that were involved in retinoic or retinoid binding proteins. Rbp1 is a retinoid binding protein, and Crabp1 and Crabp2 are retinoic acid binding proteins found to be expressed in early mouse embryos and suggested to play a role in the development of the CNS (Dencker, Annerwall, Busch & Eriksson, 1990). Also, Zfp42/Rex1, which was found to be up-regulated in SMA cells, has been shown to be repressed by RA in F9 teratocarcinoma cells (Hosler, LaRosa, Grippo & Gudas, 1989). These changes in RA related genes may be the key to explaining why within our study SMA cells had failed to differentiate. RA regulates many phases during MN differentiation (Figure 4.1). RA has been implicated in its ability to induce neurogenesis by blocking fibroblast growth factor (FGF) signaling (Diez del Corral et al., 2003). Furthermore, RA and FGF signaling is enough to induce MN differentiation independent of Shh signaling (Novitch, Wichterle, Jessell & Sockanathan, 2003). Due to the importance of RA signaling during development, a

reduction in RA binding proteins may be responsible for the neurodevelopmental defect in the SMA cells. However, from our limited data we cannot speculate whether the reduction in RA binding proteins is the cause or result of developmental abnormalities in the SMA cells. We hope that future studies will help to elucidate this question.

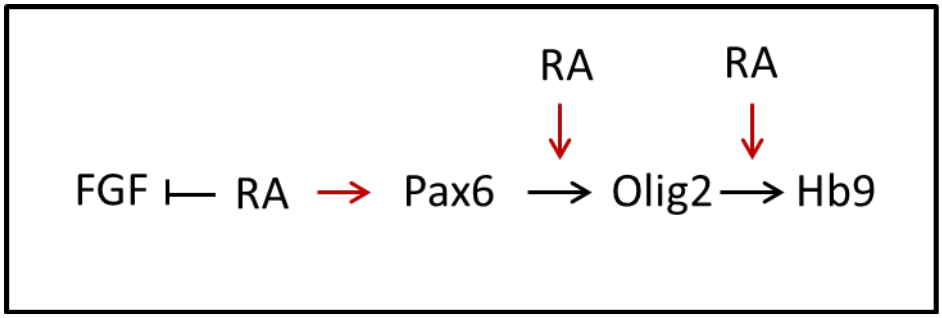


**Figure 4.1** RA regulates many phases of MN development. RA inhibits FGF signaling to induce neurogenesis. RA also is responsible for regulating each stage of neural specification. Modified from original (Appel & Eisen, 2003)

After comparing to literature, we cannot rule out the possibility that our current observations are model specific, and if so, to what degree. Preliminary studies of qRT-PCR using the same 17 genes used for confirmation were conducted (Figure S2) in an alternate mES model of SMA previously characterized in our lab (Wu et al., 2011). Although the experiment was with a n=1, overall trends of the 17 genes were in agreement with our SMA ES cell model indicating that this pattern may be specific in ES cell models from severe SMA mice. However, future experiments will need to be conducted to determine statistical significance. In addition, qRT-PCR comparison

against primary cultures obtained from severe mouse models of SMA may also shed light on model specific differences.

In conclusion we have observed distinct gene expression patterns in the SMA cell lines versus the control. Pathways up-regulated in SMA were involved in pluripotency and cell proliferation whereas common pathways found in the down-regulated genes have shown decreases in neuronal markers commonly found in mature and developing neurons. Taken together these findings suggest that SMA cells are not differentiating into MNs or are differentiating at a much slower pace than control the control line. Whether the observation of developmental issues is model specific or is indicative of the SMA disease as a whole has yet to be determined. The mechanisms behind this phenomenon need to also be addressed before further conclusions can be made.

# Chapter 5

## SUMMARY AND FUTURE DIRECTIONS

Our aim was to compare gene expression patterns between a SMA ES cell model against a control ES cell model. When mES cells were differentiated with RA and Shh, we found both control and SMA cell lines capable of producing mature motor neurons. Furthermore, when MNs were isolated and genome wide profiling was done, we observed distinct gene expression patterns. Within the genes up-regulated in SMA cells were markers for pluripotency and cell proliferation, whereas genes down-regulated in SMA cells have shown markers for axonal guidance and neuronal differentiation. However, many questions remain answered.

The first work which needs to be done is to run an actual validation of the RNA-Seq data using qRT-PCR on 3 different biological replicates from control and SMA cell lines. This would test not only the RNA-Seq technology, but the SMA vs control conditions as well. Furthermore, as originally planned, 22 genes were initially chosen and primers created, however, due to sub-optimal primer conditions, only 17 were able to be used to run qRT-PCR. Therefore, it would be ideal to design new primers, optimize them, and to run the validation on all 22 primer sets.

After validating the samples in the same cell lines used for RNA-Seq, we also would like to compare the same 22 genes in alternate models. This would help to answer the question of whether the gene expression pattern was unique to this specific cell model. As previously discussed, another mES cell model for severe SMA is also available. However, due to the lack of GFP expression in these cell lines, an alternate

method would need to be used to isolate MNs from the heterogeneous cell population. Isolation through Magnetic Activated Cell Sorting (MACS, Miltenyi Biotec, Auburn, CA, USA) would allow us to isolate MNs that would then be compared through qRT-PCR. Furthermore comparison of the 22 genes in primary cultures obtained from SMA mice MNs would allow for further confirmation. These experiments would help to answer the question of whether the observations seen are model specific.

Regardless of model specificity, the question remains why differences in development are seen in our model of SMA. We discussed the possibility that RA may hold a key to answering this question. For future studies, we would like to see if any gene expression changes in retinoic acid/retinoid metabolism or RA binding proteins are observable in earlier stages of differentiation. This would shed light on temporal patterns associated with these differences. Furthermore, we would like to address if increasing doses of RA during the differentiation process would change transcript patterns in RA related genes. Results will indicate if the problem lies in either the binding proteins themselves or if not enough signaling molecules are present to induce differentiation.

Separately, our future directions can also be taken with the RNA-Seq data analysis process. In our current study, only the annotated genes were analyzed and known gene expression patterns were examined. RNA-Seq has many applications including the ability to investigate novel genes, isoforms, and splice variants. Therefore, we would like to take advantage of this technology by also further analyzing our data set for other unique patterns in SMA. This would be done with through both TopHat and Cufflinks by turning on the option for Cuffcompare, a sub-

program within Cufflinks that analyzes transfrags to reference annotations therefore helping to identify any novel groups (Trapnell et al., 2010).

## REFERENCES


Anderson, K., Baban, D., Oliver, P., Potter, A., & Davies, K. (2004). Expression profiling in spinal muscular atrophy reveals an RNA binding protein deficit. *Neuromuscular Disorders, 14*(11), 711-722. doi: 10.1016/j.nmd.2004.08.009

Andreae, L. C., Peukert, D., Lumsden, A., & Gilthorpe, J. D. (2007). Analysis of LrrnI expression and its relationship to neuromeric boundaries during chick neural development. *Neural Development, 2,* 22. doi: 10.1186/1749-8104-2-22

Appel, B., & Eisen, J. (2003). Retinoids run rampant: Multiple roles during spinal cord and motor neuron development. *Neuron, 40*(3), 461-464. doi: 10.1016/S0896-6273(03)00688-3

Arber, S., Han, B., Mendelsohn, M., Smith, M., Jessell, T. M., & Sockanathan, S. (1999). Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. *Neuron, 23*(4) doi: 10.1016/S0896-6273(01)80026-X

Bassell, G., Zhang, H., Byrd, A., Femino, A., Singer, R., Taneja, K., . . . Kosik, K. (1998). Sorting of beta-actin mRNA and protein to neurites and growth cones in culture. *Journal of Neuroscience, 18*(1), 251-265.

Ben-Shachar, S., Orr-Urtreger, A., Bardugo, E., Shomrat, R., & Yaron, Y. (2011). Large-scale population screening for spinal muscular atrophy: Clinical implications. *Genetics in Medicine, 13*(2) doi: 10.1097/GIM.0b013e3182017c05

Bleicheret, F., & Baserga, S. J. (2010). Ribonucleoprotein multimers and their functions. *Critical Reviews in Biochemistry and Molecular Biology, 45*(5), 331-350. doi: 10.3109/10409238.2010.496772

Boyer, L., Lee, T., Cole, M., Johnstone, S., Levine, S., Zucker, J., . . . Young, R. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell, 122*(6), 947-956. doi: 10.1016/j.cell.2005.08.020

Boyne, L., Fischer, I., & Shea, T. (1996). Role of vimentin in early stages of neuritogenesis in cultured hippocampal neurons. *International Journal of*

*Developmental Neuroscience, 14*(6), 739-748. doi: 10.1016/S0736-5748(96)00053-6

Briscoe, J., Sussel, L., Serup, P., Hartigan-O'Connor, D., Jessell, T., Rubenstein, J., & Ericson, J. (1999). Homeobox gene Nkx2.2 and specification of neuronal identity by graded sonic hedgehog signalling. *Nature, 398*(6728), 622-627.

Buhler, D., Raker, V., Luhrmann, R., & Fischer, U. (1999). Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: Implications for spinal muscular atrophy. *Human Molecular Genetics, 8*(13), 2351-2357.

Burnett, B. G., Munoz, E., Tandon, A., Kwon, D. Y., Sumner, C. J., & Fischbeck, K. H. (2009). Regulation of SMN protein stability. *Molecular and Cellular Biology, 29*(5), 1107-1115. doi: 10.1128/MCB.01262-08

Burghes, A. H. M. (1997). When is a deletion not a deletion? when it is converted. *The American Journal of Human Genetics, 61*(1), 9-15. doi: 10.1086/513913

Cauchi, R. J. (2010). 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. doi: 10.1002/bies.201000088

Chang, H. C., Dimlich, D. N., Yokokura, T., Mukherjee, A., Kankel, M. W., Sen, A., . . . Artavanis-Tsakonas, S. (2008). Modeling spinal muscular atrophy in drosophila. *Plos One, 3*(9), e3209. doi: 10.1371/journal.pone.0003209

Cobben, J., Vandersteege, G., Grootscholten, P., Devisser, M., Scheffer, H., & Buys, C. (1995). Deletions of the survival motor-neuron gene in unaffected siblings of patients with spinal muscular-atrophy.*American Journal of Human Genetics, 57*(4), 805-808.

Cusco, I., Barcelo, M., Soler, C., Parra, J., Baiget, M., & Tizzano, E. (2002). Prenatal diagnosis for risk of spinal muscular atrophy. *Bjog-an International Journal of Obstetrics and Gynaecology, 109*(11), 1244-1249. doi: 10.1046/j.1471-0528.2002.02083.x

D'Amico, A., Mercuri, E., Tiziano, F. D., & Bertini, E. (2011). Spinal muscular atrophy. *Orphanet Journal of Rare Diseases, 6*, 71. doi: 10.1186/1750-1172-6-71

Dencker, L., Annerwall, E., Busch, C., & Eriksson, U. (1990). Localization of specific retinoid-binding sites and expression of cellular retinoic-acid-binding protein (crabp) in the early mouse embryo.*Development, 110*(2), 343-&.

DiDonato, C., Lorson, C., De Repentigny, Y., Simard, L., Chartrand, C., Androphy, E., & Kothary, R. (2001). Regulation of murine survival motor neuron (smn) protein levels by modifying smn exon 7 splicing. *Human Molecular Genetics, 10*(23), 2727-2736. doi: 10.1093/hmg/10.23.2727

Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M., & Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron, 40*(1), 65-79. doi: 10.1016/S0896-6273(03)00565-8

Eggert, C., Chari, A., Laggerbauer, B., & Fischer, U. (2006). Spinal muscular atrophy: The RNP connection. *Trends in Molecular Medicine, 12*(3), 113-121. doi: 10.1016/j.molmed.2006.01.005

Fang, Z., & Cui, X. (2011). Design and validation issues in RNA-seq experiments. *Briefings in Bioinformatics, 12*(3), 280-287. doi: 10.1093/bib/bbr004

Feng, L., & Heintz, N. (1995). Differentiating neurons activate transcription of the brain lipid-binding protein gene in radial glia through a novel regulatory element. *Development, 121*(6), 1719-1730.

Forrest, A. R. R., & Carninci, P. (2009). Whole genome transcriptome analysis. *Rna Biology, 6*(2), 107-112.

Frugier, T., Tiziano, F. D., Cifuentes-Diaz, C., Miniou, P., Roblot, N., Dierich, A., . . . Melki, J. (2000). Nuclear targeting defect of SMN lacking the C-terminus in a mouse model of spinal muscular atrophy. *Human Molecular Genetics, 9*(5) doi: 10.1093/hmg/9.5.849

Garber, M., Grabherr, M. G., Guttman, M., & Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods, 8*(6), 469-477. doi: 10.1038/NMETH.1613

Glinka, M., Herrmann, T., Funk, N., Havlicek, S., Rossoll, W., Winkler, C., & Sendtner, M. (2010). The heterogeneous nuclear ribonucleoprotein-R is necessary for axonal beta-actin mRNA translocation in spinal motor neurons. *Human Molecular Genetics, 19*(10), 1951-1966. doi: 10.1093/hmg/ddq073

Guo, W., Burris, T., & McCabe, E. (1995). Expression of dax-1, the gene responsible for X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism, in the hypothalamic-pituitary-adrenal gonadal axis. *Biochemical and Molecular Medicine, 56*(1), 8-13. doi: 10.1006/bmme.1995.1049

Hosler, B., LaRosa, G., Grippo, J., & Gudas, L. (1989). Expression of rex-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9-teratocarcinoma cells. *Molecular and Cellular Biology, 9*(12), 5623-5629.

Hsieh-Li, H., Chang, J., Jong, Y., Wu, M., Wang, N., Tsai, C., & Li, H. (2000). A mouse model for spinal muscular atrophy. *Nature Genetics, 24*(1), 66-70.

Ikeda, Y., Takeda, Y., Shikayama, T., Mukai, T., Hisano, S., & Morohashi, K. (2001). Comparative localization of dax-1 and Ad4BP/SF-1 during development of the hypothalamic-pituitary-gonadal axis suggests their closely related and distinct functions. *Developmental Dynamics, 220*(4), 363-376. doi: 10.1002/dvdy.1116

Kambach, C., Walke, S., Young, R., Avis, J., de la Fortelle, E., Raker, V., . . . Nagai, K. (1999). Crystal structures of two sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell, 96*(3), 375-387. doi: 10.1016/S0092-8674(00)80550-4

Kariya, S., Park, G., Maeno-Hikichi, Y., Leykekhman, O., Lutz, C., Arkovitz, M. S., . . . Monani, U. R. (2008). Reduced SMN protein impairs maturation of the neuromuscular junctions in mouse models of spinal muscular atrophy. *Human Molecular Genetics, 17*(16), 2552-2569. doi: 10.1093/hmg/ddn156

Le, T., Pham, L., Butchbach, M., Zhang, H., Monani, U., Coovert, D., . . . Burghes, A. (2005). SMN delta 7, the major product of the centromeric survival motor neuron (SMN2) gene, extends survival in mice with spinal muscular atrophy and associates with full-length SMN. *Human Molecular Genetics, 14*(6), 845-857. doi: 10.1093/hmg/ddi078

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., . . . Melki, J. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell, 80*(1), 155-165. doi: 10.1016/0092-8674(95)90460-3

Lim, L. S., Loh, Y., Zhang, W., Li, Y., Chen, X., Wang, Y., . . . Stanton, L. W. (2007). Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Molecular Biology of the Cell, 18*(4), 1348-1358. doi: 10.1091/mbc.E06-07-0624

Liu, H., Shafey, D., Moores, J. N., & Kothary, R. (2010). Neurodevelopmental consequences of smn depletion in a mouse model of spinal muscular atrophy. *Journal of Neuroscience Research, 88*(1), 111-122. doi: 10.1002/jnr.22189

Liu, Q., & Dreyfuss, G. (1996). A novel nuclear structure containing the survival of motor neurons protein. *Embo Journal, 15*(14), 3555-3565.

Lorson, C., Hahnen, E., Androphy, E., & Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proceedings of the National Academy of Sciences of the United States of America, 96*(11), 6307-6311. doi: 10.1073/pnas.96.11.6307

Lyahyai, J., Sbiti, A., Barkat, A., Ratbi, I., & Sefiani, A. (2012). Spinal muscular atrophy carrier frequency and estimated prevalence of the disease in moroccan newborns. *Genetic Testing and Molecular Biomarkers, 16*(3), 215-218. doi: 10.1089/gtmb.2011.0149

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal, 17*(1), pp. 10-12.

Masui, S., Ohtsuka, S., Yagi, R., Takahashi, K., Ko, M. S. H., & Niwa, H. (2008). Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *Bmc Developmental Biology, 8*, 45. doi: 10.1186/1471-213X-8-45

Matera, A. G., Terns, R. M., & Terns, M. P. (2007). Non-coding RNAs: Lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews Molecular Cell Biology, 8*(3), 209-220. doi: 10.1038/nrm2124

Mattaj, I. W., & De Robertis, E. M. (1985). Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell, 40*(1), 111-118. doi: 10.1016/0092-8674(85)90314-9

McGovern, V. L., Gavrilina, T. O., Beattie, C. E., & Burghes, A. H. M. (2008). Embryonic motor axon development in the severe SMA mouse. *Human Molecular Genetics, 17*(18), 2900-2909. doi: 10.1093/hmg/ddn189

McWhorter, M., Monani, U., Burghes, A., & Beattie, C. (2003). Knockdown of the survival motor neuron (smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. *Journal of Cell Biology,162*(5), 919-931. doi: 10.1083/job.200303168

Miles, G., Yohn, D., Wichterle, H., Jessell, T., Rafuse, V., & Brownstone, R. (2004). Functional properties of motoneurons derived from mouse embryonic stem cells. *Journal of Neuroscience, 24*(36), 7848-7858. doi: 10.1523/JNEUROSCI.1972-04.2004

Monani, U., Lorson, C., Parsons, D., Prior, T., Androphy, E., Burghes, A., & McPherson, J. (1999). A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2.*Human Molecular Genetics, 8*(7), 1177-1183. doi: 10.1093/hmg/8.7.1177

Monani, U., Sendtner, M., Coovert, D., Parsons, D., Andreassi, C., Le, T., . . . Burghes, A. (2000). The human centromeric survival motor neuron gene (SMN2) rescues embryonic lethality in smn(-/-) mice and results in a mouse with spinal muscular atrophy. *Human Molecular Genetics, 9*(3), 333-339. doi: 10.1093/hmg/9.3.333

Morozova, O., Hirst, M., & Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics, 10*, 135-151. doi: 10.1146/annurev-genom-082908-145957

Morrison, K. E. (1996). Advances in SMA research: Review of gene deletions. *Neuromuscular Disorders, 6*(6), 397-408. doi: 10.1016/S0960-8966(96)00368-9

Mourelatos, Z., Abel, L., Yong, J., Kataoka, N., & Dreyfuss, G. (2001). SMN interacts with a novel family of hnRNP and spliceosomal proteins. *Embo Journal, 20*(19), 5443-5452. doi: 10.1093/emboj/20.19.5443

Murray, L. M., Lee, S., Baeumer, D., Parson, S. H., Talbot, K., & Gillingwater, T. H. (2010). Pre-symptomatic development of lower motor neuron connectivity in a mouse model of severe spinal muscular atrophy. *Human Molecular Genetics, 19*(3), 420-433. doi: 10.1093/hmg/ddp506

Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., . . . Niwa, H. (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells.*Molecular and Cellular Biology, 26*(20), 7772-7782. doi: 10.1128/MCB.00468-06

Novitch, B., Wichterle, H., Jessell, T., & Sockanathan, S. (2003). A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification.*Neuron, 40*(1), 81-95. doi: 10.1016/j.neuron.2003.08.006

Ogg, S., & Lamond, A. (2002). Cajal bodies and coilin - moving towards function. *Journal of Cell Biology, 159*(1), 17-21. doi: 10.1083/jcb.200206111

Ohtsuka, T., Ishibashi, M., Gradwohl, G., Nakanishi, S., Guillemot, F., & Kageyama, R. (1999). Hes1 and Hes5 as notch effectors in mammalian neuronal differentiation. *Embo Journal, 18*(8), 2196-2207. doi: 10.1093/emboj/18.8.2196

O'Shea, K. (1999). Embryonic stem cell models of development. *Anatomical Record, 257*(1), 32-41. doi: 10.1002/(SICI)1097-0185(19990215)257:1<32::AID-AR6>3.0.CO;2-2

Pagliardini, S., Giavazzi, A., Setola, V., Lizier, C., Di Luca, M., DeBiasi, S., & Battaglia, G. (2000). Subcellular localization and axonal transport of the survival motor neuron (SMN) protein in the developing rat spinal cord. *Human Molecular Genetics, 9*(1) doi: 10.1093/hmg/9.1.47

Papetti, M., & Augenlicht, L. H. (2011). MYBL2, a link between proliferation and differentiation in maturing colon epithelial cells. *Journal of Cellular Physiology, 226*(3), 785-791. doi: 10.1002/jcp.22399

Park, D., Xiang, A. P., Mao, F. F., Zhang, L., Di, C., Liu, X., . . . Lahn, B. T. (2010). Nestin is required for the proper self-renewal of neural stem cells. *Stem Cells, 28*(12), 2162-2171. doi: 10.1002/stem.541

Pfaff, S., Mendelsohn, M., Stewart, C., Edlund, T., & Jessell, T. (1996). Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell, 84*(2), 309-320. doi: 10.1016/S0092-8674(00)80985-X

Prior, T. W. (2010). Perspectives and diagnostic considerations in spinal muscular atrophy. *Genetics in Medicine, 12*(3), 145-152. doi: 10.1097/GIM.0b013e3181c5e713

Purandare, S., Ware, S., Kwan, K., Gebbia, M., Bassi, M., Deng, J., . . . Casey, B. (2002). A complex syndrome of left-right axis, central nervous system and axial skeleton defects in Zic3 mutant mice.*Development, 129*(9), 2293-2302.

Redmond, L. C., Dumur, C. I., Archer, K. J., Grayson, D. R., Haar, J. L., & Lloyd, J. A. (2011). Kruppel-like factor 2 regulated gene expression in mouse embryonic yolk sac erythroid cells. *Blood Cells Molecules and Diseases, 47*(1), 1-11. doi: 10.1016/j.bcmd.2011.03.002

Rodrigues, N., Owen, N., Talbot, K., Ignatius, J., Dubowitz, V., & Davies, K. (1995). Deletions in the survival motor-neuron gene on 5q13 in autosomal recessive spinal muscular-atrophy. *Human Molecular Genetics, 4*(4), 631-634. doi: 10.1093/hmg/4.4.631

Rossoll, W., Jablonka, S., Andreassi, C., Kroning, A., Karle, K., Monani, U., & Sendtner, M. (2003). Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons. *Journal of Cell Biology, 163*(4), 801-812. doi: 10.1083/jcb.200304128

Russman, B. S. (2007). Spinal muscular atrophy: Clinical classification and disease heterogeneity. *Journal of Child Neurology, 22*(8) doi: 10.1177/0883073807305673

Schrank, B., Gotz, R., Gunnersen, J., Ure, J., Toyka, K., Smith, A., & Sendtner, M. (1997). Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. *Proceedings of the National Academy of Sciences of the United States of America, 94*(18), 9920-9925. doi: 10.1073/pnas.94.18.9920

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., & Sattler, M. (2001). SMN tudor domain structure and its interaction with the sm proteins. *Nature Structural Biology, 8*(1), 27-31.

Smirnova, L., Grafe, A., Seiler, A., Schumacher, S., Nitsch, R., & Wulczyn, F. (2005). Regulation of miRNA expression during neural cell specification. *European Journal of Neuroscience, 21*(6), 1469-1477. doi: 10.1111/j.1460-9568.2005.03978.x

Su, Y., Hung, C., Lin, S., Chen, F., Chern, J. P. S., Tsai, C., . . . Lee, C. (2011). Carrier screening for spinal muscular atrophy (SMA) in 107,611 pregnant women during the period 2005-2009: A prospective population-based cohort study. *Plos One, 6*(2), e17067. doi: 10.1371/journal.pone.0017067

Sugarman, E. A., Nagan, N., Zhu, H., Akmaev, V. R., Zhou, Z., Rohlfs, E. M., . . . Allitto, B. A. (2012). Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy: Clinical laboratory analysis of > 72 400 specimens. *European Journal of Human Genetics, 20*(1), 27-32. doi: 10.1038/ejhg.2011.134

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics, 25*(9), 1105-1111. doi: 10.1093/bioinformatics/btp120

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., . . . Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq

experiments with TopHat and cufflinks. *Nature Protocols, 7*(3), 562-578. doi: 10.1038/nprot.2012.016

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., . . . Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology, 28*(5), 511-U174. doi: 10.1038/nbt.1621

Tsukui, T., Capdevila, J., Tamura, K., Ruiz-Lozano, P., Rodriguez-Esteban, C., Yonei-Tamura, S., . . . Belmonte, J. (1999). Multiple left-right asymmetry defects in shh(-/-) mutant mice unveil a convergence of the shh and retinoic acid pathways in the control of lefty-1. *Proceedings of the National Academy of Sciences of the United States of America, 96*(20), 11376-11381. doi: 10.1073/pnas.96.20.11376

Wichterle, H., Lieberam, I., Porter, J., & Jessell, T. (2002). Directed differentiation of embryonic stem cells into motor neurons. *Cell, 110*(3), 385-397. doi: 10.1016/S0092-8674(02)00835-8

Wu, C., Whye, D., Glazewski, L., Choe, L., Kerr, D., Lee, K. H., . . . Wang, W. (2011). Proteomic assessment of a cell model of spinal muscular atrophy. *Bmc Neuroscience, 12*, 25. doi: 10.1186/1471-2202-12-25

Zhang, H. L. L., Pan, F., Hong, D., Shenoy, S. M., Singer, R. H., & Bassell, G. J. (2003). Active transport of the survival motor neuron protein and the role of exon-7 in cytoplasmic localization. *Journal of Neuroscience,23*(16)

Zhang, H., Xing, L., Rossoll, W., Wichterle, H., Singer, R. H., & Bassell, G. J. (2006). Multiprotein complexes of the survival of motor neuron protein SMN with gemins traffic to neuronal processes and growth cones of motor neurons (vol 26, pg 8622, 2006). *Journal of Neuroscience, 26*(35)

Zhang, W., Walker, E., Tamplin, O. J., Rossant, J., Stanford, W. L., & Hughes, T. R. (2006). Zfp206 regulates ES cell gene expression and differentiation. *Nucleic Acids Research, 34*(17), 4780-4790. doi: 10.1093/nar/gkl631

## Appendix

## SUPPLEMENTARY FIGURES

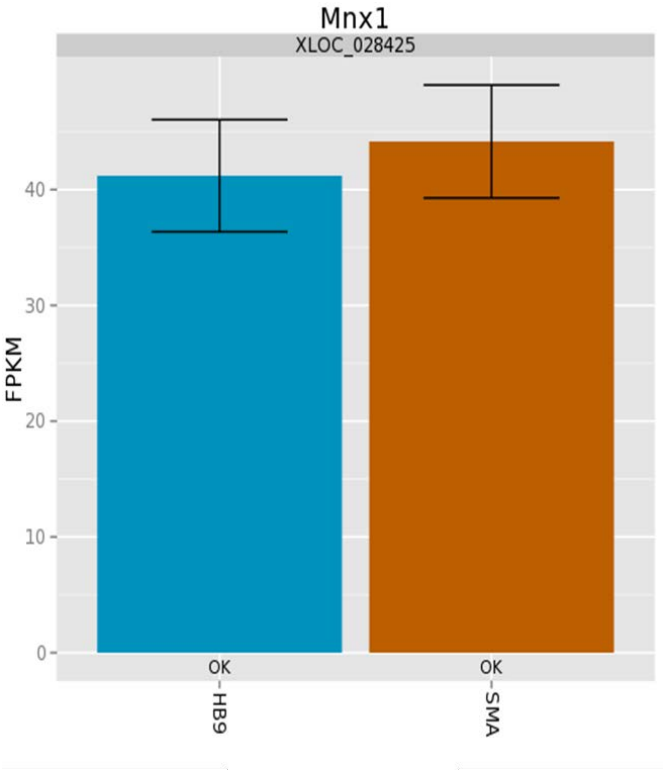


**Figure A.1** Expression of Hb9 in control (Hb9) and SMA cells from RNA-Seq results. Graphs were generated using the CummeRbund package in R. Graphs show no statistical significant changes in Hb9 expression between control and SMA cells.

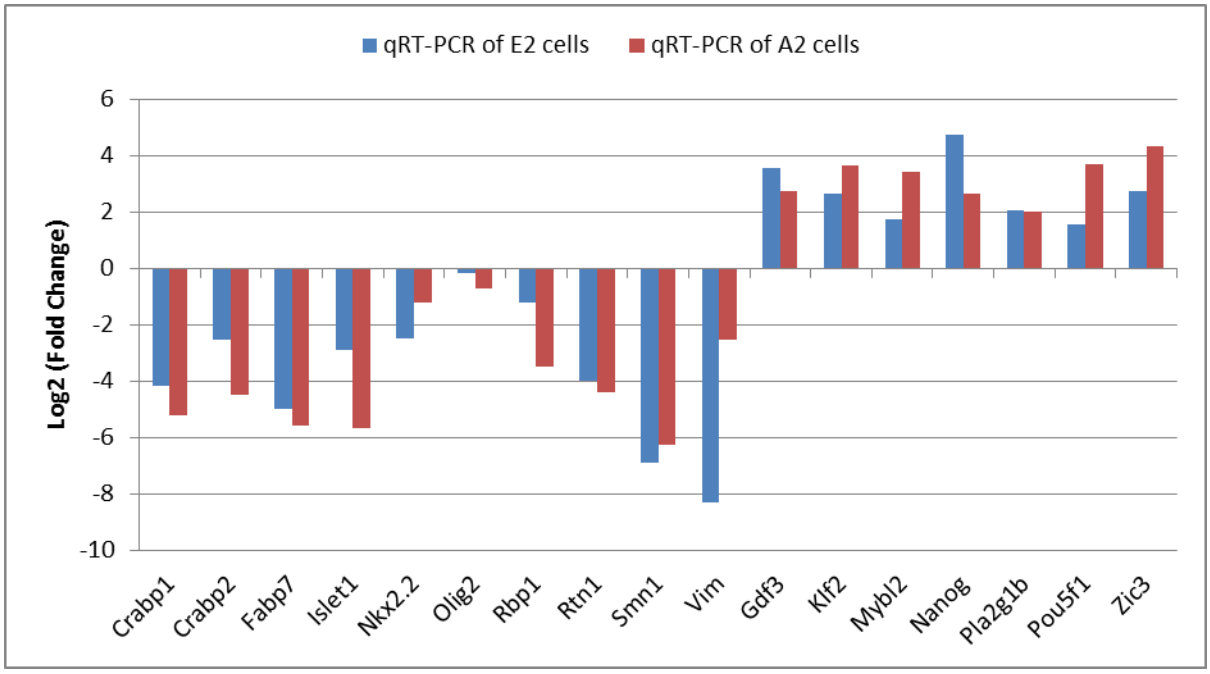


**Figure A.2** Preliminary qRT-PCR comparison between two different ES cell lines both from severe mouse models of SMA. Cells were grown in similar manner. A2 cell lines are the cells used within the thesis. E2 cell lines do not contain the GFP transgene. The E2 analysis was done on n=1.