2016

Novel Functions of the Survival Motor Neuron Protein

Eric William Ottesen
Iowa State University

Novel functions of the Survival Motor Neuron protein

by

Eric W Ottesen

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Molecular, Cellular, and Developmental Biology

Program of Study Committee:
Ravindra N. Singh, Major Professor
Jeffrey K Beetham
Jeffrey Trimarchi
Peng Liu
Brian M. Lee

Iowa State University

Ames, Iowa

2016

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

There are many people that I would like to thank for their support and guidance throughout my life and graduate career. First and foremost, I would like to thank my advisor, Dr. Ravindra Singh, for guiding me through my graduate research and constantly challenging me to see beyond the day-to-day matters of research to the larger whole. I would like to thank all the members of the Singh lab, both past and present, who have contributed to the research contained in this text, both materially and intellectually. I would also like to thank everybody who has served on my committee, Drs. Jeffrey Beetham, Jeffrey Trimarchi, Peng Liu, Brian Lee, and Drena Dobbs, for their guidance and advice. Finally, I would like to thank my parents, Ruth and James Ottesen, for raising an inquisitive little daydreamer to become the inquisitive not-so-little daydreamer that I am today.

ABSTRACT

The Survival Motor Neuron (SMN) protein is a multi-functional protein that participates in a wide variety of critical pathways. Low levels of SMN cause spinal muscular atrophy (SMA), the most common genetic cause of infant mortality. While the role of SMN in the assembly of small nuclear ribonucleoproteins (snRNPs) has been well characterized, many of its other diverse functions have not been thoroughly explored. Here, we examine the critical role of SMN in the growth and development of male mammalian sex organs. We show that low levels of SMN in a mild mouse model of SMA cause impaired testis development, degenerated seminiferous tubules, reduced sperm count, and low fertility. Underscoring an increased requirement for SMN expression, wild type testis showed extremely high levels of SMN protein compared to other tissues. The testis phenotype is linked to increased apoptosis in seminiferous tubules and extreme perturbations in the testis transcriptome. We examine the RNA binding function of SMN by Systematic Evolution of Ligands by Exponential Enrichment (SELEX) to identify RNA sequence and structural motif(s) of SMN. Our results reveal a combination of sequence motifs and structural contexts that drive the specificity of RNA-SMN interactions. Our results of truncation and substitution experiments suggest a requirement for multiple contacts between SMN and RNA to maintain the high affinity. We demonstrate that both affinity and specificity of RNA-SMN interaction are influenced by salt concentrations. To identify in vivo RNA targets of SMN, we performed crosslinking and immunoprecipitation coupled with high-throughput sequencing (HITS-CLIP). HITS-CLIP identified a variety of RNA targets of SMN with an enrichment of mRNAs

participating in a number of pathways, including ribosome function and actin cytoskeleton regulation. In order to determine whether expression levels of target RNAs are regulated by SMN, we performed knockdown of SMN levels followed by RNA-Seq. *SPON2*, *LAMB2*, and *EEF1A2* in particular were all predicted by HITS-CLIP to be bound by SMN and were downregulated upon SMN knockdown, indicating a direct regulatory role for SMN on expression of these genes.

CHAPTER I

GENERAL INTRODUCTION

Spinal Muscular Atrophy and the SMN Protein

Spinal Muscular Atrophy (SMA) is a genetic disease, primarily affecting infants and small children. It is the most common genetic cause of infant mortality, affecting 1 in 10,000 live births, with a carrier frequency of 1 in 35 (Wirth et al 2006). SMA is caused by low levels of the Survival Motor Neuron (SMN) due to mutation, deletion, or gene conversion of the *SMN1* gene, resulting in progressive deterioration of motor neurons (Lefebvre et al 1995). In mice, deletion of the *Smn* gene, which is equivalent to *SMN1*, is homozygous lethal (Schrank et al 1997). However, due to a duplication of the region of chromosome 5 containing contains *SMN1*, humans have a second copy of the *SMN* gene, *SMN2*. *SMN2* is only a partially functional copy, due to the presence of a C-T mutation in the 6th position of exon 7. Although the C-T mutation is translationally silent, it alters the splicing pattern of *SMN2* pre-mRNA, causing predominant skipping of *SMN2* exon 7 (Lorson et al 1999). The mRNA isoform produced by skipping of exon 7 produces a truncated protein with an altered C-terminus, SMNΔ7, which is highly unstable and deficient in several functions (Cho and Dreyfuss 2010, Gupta et al 2015). However, SMNΔ7 retains the ability to extend survival in a severe mouse model of SMA when overexpressed (Le et al 2005), indicating that functionality of the protein is not completely compromised in the absence of exon 7.

Domain Structure of the SMN Protein

SMN is a multifunctional protein and has many interacting partners (reviewed later in the text), and as such, has multiple functional domains (Figure 1A-1B). The region coded by exons 2A and 2B mediate the interaction with the SMN binding partner Gemin2 (Zhang et al 2011), which is crucial for a number of critical functions of SMN (Fischer et al 1997, Liu et al 1997, Sanchez et al 2013, Takaku et al 2011, Piazzon et al 2013). In addition, an overlapping region is also required for the nucleic acid binding activity of the SMN protein (Bertrandy et al 1999, Lorson and Androphy 1998). Exon 3 codes for the Tudor domain. Tudor domains, mediate interactions with methylated arginine residues (Chen et al 2011). Specifically, the SMN Tudor domain binds methylated arginine residues on the Sm proteins (Selenko et al 2001), as well as a number of other proteins critical for SMN function (Hebert et al 2001, Sabra et al 2013, Zhao et al 2016). Consistently, mutations within and antibodies directed against the SMN Tudor domain both dramatically impair snRNP assembly (Buhler et al 1999). The domain coded for by the region spanning exon 4 through the first portion of exon 6 is predicted to be disordered, with a number of poly-proline stretches. This portion of the SMN protein is predicted to interact with the actin cytoskeleton-associated protein Profilin (Giesemann et al 1999, Nolle et al 2011). The C-terminal region, in particular the region spanning the end of exon 6 and the beginning of exon 7, contains the YG box, which mediates SMN protein oligomerization (Gupta et al 2015). Isoforms of SMN with disrupted YG box structure also have reduced stability (Seo et al 2016).

Although deletion of part or all of *SMN1* is the most common cause of SMA, a small proportion of patients suffer from point mutations which cause coding changes within the SMN protein (Wirth et al 2000, Cusco et al 2004, Sun et al 2005, Zapletalova et al 2007, Howell et al 2014, Figure 1C). These mutations are a critical tool for understanding the different domains of SMN because they indicate essential sequences that are required for either SMN function or stability. The majority of these point mutations localize to either the Tudor domain or the YG box, confirming the critical nature of these two domains (Figure 1C).

The SMN Complex: A Multipart, Multifunctional RNP Assembly Hub

SMN has multiple protein-interacting domains, and, as such, has a number of binding partners that operate within numerous processes (Table 1). The most well-characterized of these by far is the SMN complex. The SMN complex, as described, consists of 9 different proteins: SMN, Gemins 2-8, and UNRIP (Cauchi 2010). Although it is generally referred to as a single entity, in reality, the SMN complex consists of multiple sub-complexes which can function independently (Battle et al 2007). Proteins within the SMN complex contain signals for both nuclear localization and export (Zhang et al 2003, Lorson et al 2008). Consistently, SMN and other complex members are generally located in both the cytoplasm and the nucleus (Zhang et al 2003). However, the cytoplasm is the primary site of snRNP assembly (Massenet et al 2002). At its most fundamental core, the SMN functional unit consists of SMN and Gemin2 (Zhang et al 2011). Consistent with the concept of SMN-Gemin2 as a functional unit, it was recently reported that U1-70K can "hijack"

SMN-Gemin2 to facilitate U1 snRNP assembly without the assistance of other Gemins (So et al 2016). Underscoring its critical role, Gemin2 is even more widely evolutionarily conserved than SMN itself. In *Arabidopsis*, for example, Gemin2 is required for proper circadian rhythm regulation due to its effect on splicing; however, the protein that is predicted to fill the role of SMN is actually a homolog of SPF30 rather than SMN (Schlaen et al 2015). Although SMN and Gemin2 are almost universally co-regulated and found interacting with one another, there is slowly mounting evidence that this may not always be the case. For example, in a fluorescence microscopy-based study of axonal mRNP transport, a portion of the particles observed contained SMN but lacked Gemin2, indicating a possible functional SMN unit that does not contain this critical interacting partner (Zhang et al 2006). Supporting a lack of requirement for Gemin2 in an axonal function for SMN, reduction of Gemin2 had no effect on the outgrowth of motor neuron axons, unlike SMN (McWhorter et al 2008).

Gemin3 and Gemin4 form another subunit of the SMN complex (Battle et al 2007). Gemin3, otherwise known as DDX20 and/or DP103, is a putative DEAD-box helicase based on sequence homology (Charroux et al 1999, Campbell et al 2000). The precise function of Gemin3-Gemin4 within the SMN complex is currently unknown, although they independently interact with Gemin5 (Battle et al 2007), and a requirement for ATP hydrolysis for snRNP assembly (Meister et al 2001, Pellizzoni et al 2002b) coupled with Gemin3's identity as a putative helicase suggests an RNA remodeling function for this subcomplex. Independently of the SMN complex, Gemin3 has also been implicated in miRNA biogenesis (Mourelatos et

al 2002). Gemin4 immunoreactivity is localized to both gems and the nucleolus, indicating a possible secondary function for this protein in ribosome biogenesis (Charroux et al 2000).

Gemin5 is a large (~169 kDa) RNA binding protein (RBP) containing 13 WD repeat domains which mediate its RNA binding activity (Lau et al 2009). Gemin5 serves as the primary snRNA recruitment factor of the SMN complex (Lau et al 2009, Yong et al 2010). Supporting its function as a recruitment factor rather than as a core member of the SMN complex, the majority of Gemin5 protein exists outside of the SMN complex (Battle et al 2007). In order to deliver snRNAs to the SMN complex, Gemin5 interacts directly with both Gemin2 and the Gemin3-Gemin4 subcomplex (Otter et al 2007). Gemin5 also functions outside of the SMN complex in the regulation of translation. Gemin5 is an interacting partner of eIF4E and is essential for cap-independent translation of a number of viral mRNAs (Fierro-Monti et al 2006, Pacheco et al 2009). Gemin5 is also recruited to stress granules upon induction of their formation by arsenite treatment (Battle et al 2007).

The functions of Gemins6-8 and UNRIP are less well characterized than the other subunits of the SMN complex. Like Gemin5, the majority of these proteins are found outside of the SMN complex (Battle et al 2007). Also like Gemin5, UNRIP has been implicated in cap-independent translation (Hunt et al 1999) during viral infection. However, additional roles for these proteins in the absence of viral infection have not been investigated.

SMN Plays a Role in DNA Repair and Transcription

In the event of DNA damage, lysine 79 of Histone H3 becomes methylated, which in turn recruits SMN through its Tudor domain (Sabra et al 2013). Further supporting a role in DNA repair, SMN-Gemin2 complexes were found to stimulate RAD51-mediated homologous recombination and double-stranded break repair (Takaku et al 2011). Another nuclear target of arginine methylation is the C-terminal domain of RNA polymerase II (PolII-CTD) (Zhao et al 2016). Recently it was found that SMN is recruited to the transcription elongation complex by methylation of R1810 within PolII CTD and interacts with senataxin in order to resolve the RNA-DNA duplex formed during transcription and properly execute termination (Zhao et al 2016). Interestingly, this is the second case of SMN assisting in resolving nucleic acid duplexes.

SMN and Associated Proteins Form a Nuclear Structure Known as Gems

In the nucleus, SMN is primarily located in Cajal bodies and in nuclear structures known as Gems (Liu and Dreyfuss, 1996). It is not currently known exactly what the function(s) of these nuclear bodies may be, but a loss of SMN in gems is correlated with incidence of amyotrophic lateral sclerosis (ALS) (Shan et al 2010, Yamazaki et al 2012). The signature protein marking Cajal bodies is coilin, which contains arginines which serve as methylation targets, thereby recruiting SMN through its Tudor domain (Hebert et al 2001, Boisvert et al 2002). A number of other SMN-interacting proteins also become targeted to Cajal bodies/gems, such as ZPR1, all of the Gemins, and snoRNP markers Fibrillarin and GAR1 (Ahmad et al 2012, Pellizzoni

et al 2001). In addition to Gems/Cajal bodies, SMN and associated proteins have been shown to localize to the nucleolus in certain tissues and conditions (Charroux et al 2000, Young et al 2001, Wehner et al 2002).

SMN Functions in the Biogenesis of the Signal Recognition Particle

The signal recognition particle (SRP) is an RNP complex that interacts with elongating polypeptides with N-terminal signal peptides and targets them for translocation into the endoplasmic reticulum (Keenan et al 2001, Saraogi and Shan 2011). The SRP consists of the 7S RNA that has two domains, the Alu domain, and the S-domain, which are connected by a long stem (Keenan et al 2001) and several protein components, named SRP9, 14, 19, 54, 68, and 72 (Saraogi and Shan 2011). The 7S RNA was initially identified in a screen of potential in vitro interacting partners with purified SMN complex (Piazzon et al 2012). It was further shown that Gemin5 directly binds the S-domain and that the SMN complex is required for the assembly of the SRP54 protein onto the 7S RNA (Piazzon et al 2012). In addition, the level of 7S RNA was significantly reduced in the spinal cord of SMA model mice, indicating a requirement for high levels of SMN for proper expression of functional SRP (Piazzon et al 2012).

Neuron-Specific Functions of the SMN Protein

SMA is primarily a disease of motor neurons (Burghes and Beattie 2009); therefore, when considering functions of the SMN protein, any processes that are specific to neuronal cells are of particular interest. Within axons of cultured

neuronal cells, SMN is localized to actively transported granules (Zhang et al 2006). These SMN-positive granules colocalize with a number of the Gemin proteins, but not the Sm proteins or Fibrillarin (Zhang et al 2006, Todd et al 2010). SMN interacts with the vesicle coat protein COPI (Peter et al 2011, Custer et al 2013), which may be the mechanism by which it is targeted to axons. In elongating axons, the growth cones are the site of local translation (Steward and Schuman 2003, Colak et al 2013, Fallini et al 2016). Interestingly, SMN is known to interact with a number of RNA-binding proteins (RBPs) (covered later) and knockdown of SMN in primary motor neurons results in redistribution of poly(A)+ mRNA away from the axons to the cell body (Fallini et al 2011). Therefore, a new model for SMN function has emerged, in which SMN interacts with mRNPs, either through interactions with RBPs or a potential direct interaction with mRNA itself, and assists in targeting these complexes to axons for transport and eventually local translation. Supporting this hypothesis, the first mRNA found to colocalize with SMN within axons is β-actin (Rossoll et al 2003), whose local translation has been shown to be critical for axon outgrowth and steering (Willis et al 2005, Leung et al 2006). More recently, *CPG15* and *GAP43* mRNAs, both critical for axon outgrowth, have been identified as colocalizing with SMN and are dependent on SMN expression for transport (Akten et al 2011, Fallini et al 2016).

In addition to affecting axon outgrowth through local mRNA translation, SMN can also play a more direct role in axon growth by affecting actin filament dynamics. SMN interacts directly with Profilin (Giesemann et al 1999, Bowerman et al 2007, Nolle et al 2011), which plays a critical role in actin polymerization (Pollard and

Borisy 2003), Annexin A2, which mediates interactions between cell membranes and the actin cytoskeleton (Shafey et al 2010), and myosin regulatory light chain (MRLC) (Shafey et al 2010). Supporting a direct role for actin dynamics and axonogenesis in the pathogenesis of SMA, actin bundling protein Plastin 3 has been identified as a protective disease modifier of SMA (Oprea et al 2008).

SMN Interacts with Numerous RNA Binding Proteins

Of all of the SMN protein's diverse interactions, one overriding theme, though not necessarily universal, is the presence of both nucleic acids and protein. In fact, the SMN complex has been referred to as a general-purpose RNP assembly machine (Battle et al 2006). To that end, it is not surprising that SMN interacts with a large number of RBPs. Gemin5 and U1-70K have already been adressed (Yong et al 2010, So et al 2016). SMN also interacts directly with the RBP FUS, which is often mutated in amyotrophic lateral sclerosis (ALS) (Yamazaki et al 2012). In addition, expression of mutant forms of FUS or another ALS-causing protein TDP-43 results in redistribution of SMN from nuclear gems into cytoplasmic granules, implying that ALS and SMA may share similar mechanisms of pathogenesis (Yamazaki et al 2012). Staying within the realm of RBPs containing small RNAs, SMN interacts directly with both Fibrillarin and GAR1, the primary protein components of snoRNPs (Pellizzoni et al 2001). The presence of this interaction, in addition to a proposed role for SMN in loading SRP54 onto the 7S RNA of the signal recognition particle (SRP) (Piazzon et al 2013), cements a general role for SMN in the assembly of functional RNPs.

SMN also interacts with a number of mRNA binding proteins which are involved in pre-mRNA splicing, as well as mRNA stability and/or transport. In terms of splicing factors, SMN interacts with hnRNP Q and KH-type splicing regulatory protein (KSRP) (Mourelatos et al 2001, Rossoll et al 2002, Tadesse et al 2008). Numerous RBPs have been implicated in axonal RNP transport along with SMN. These include zipcode binding protein (ZBP) (Fallini et al 2014), hnRNP R (Rossoll et al 2002, Glinka et al 2010), and HuD (Fallini et al 2011).

Mouse Models of Spinal Muscular Atrophy

In order to study the progression of any disease or to test potential treatments, it is critical to have a model that recapitulates the human disease as closely as possible. However, despite its simple genetic origins as an autosomal recessive disease, SMA is quite challenging to model successfully. This is due to the presence of *SMN2*, which is specific to humans. Without the presence of *SMN2* to compensate, deletion of *Smn* in mice is homozygous lethal in early embryonic stages (Schrank et al 1997). Therefore, most mouse models of SMA are generated by reducing the level of SMN protein expression without eliminating it entirely, either by generating transgenic lines containing human *SMN2* or by modulating the expression of protein generated from the mouse *Smn* locus, or some combination of the two.

The first SMA mouse models were generated by producing mice transgenic for the *SMN2* locus and crossing them with lines with null mutations of the *Smn* gene (Hsieh-Li et al 2000, Monani et al 2000). Depending on the copy number of the

transgene in question, mice were either phenotypically normal (in high copy number lines) or experienced severe SMA-like symptoms, including a smaller size, decreased motor ability, and a near complete loss of the righting reflex, and generally died within 4-6 days of birth (Monani et al 2000).

Due to the severity of their symptoms, the initial mouse models of SMA proved difficult to work with. In order to extend the survival of SMA mice, a new model was generated in which mice carrying multiple copies of a cDNA construct of SMN with exon 7 deleted (*SMNΔ7*) were crossed with existing transgenic mice harboring deletion of *Smn* and a single copy of *SMN2* (Le et al 2005). The resulting mice (*Smn*$^{-/-}$*;SMN2*;*SMNΔ7*, commonly referred to as Δ7 mice) had a slightly ameliorated phenotype, with a mean survival of ~10 days (Le et al 2005). To this date, the Δ7 model remains the most widely used mouse model of SMA.

A number of conditional mouse models have also been generated. These usually fall into two categories: conditional rescue and conditional knockout. One of the earliest studies used conditional rescue using transgenic SMN under different promoters to show that neuronally expressed SMN under the control of the prion protein promoter was able to rescue knockout of *Smn*, but not SMN expressed in muscles under the control of the human skeletal actin promoter (Gavrilina et al 2008). Later models used an inducible rescue by placing the neomycin resistance cassette within the mouse *Smn* gene flanked by LoxP sites, with Cre recombinase driven by different promoters (Hammond et al 2010). On the flip side, conditional knockouts have also been generated. Mice deficient in SMN expression in motor neurons alone (Cifuentes-Diaz et al 2002) or in skeletal muscle alone (Cifuentes-

Diaz et al 2001) both experience severe phenotypes, indicating that at least some expression of SMN is required in both tissues.

More recently, an alternative approach was taken to generate a range of mouse models with simplified genetics. Rather than rely on separate knockout of *Smn* and expression of SMN protein from transgenic inserts, the *Smn* locus itself was modified to modulate SMN expression using 4 alternative constructs which could be crossed to each other to produce a range of phenotypes (Osborne et al 2012). This approach, termed the allelic series, was the first to generate a truly intermediate SMA mouse in the C allele, which was generated by replacing the last 2 exons of *Smn* with those of human *SMN2*, followed by a genomic fragment containing the entire *SMN2* gene (exons+introns) (Osborne et al 2012). Mice homozygous for the C allele (C/C) have a normal lifespan, but suffer from peripheral necrosis of the ears and tail, and have reduced body weight (Osborne et al 2012). Later in life, C/C mice have altered electrophysiological properties at the neuromuscular junction, but no apparent defects in coordination or muscular strength (Osborne et al 2012).

References

Ahmad, S., Wang, Y., Shaik, G.M., Burghes, A.H., and Gangwani, L. (2012). The zinc finger protein ZPR1 is a potential modifier of spinal muscular atrophy. Human Molecular Genetics *21*, 2745-2758.

Akten, B., Kye, M.J., Hao, L.T., Wertz, M.H., Singh, S., Nie, D.Y., Huang, J., Merianda, T.T., Twiss, J.L., Beattie, C.E.*, et al.* (2011). Interaction of survival of motor neuron (SMN) and HuD proteins with mRNA cpg15 rescues motor neuron axonal deficits. Proceedings of the National Academy of Sciences of the United States of America *108*, 10337-10342.

Baccon, J., Pellizzoni, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). Identification and characterization of Gemin7, a novel component of the survival of motor neuron complex. Journal of Biological Chemistry *277*, 31957-31962.

Battle, D.J., Kasim, M., Wang, J., and Dreyfuss, G. (2007). SMN-independent subunits of the SMN complex - Identification of a small nuclear ribonucleoprotein assembly intermediate. Journal of Biological Chemistry *282*, 27953-27959.

Battle, D.J., Kasim, M., Yong, J., Lotti, F., Lau, C.K., Mouaikel, J., Zhang, Z., Han, K., Wan, L., and Dreyfuss, G. (2006). The SMN complex: An assembly machine for RNPs. Cold Spring Harbor Symposia on Quantitative Biology *71*, 313-320.

Bertrandy, S., Burlet, P., Clermont, O., Huber, C., Fondrat, C., Thierry-Mieg, D., Munnich, A., and Lefebvre, S. (1999). The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. Human Molecular Genetics *8*, 775-782.

Boisvert, F.M., Cote, J., Boulanger, M.C., Cleroux, P., Bachand, F., Autexier, C., and Richard, S. (2002). Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing. Journal of Cell Biology *159*, 957-969.

Bowerman, M., Shafey, D., and Kothary, R. (2007). Smn depletion alters profilin II expression and leads to upregulation of the RhoA/ROCK pathway and defects in neuronal integrity. Journal of Molecular Neuroscience *32*, 120-131.

Buhler, D., Raker, V., Luhrmann, R., and 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*, 2351-2357.

Burghes, A.H.M., and Beattie, C.E. (2009). Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? Nature Reviews Neuroscience *10*, 597-609.

Campbell, L., Hunter, K.M.D., Mohaghegh, P., Tinsley, J.M., Brasch, M.A., and Davies, K.E. (2000). Direct interaction of Smn with dp103, a putative RNA helicase: a role for Smn in transcription regulation? Human Molecular Genetics *9*, 1093-1100.

Carissimi, C., Baccon, J., Straccia, M., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J., and Pellizzoni, L. (2005). Unrip is a component of SMN complexes active in snRNP assembly. Febs Letters *579*, 2348-2354.

Carissimi, C., Saieva, L., Baccon, J., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J., and Pellizzoni, L. (2006). Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly. Journal of Biological Chemistry *281*, 8126-8134.

Carnegie, G.K., Sleeman, J.E., Morrice, N., Hastie, C.J., Peggie, M.W., Philp, A., Lamond, A.I., and Cohen, P.T.W. (2003). Protein phosphatase 4 interacts with the Survival of Motor Neurons complex and enhances the temporal localisation of snRNPs. Journal of Cell Science *116*, 1905-1913.

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*, 1077-1089.

Charroux, B., Pellizzoni, L., Perkinson, R.A., Shevchenko, A., Mann, M., and Dreyfuss, G. (1999). Gemin3: A novel DEAD box protein that interacts with SMN, the spinal muscular atrophy gene product, and is a component of gems. Journal of Cell Biology *147*, 1181-1193.

Charroux, B., Pellizzoni, L., Perkinson, R.A., Yong, J., Shevchenko, A., Mann, M., and Dreyfuss, G. (2000). Gemin4: A novel component of the SMN complex that is found in both gems and nucleoli. Journal of Cell Biology *148*, 1177-1186.

Chen, C., Nott, T.J., Jin, J., and Pawson, T. (2011). Deciphering arginine methylation: Tudor tells the tale. Nature Reviews Molecular Cell Biology *12*, 629-642.

Chen, H.H., Yu, H.I., Chiang, W.C., Lin, Y.D., Shia, B.C., and Tarn, W.Y. (2012). hnRNP Q Regulates Cdc42-Mediated Neuronal Morphogenesis. Molecular and Cellular Biology *32*, 2224-2238.

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

Cifuentes-Diaz, C., Frugier, T., Tiziano, F.D., Lacene, E., Roblot, N., Joshi, V., Moreau, M.H., and Melki, J. (2001). Deletion of murine SMN exon 7 directed to skeletal muscle leads to severe muscular dystrophy. Journal of Cell Biology *152*,

1107-1114.

Cifuentes-Diaz, C., Nicole, S., Velasco, M.E., Borra-Cebrian, C., Panozzo, C., Frugier, T., Millet, G., Roblot, N., Joshi, V., and Melki, J. (2002). Neurofilament accumulation at the motor endplate and lack of axonal sprouting in a spinal muscular atrophy mouse model. Human Molecular Genetics *11*, 1439-1447.

Claus, P., Doring, F., Gringel, S., Muller-Ostermeyer, F., Fuhlrott, J., Kraft, T., and Grothe, C. (2003). Differential intranuclear localization of fibroblast growth factor-2 isoforms and specific interaction with the survival of motoneuron protein. Journal of Biological Chemistry *278*, 479-485.

Colak, D., Ji, S.J., Porse, B.T., and Jaffrey, S.R. (2013). Regulation of Axon Guidance by Compartmentalized Nonsense-Mediated mRNA Decay. Cell *153*, 1252-1265.

Cusco, I., Barcelo, M.J., del Rio, E., Baiget, M., and Tizzano, E.F. (2004). Detection of novel mutations in the SMN Tudor domain in type ISMA patients. Neurology *63*, 146-149.

Custer, S.K., Todd, A.G., Singh, N.N., and Androphy, E.J. (2013). Dilysine motifs in exon 2b of SMN protein mediate binding to the COPI vesicle protein -COP and neurite outgrowth in a cell culture model of spinal muscular atrophy. Human Molecular Genetics *22*, 4043-4052.

Fallini, C., Donlin-Asp, P.G., Rouanet, J.P., Bassell, G.J., and Rossoll, W. (2016). Deficiency of the Survival of Motor Neuron Protein Impairs mRNA Localization and Local Translation in the Growth Cone of Motor Neurons. Journal of Neuroscience *36*, 3811-3820.

Fallini, C., Rouanet, J.P., Donlin-Asp, P.G., Guo, P., Zhang, H.L., Singer, R.H., Rossoll, W., and Bassell, G.J. (2014). Dynamics of Survival of Motor Neuron (SMN) Protein Interaction with the mRNA-Binding Protein IMP1 Facilitates Its Trafficking into Motor Neuron Axons. Developmental Neurobiology *74*, 319-332.

Fallini, C., Zhang, H.L., Su, Y.H., Silani, V., Singer, R.H., Rossoll, W., and Bassell, G.J. (2011). The Survival of Motor Neuron (SMN) Protein Interacts with the mRNA-Binding Protein HuD and Regulates Localization of Poly(A) mRNA in Primary Motor Neuron Axons. Journal of Neuroscience *31*, 3914-3925.

Fierro-Monti, I., Mohammed, S., Matthiesen, R., Santoro, R., Burns, J.S., Williams, D.J., Proud, C.G., Kassem, M., Jensen, O.N., and Roepstorff, P. (2006). Quantitative proteomics identifies Gemin5, a scaffolding protein involved in ribonucleoprotein assembly, as a novel partner for eukaryotic initiation factor 4E. Journal of Proteome Research *5*, 1367-1378.

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

Gangwani, L., Mikrut, M., Theroux, S., Sharma, M., and Davis, R.J. (2001). Spinal muscular atrophy disrupts the interaction of ZPR1 with the SMN protein. Nature Cell Biology *3*, 376-383.

Gavrilina, T.O., McGovern, V.L., Workman, E., Crawford, T.O., Gogliotti, R.G., DiDonato, C.J., Monani, U.R., Morris, G.E., and Burghes, A.H.M. (2008). Neuronal SMN expression corrects spinal muscular atrophy in severe SMA mice while muscle-specific SMN expression has no phenotypic effect. Human Molecular Genetics *17*, 1063-1075.

Giesemann, T., Rathke-Hartlieb, S., Rothkegel, M., Bartsch, J.W., Buchmeier, S., Jockusch, B.M., and Jockusch, H. (1999). A role for polyproline motifs in the spinal muscular atrophy protein SMN - Profilins bind to and colocalize with SMN in nuclear gems. Journal of Biological Chemistry *274*, 37908-37914.

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

Gubitz, A.K., Mourelatos, Z., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). Gemin5, a novel WD repeat protein component of the SMN complex that binds Sm proteins. Journal of Biological Chemistry *277*, 5631-5636.

Gupta, K., Martin, R., Sharp, R., Sarachan, K.L., Ninan, N.S., and Van Duyne, G.D. (2015). Oligomeric Properties of Survival Motor Neuron.Gemin2 Complexes. Journal of Biological Chemistry *290*, 20185-20199.

Hammond, S.M., Gogliotti, R.G., Rao, V., Beauvais, A., Kothary, R., and DiDonato, C.J. (2010). Mouse Survival Motor Neuron Alleles That Mimic SMN2 Splicing and Are Inducible Rescue Embryonic Lethality Early in Development but Not Late. Plos One *5*, 12.

Hebert, M.D., Szymczyk, P.W., Shpargel, K.B., and Matera, A.G. (2001). Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein. Genes & Development *15*, 2720-2729.

Howell, M.D., Singh, N.N., and Singh, R.N. (2014). Advances in therapeutic development for spinal muscular atrophy. Future Medicinal Chemistry *6*, 1081-1099.

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

Hunt, S.L., Hsuan, J.J., Totty, N., and Jackson, R.J. (1999). unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA. Genes & Development

13, 437-448.

Joshi, P., Greco, T.M., Guise, A.J., Luo, Y., Yu, F., Nesvizhskii, A.I., and Cristea, I.M. (2013). The functional interactome landscape of the human histone deacetylase family. Molecular Systems Biology *9*, 21.

Keenan, R.J., Freymann, D.M., Stroud, R.M., and Walter, P. (2001). The signal recognition particle. Annual Review of Biochemistry *70*, 755-775.

Lau, C.K., Bachorik, J.L., and Dreyfuss, G. (2009). Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains. Nature Structural & Molecular Biology *16*, 486-491.

Le, T.T., Pham, L.T., Butchbach, M.E.R., Zhang, H.L., Monani, U.R., Coovert, D.D., Gavrilina, T.O., Xing, L., Bassell, G.J., and Burghes, A.H.M. (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*, 845-857.

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M.*, et al.* (1995). IDENTIFICATION AND CHARACTERIZATION OF A SPINAL MUSCULAR ATROPHY-DETERMINING GENE. Cell *80*, 155-165.

Leung, K.M., van Horck, F.P.G., Lin, A.C., Allison, R., Standart, N., and Holt, C.E. (2006). Asymmetrical beta-actin mRNA translation in growth cones mediates attractive turning to netrin-1. Nature Neuroscience *9*, 1247-1256.

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

Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997). The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. Cell *90*, 1013-1021.

Lorson, C.L., and Androphy, E.J. (1998). The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. Human Molecular Genetics *7*, 1269-1275.

Lorson, C.L., Hahnen, E., Androphy, E.J., and 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*, 6307-6311.

Lorson, M.A., Dickson, A.M., Shaw, D.J., Todd, A.G., Young, E.C., Morse, R., Wolstencroft, C., Lorson, C.L., and Young, P.J. (2008). Identification and characterisation of a nuclear localisation signal in the SMN associated protein,

Gemin4. Biochemical and Biophysical Research Communications *375*, 33-37.

Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I.W., and Dreyfuss, G. (2002). The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. Molecular and Cellular Biology *22*, 6533-6541.

McWhorter, M.L., Boon, K.L., Horan, E.S., Burghes, A.H.M., and Beattie, C.E. (2008). The SMN binding protein Gemin2 is not involved in motor axon outgrowth. Developmental Neurobiology *68*, 182-194.

Meister, G., Buhler, D., Pillai, R., Lottspeich, F., and Fischer, U. (2001). A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nature Cell Biology *3*, 945-949.

Monani, U.R., Sendtner, M., Coovert, D.D., Parsons, D.W., Andreassi, C., Le, T.T., Jablonka, S., Schrank, B., Rossol, W., Prior, T.W*., et al.* (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*, 333-339.

Mourelatos, Z., Abel, L., Yong, J.S., Kataoka, N., and Dreyfuss, G. (2001). SMN interacts with a novel family of hnRNP and spliceosomal proteins. Embo Journal *20*, 5443-5452.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes & Development *16*, 720-728.

Nolle, A., Zeug, A., van Bergeijk, J., Tonges, L., Gerhard, R., Brinkmann, H., Al Rayes, S., Hensel, N., Schill, Y., Apkhazava, D*., et al.* (2011). The spinal muscular atrophy disease protein SMN is linked to the rho-kinase pathway via profilin. Human Molecular Genetics *20*, 4865-4878.

Oprea, G.E., Kroeber, S., McWhorter, M.L., Rossoll, W., Mueller, S., Krawczak, M., Bassell, G.J., Beattie, C.E., and Wirth, B. (2008). Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy. Science *320*, 524-527.

Osborne, M., Gomez, D., Feng, Z.H., McEwen, C., Beltran, J., Cirillo, K., El-Khodor, B., Lin, M.Y., Li, Y., Knowlton, W.M*., et al.* (2012). Characterization of behavioral and neuromuscular junction phenotypes in a novel allelic series of SMA mouse models. Human Molecular Genetics *21*, 4431-4447.

Otter, S., Grimmler, M., Neuenkirchen, N., Chari, A., Sickmann, A., and Fischer, U. (2007). A comprehensive interaction map of the human survival of motor neuron (SMN) complex. Journal of Biological Chemistry *282*, 5825-5833.

Pacheco, A., de Quinto, S.L., Ramajo, J., Fernandez, N., and Martinez-Salas, E. (2009). A novel role for Gemin5 in mRNA translation. Nucleic Acids Research *37*, 582-590.

Pellizzoni, L., Baccon, J., Charroux, B., and Dreyfuss, G. (2001). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. Current Biology *11*, 1079-1088.

Pellizzoni, L., Baccon, J., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002a). Purification of native survival of motor neurons complexes and identification of Gemin6 as a novel component. Journal of Biological Chemistry *277*, 7540-7545.

Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002b). Essential role for the SMN complex in the specificity of snRNP assembly. Science *298*, 1775-1779.

Peter, C.J., Evans, M., Thayanithy, V., Taniguchi-Ishigaki, N., Bach, I., Kolpak, A., Bassell, G.J., Rossoll, W., Lorson, C.L., Bao, Z.Z.*, et al.* (2011). The COPI vesicle complex binds and moves with survival motor neuron within axons. Human Molecular Genetics *20*, 1701-1711.

Piazzon, N., Schlotter, F., Lefebvre, S., Dodre, M., Mereau, A., Soret, J., Besse, A., Barkats, M., Bordonne, R., Branlant, C.*, et al.* (2013). Implication of the SMN complex in the biogenesis and steady state level of the Signal Recognition Particle. Nucleic Acids Research *41*, 1255-1272.

Pollard, T.D., and Borisy, G.G. (2003). Cellular motility driven by assembly and disassembly of actin filaments. Cell *112*, 453-465.

Rossoll, W., Jablonka, S., Andreassi, C., Kroning, A.K., Karle, K., Monani, U.R., and 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*, 801-812.

Rossoll, W., Kroning, A.K., Ohndorf, U.M., Steegborn, C., Jablonka, S., and Sendtner, M. (2002). Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons? Human Molecular Genetics *11*, 93-105.

Sabra, M., Texier, P., El Maalouf, J., and Lomonte, P. (2013). The Tudor protein survival motor neuron (SMN) is a chromatin-binding protein that interacts with methylated lysine 79 of histone H3. Journal of Cell Science *126*, 3664-3677.

Sanchez, G., Dury, A.Y., Murray, L.M., Biondi, O., Tadesse, H., El Fatimy, R., Kothary, R., Charbonnier, F., Khandjian, E.W., and Cote, J. (2013). A novel function for the survival motoneuron protein as a translational regulator. Human Molecular Genetics *22*, 668-684.

Saraogi, I., and Shan, S.O. (2011). Molecular Mechanism of Co-translational Protein Targeting by the Signal Recognition Particle. Traffic *12*, 535-542.

Saunders, L.R., Perkins, D.J., Balachandran, S., Michaels, R., Ford, R., Mayeda, A., and Barber, G.N. (2001). Characterization of two evolutionarily conserved, alternatively spliced nuclear phosphoproteins, NFAR-1 and -2, that function in mRNA processing and interact with the double-stranded RNA-dependent protein kinase, PKR. Journal of Biological Chemistry *276*, 32300-32312.

Schlaen, R.G., Mancini, E., Sanchez, S.E., Perez-Santangelo, S., Rugnone, M.L., Simpson, C.G., Brown, J.W.S., Zhang, X., Chernomoretz, A., and Yanovsky, M.J. (2015). The spliceosome assembly factor GEMIN2 attenuates the effects of temperature on alternative splicing and circadian rhythms. Proceedings of the National Academy of Sciences of the United States of America *112*, 9382-9387.

Schrank, B., Gotz, R., Gunnersen, J.M., Ure, J.M., Toyka, K.V., Smith, A.G., and 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*, 9920-9925.

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

Shafey, D., Boyer, J.G., Bhanot, K., and Kothary, R. (2010). Identification of Novel Interacting Protein Partners of SMN Using Tandem Affinity Purification. Journal of Proteome Research *9*, 1659-1669.

Shan, X., Chiang, P.M., Price, D.L., and Wong, P.C. (2010). Altered distributions of Gemini of coiled bodies and mitochondria in motor neurons of TDP-43 transgenic mice. Proceedings of the National Academy of Sciences of the United States of America *107*, 16325-16330.

So, B.R., Wan, L.L., Zhang, Z.X., Li, P.L., Babiash, E., Duan, J.Q., Younis, I., and Dreyfuss, G. (2016). A U1 snRNP-specific assembly pathway reveals the SMN complex as a versatile hub for RNP exchange. Nature Structural & Molecular Biology *23*, 225-230.

Steward, O., and Schuman, E.M. (2003). Compartmentalized synthesis and degradation of proteins in neurons. Neuron *40*, 347-359.

Sun, Y., Grimmler, M., Schwarzer, V., Schoenen, F., Fischer, U., and Wirth, B. (2005). Molecular and functional analysis of intragenic SMN1 mutations in patients with spinal muscular atrophy. Human Mutation *25*, 64-71.

Tadesse, H., Deschenes-Furry, J., Boisvenue, S., and Cote, J. (2008). KH-type

splicing regulatory protein interacts with survival motor neuron protein and is misregulated in spinal muscular atrophy. Human Molecular Genetics *17*, 506-524.

Takaku, M., Tsujita, T., Horikoshi, N., Takizawa, Y., Qing, Y., Hirota, K., Ikura, M., Ikura, T., Takeda, S., and Kurumizaka, H. (2011). Purification of the Human SMN-GEMIN2 Complex and Assessment of Its Stimulation of RAD51-Mediated DNA Recombination Reactions. Biochemistry *50*, 6797-6805.

Todd, A.G., Morse, R., Shaw, D.J., Stebbings, H., and Young, P.J. (2010). Analysis of SMN-neurite granules: Core Cajal body components are absent from SMN-cytoplasmic complexes. Biochemical and Biophysical Research Communications *397*, 479-485.

Wang, I.F., Reddy, N.M., and Shen, C.K.J. (2002). Higher order arrangement of the eukaryotic nuclear bodies. Proceedings of the National Academy of Sciences of the United States of America *99*, 13583-13588.

Wehner, K.A., Ayala, L., Kim, Y., Young, P.J., Hosler, B.A., Lorson, C.L., Baserga, S.J., and Francis, J.W. (2002). Survival motor neuron protein in the nucleolus of mammalian neurons. Brain Research *945*, 160-173.

Willis, D., Li, K.W., Zheng, J.Q., Chang, J.H., Smit, A., Kelly, T., Merianda, T.T., Sylvester, J., van Minnen, J., and Twiss, J.L. (2005). Differential transport and local translation of cytoskeletal, injury-response, and neurodegeneration protein mRNAs in axons. Journal of Neuroscience *25*, 778-791.

Wirth, B. (2000). An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). Human Mutation *15*, 228-237.

Wirth, B., Brichta, L., and Hahnen, E. (2006). Spinal muscular atrophy: from gene to therapy. Seminars in pediatric neurology *13*, 121-131.

Yamazaki, T., Chen, S., Yu, Y., Yan, B.A., Haertlein, T.C., Carrasco, M.A., Tapia, J.C., Zhai, B., Das, R., Lalancette-Hebert, M.*, et al.* (2012). FUS-SMN Protein Interactions Link the Motor Neuron Diseases ALS and SMA. Cell Reports *2*, 799-806.

Yong, J., Kasim, M., Bachorik, J.L., Wan, L.L., and Dreyfuss, G. (2010). Gemin5 Delivers snRNA Precursors to the SMN Complex for snRNP Biogenesis. Molecular Cell *38*, 551-562.

Young, P.J., Le, T.T., Dunckley, M., Man, N.T., Burghes, A.H.M., and Morris, G.E. (2001). Nuclear gems and cajal (coiled) bodies in fetal tissues: Nucleolar distribution of the spinal muscular atrophy protein, SMN. Experimental Cell Research *265*, 252-261.

Zapletalova, E., Hedvicakova, P., Kozak, L., Vondracek, P., Gaillyova, R., Marikova, T., Kalina, Z., Juttnerova, V., Fajkus, J., and Fajkusova, L. (2007). Analysis of point mutations in the SMN1 gene in SMA patients bearing a single SMN1 copy. Neuromuscular Disorders *17*, 476-481.

Zhang, H.L., Xing, L., Rossoll, W., Wichterle, H., Singer, R.H., and 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. Journal of Neuroscience *26*, 8622-8632.

Zhang, H.L.L., Pan, F., Hong, D., Shenoy, S.M., Singer, R.H., and 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*, 6627-6637.

Zhang, R.D., So, B.R., Li, P.L., Yong, J., Glisovic, T., Wan, L.L., and Dreyfuss, G. (2011). Structure of a Key Intermediate of the SMN Complex Reveals Gemin2's Crucial Function in snRNP Assembly. Cell *146*, 384-395.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. Cell *133*, 585-600.

Zhang, Z.X., Pinto, A.M., Wan, L.L., Wang, W., Berg, M.G., Oliva, I., Singh, L.N., Dengler, C., Wei, Z., and Dreyfuss, G. (2013). Dysregulation of synaptogenesis genes antecedes motor neuron pathology in spinal muscular atrophy. Proceedings of the National Academy of Sciences of the United States of America *110*, 19348-19353.

Zhao, D.Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z.Y., Schmitges, F.W., Zhong, G.Q., Liu, K., Li, W.G., Moffat, J.*, et al.* (2016). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature *529*, 48-+.

Zou, J.H., Barahmand-pour, F., Blackburn, M.L., Matsui, Y., Chansky, H.A., and Yang, L. (2004). Survival motor neuron (SMN) protein interacts with transcription corepressor mSin3A. Journal of Biological Chemistry *279*, 14922-14928.

Zou, T., Yang, X.M., Pan, D.M., Huang, J., Sahin, M., and Zhou, J.H. (2011). SMN Deficiency Reduces Cellular Ability to Form Stress Granules, Sensitizing Cells to Stress. Cellular and Molecular Neurobiology *31*, 541-550.

Figure Legends

Figure 1. Overview of the SMN Protein. (A) Domains of the SMN protein. Colored boxes represent different exons, with exon designation given within each box. Locations of individual domains are indicated. **(B)** Interaction sites of the SMN protein. Labeling of exons is the same as in (A). Primary interacting partner of each designated region is given. **(C)** Locations of known SMA-causing point mutations. Labeling of exons is the same as in (A). Position of each known mutation is indicated, labels give the specific amino acid change.

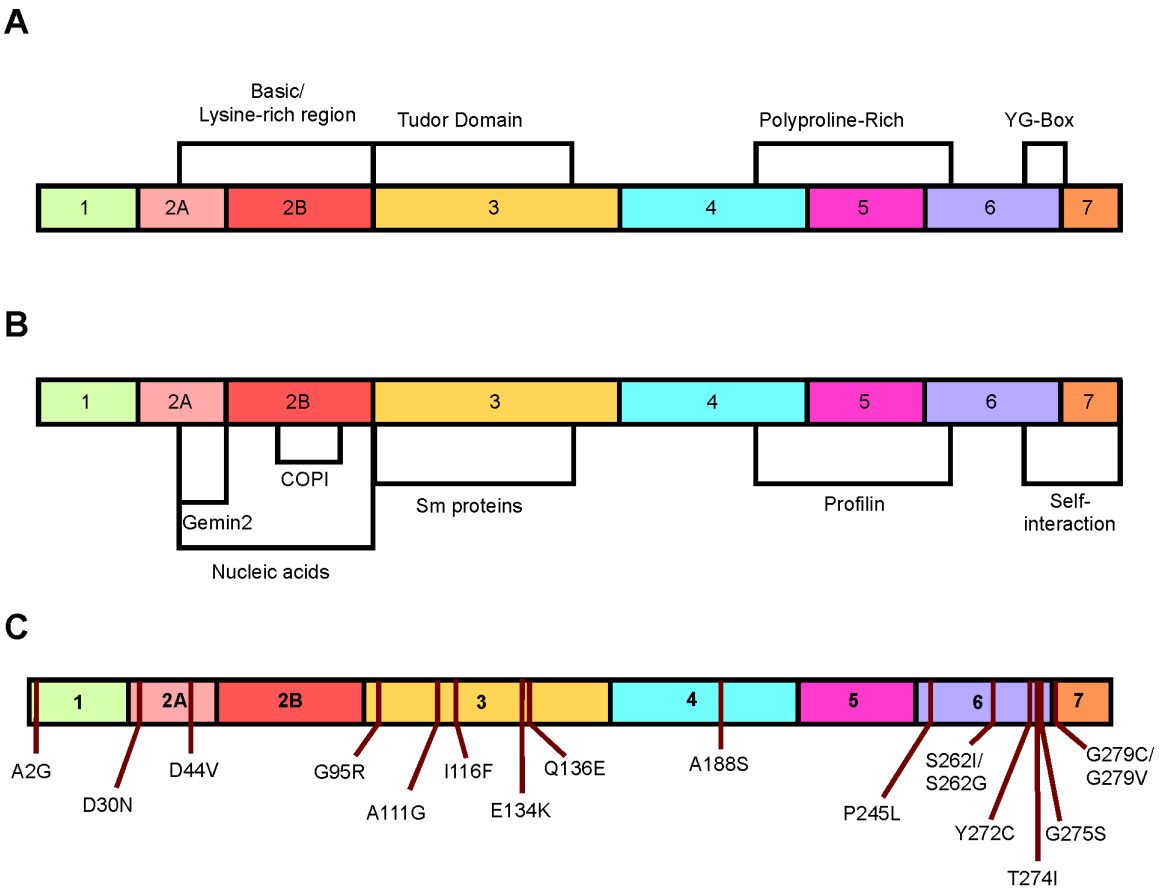


Figure 1. Overview of the SMN Protein

Table 1. Interacting partners of the SMN Protein

Protein	Function of Interaction	Location	Reference
Gemin2	snRNP assembly, DNA recombination	Nucleus, cell body, axon	Fischer et al 1997, Liu et al 1997, Zhang et al 2003, Zhang et al 2006, Todd et al 2010, Takaku et al 2011, Zhang et al 2011
Gemin3	snRNP assembly	Nucleus, cell body, axon	Charroux et al 1999, Todd et al 2010
Gemin4	snRNP assembly	Nucleus, cell body, axon	Charroux et al 2000, Todd et al 2010
Gemin5	snRNP assembly	Nucleus, cell body, axon	Gubitz et al 2002, Todd et al 2010, Yong et al 2010
Gemin6	snRNP assembly	Nucleus, cell body, axon	Pellizzoni et al 2002a, Pellizzoni et al 2002b, Todd et al 2010
Gemin7	snRNP assembly	Nucleus, cell body, axon	Baccon et al 2002, Todd et al 2010
Gemin8	snRNP assembly	Nucleus, cell body, axon	Carissimi et al 2006, Todd et al 2010
Unrip	snRNP assembly	Nucleus, cell body, axon	Carissimi et al 2005, Todd et al 2010
Fibrillarin	snoRNP biogenesis	Nucleus	Pellizzoni et al 2001
GAR1	snoRNP biogenesis	Nucleus	Pellizzoni et al 2001
hnRNP Q	pre-mRNA splicing, mRNA transport, axon outgrowth	Nucleus, axon	Mourelatos et al 2001, Rossoll et al 2002, Chen et al 2012
hnRNP R	pre-mRNA splicing, mRNA transport, axon outgrowth	Nucleus, axon	Rossoll et al 2002, Glinka et al 2010
huD	pre-mRNA splicing, RNA stability, mRNA transport, axon outgrowth	Nucleus, axon	Fallini et al 2011
IMP1	mRNA transport	Axon	Fallini et al 2014
KSRP	pre-mRNA splicing	Nucleus	Tadesse et al 2008
MRLC	Muscle function	Muscle (C2C12 cells)	Shafey et al 2010

Table 1 Continued

ANXA2	Cytoskeletal function	Muscle (C2C12 cells), neurons (PC12 cells)	Shafey etal 2010
FGF-2(23)	Cell growth/differentiation	Nucleus	Claus et al 2003
FMRP	mRNA transport, translation regulation	Nucleus, cell body, axon	Piazzon et al 2008
HDAC11	Chromatin modification, pre-mRNA splicing	Nucleus	Joshi et al 2013
mSin3A	Transcription	Nucleus	Zou et al 2004
NFAR	Transcription	Nucleus	Saunders et al 2001
Ppp4c	snRNP localization	Nucleus, cytoplasm	Carnegie et al 2003
TDP43	RNA processing, nuclear architecture	Nucleus	Wang et al 2002
ZPR1	Nuclear architecture	Nucleus, cytoplasm	Gangwani et al 2001, Ahmad et al 2012

CHAPTER II

SEVERE IMPAIRMENT OF MALE REPRODUCTIVE ORGAN DEVELOPMENT IN A LOW SMN EXPRESSING MOUSE MODEL OF SPINAL MUSCULAR ATROPHY

Eric W. Ottesen,[1¶] Matthew D. Howell,[1¶] Natalia N. Singh,[1] Joonbae Seo,[1] Elizabeth M. Whitley[2] and Ravindra N. Singh[1*]

A paper published in Scientific Reports in 2016

[1]Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA

[2]Department of Veterinary Pathology, Iowa State University, Ames, Iowa 50011, USA

Running Title: Impaired testis development in SMA

¶These authors contributed equally

*Corresponding author: Ravindra N. Singh, Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA; Phone: 1 (515) 294-8505; Email: singhr@iastate.edu

Abstract

Spinal muscular atrophy (SMA) is caused by low levels of survival motor neuron (SMN), a multifunctional protein essential for higher eukaryotes. While SMN is one of the most scrutinized proteins associated with neurodegeneration, its gender-specific role in vertebrates remains unknown. We utilized a mild SMA model (C/C model) to examine the impact of low SMN on growth and development of mammalian sex organs. We show impaired testis development, degenerated seminiferous tubules, reduced sperm count and low fertility in C/C males, but no overt sex organ phenotype in C/C females. Underscoring an increased requirement for SMN expression, wild type testis showed extremely high levels of SMN protein compared to other tissues. Our results revealed severe perturbations in pathways critical to C/C male reproductive organ development and function, including steroid biosynthesis, apoptosis, and spermatogenesis. Consistent with enhanced apoptosis in seminiferous tubules of C/C testes, we recorded a drastic increase in cells with DNA fragmentation. SMN was expressed at high levels in adult C/C testis due to an adult-specific splicing switch, but could not compensate for low levels during early testicular development. Our findings uncover novel hallmarks of SMA disease progression and link *SMN* to general male infertility.

Introduction

Spinal muscular atrophy (SMA) is a leading genetic cause of infant mortality (Awano et al 2014) and results from deletion and/or mutation of *survival motor neuron 1* (*SMN1*), a gene that codes for SMN protein (Awano et al 2014). A nearly identical gene, *SMN2*, cannot compensate for the loss of *SMN1* due to predominant skipping of exon 7, producing SMNΔ7, a truncated protein that is partially functional and highly unstable (Cho and Dreyfuss 2010). The severity of SMA correlates with the level of SMN, a multifunctional protein implicated in regulation of small nuclear heteronuclear ribonucleoprotein (snRNP) biogenesis, transcription, translation, stress granule formation, signal transduction and axonal transport of mRNA (references in Seo *et al*, 2013). Motor neurons are particularly sensitive to the loss of SMN, although reduced SMN independently affects non-neuronal tissues, including muscle (Boyer et al 2013), heart (Heier et al 2010), lungs, and intestine (Schreml et al 2013). SMN is required for male germ cell maintenance in *Drosophila* (Grice et al 2011), however, no parallel can be drawn with mammalian spermatogenesis, which occurs within the specialized microenvironment of seminiferous tubules.

Multiple mouse models recapitulating various aspects of SMA have been generated (Bebee et al 2012). The Taiwanese and Δ7 mice, the best characterized and most widely utilized models, exhibit severe phenotypes, including early postnatal lethality, impaired maturation of neuromuscular junctions and overall deficient motor function (Awano et al 2014, Bebee et al 2012). However, due to

their short lifespan, these models are not appropriate for examining the role of SMN in male reproductive organ development. The testis is unique in producing higher SMN levels due to predominant expression of full-length *SMN2* transcripts compared to all other adult tissues and organs of a mild mouse model of SMA (Chen et al 2008, Chen et al 2015). Alternative splicing of several genes are switched during testicular development (Schmid et al 2013). However, it is not known if *SMN2* exon 7 undergoes a similar switch.

The recently reported $Smn^{C/C}$ (C/C) model expresses a reduced amount of SMN (~25-50% of WT) and displays a mild SMA-like phenotype, including peripheral necrosis, autonomic nervous system dysfunction and allodynia (Osborne et al 2012). Here we employ the C/C mouse to examine the role of SMN in reproductive organ development. We observed reduced testis size and impaired spermatogenesis in C/C mice, despite high SMN expression in testis. We show severe perturbations of the testicular transcriptome in young adult C/C mice, suggesting massive reprogramming of transcription and posttranscriptional regulation. Our results uncover a surprising shift in splicing regulation of various *SMN2* exons during the first wave of spermatogenesis. Our findings implicate for the first time a role for SMN in mammalian reproductive organ development and further elucidate the complex physiological role of this essential protein.

Results

C/C males exhibit small testes with impaired spermatogenesis and reduced fertility

To determine the effect of low SMN on reproductive organ development, we examined wild type (WT) and C/C mice at postnatal day 42 (P42). At this age, mice have completed the first wave of spermatogenesis. Notably, we observed substantially smaller testes in C/C males even after correcting for body weight (Figs.1a-c). Testosterone is a crucial hormone regulating testis development, however, we observed no significant difference in serum testosterone between WT and C/C mice (Fig. 1d). Histological analyses of C/C testes showed heterogeneity of seminiferous tubules, with evidence of degeneration, including vacuolization, multinucleated bodies and sloughed cells (Fig. 1e, right panel). We assessed the health of seminiferous tubules using an established scale (Daigle et al 2009). The average score for C/C testes indicated an overall reduction in post-meiotic cells and disrupted spermatogenesis (Fig. 1f). Histology of C/C epididymis showed a lack of or reduced number of spermatozoa in tubules of all regions (Fig. 1g). Consistently, we observed a ~10-fold reduction in epididymal sperm count in P60 C/C males compared to WT males (Fig. 1h), indicating that spermatogenesis was impaired, but not globally arrested.

Unlike testes, C/C female reproductive organs appeared normal, although variable in size (Fig. 2a). Gross uterus/ovaries mass was reduced for C/C females (Fig. 2b), but not after correction by total body weight (Fig. 2c). Beyond the gross

weight difference, there were no overt histological changes in C/C uterus and ovaries (Fig. 2d). Ovaries from females of both genotypes exhibited follicles with oocytes in various stages, and the corpus luteum and uterus appeared normal (Fig. 2d).

In order to evaluate the fertility of C/C mice, we set up breeding cages with different combinations of WT, C/C, and heterozygous (C/+) mice (Supplementary Table 1) and followed them for ninety days. WT males all sired multiple litters of pups (Supplementary Table 1). In contrast, only two of the eight C/C males sired a litter: one male paired with a WT female and another paired with a C/C female (Supplementary Table 1). Both sired only one litter and neither female appeared pregnant again after giving birth. The fact that two C/C males were able to sire at least one litter despite very low sperm count suggests no debilitating defect in sperm motility and/or fertilization. In contrast, C/C females were fertile; over a 90 day period, there was no significant difference in number of litters born (Fig. 2e) or litter size (Fig. 2f) compared to heterozygote females (Supplementary Table 1).

High levels of SMN are expressed in C/C testis

The recently developed multi-exon-skipping detection assay (MESDA) determines the relative abundance of all *SMN* splice isoforms in a single reaction (Singh et al 2012). We employed MESDA on various tissues of C/C mice. Consistent with previous reports (Chen et al 2008, Chen et al 2015), testis emerged as the only tissue to have full-length *SMN2* transcript as the major product (Fig. 3b), and the

only tissue to express appreciable amounts of full-length product from the hybrid *Smn* gene (Fig. 3c). Supporting a heightened requirement for SMN in testis, we confirmed that there was a very high expression of SMN protein in WT testis compared to all other tissues, including brain and spinal cord (Fig. 3d). When we compared SMN levels in C/C and WT tissues, we observed a ~50% reduction in brain, liver, heart, and uterus/ovaries of C/C mice (Fig. 3e-h). However, there was no difference in SMN protein between WT and C/C testes (Fig. 3h). Although consistent with the results of MESDA, the lack of SMN reduction in testis was at odds with the drastic phenotype and may indicate one or more of the following: (i) SMN expression is reduced at an earlier developmental stage which causes defects at later time points; (ii) the cell types with low SMN expressions are already lost by P42; (iii) or the testis phenotype is an indirect result of low SMN levels in other tissues, such as the nervous system.

C/C testis transcriptome is dramatically altered

We performed deep sequencing of the testis as well as brain and liver transcriptome of WT and C/C mice (Supplementary Table 2) to characterize molecular changes in these tissues (Sequence Read Archive accession number SRP062636). During initial quality control, we determined that one sample derived from a C/C testis was an outlier; it was removed from all further analysis. Using the remaining replicates, we identified 3,724 differentially expressed genes using a false discovery rate (FDR) cutoff of 0.05, with 2,186 genes upregulated and 1,538 downregulated in C/C testes (Fig. 4a). In contrast, very few genes showed aberrant expression in C/C brain and

liver (Fig. 4a) Summary statistics for the 50 most significantly upregulated and downregulated genes are given in Supplementary Tables 3 and 4.

Several Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and gene ontology (GO) terms were disproportionately affected in C/C testes. The top 25 enriched pathways in C/C testis are shown in Fig. 4b, including apoptosis, cell and extracellular matrix interaction (potential problems with tissue organization and/or integrity) and cardiomyopathy and smooth muscle contraction (potential problems with blood flow and/or transport of spermatozoa to epididymis). GO term analysis was somewhat less informative (Supplementary Figs. S1-S3), although several GO terms associated with spermatogenesis were enriched in the list of downregulated genes (Supplementary Fig. S3). We also examined expression of several genes that are previously identified to be mutated in cases of oligospermia (Hwang et al 2010); of the fourteen genes, six (*H19*, *Klhl10*, *Prm1*, *Shbg*, *Tssk4*, and *Vasa*) showed significantly altered expression in C/C testes (Supplementary Fig. 4). Aside from pathways and genes, we also detected changes suggesting altered enrichment of genes in specific groups of testicular cell types (Margolin et al 2014). Consistent with our histological analyses, genes that are highly expressed in somatic and early spermatogenic cell types were strongly enriched, and genes expressed in late spermatocytes and spermatids were found to be downregulated (Fig. 4c). We also examined the relative expression of individual markers specific for testicular cell types (Supplementary Fig. 5). These results suggest extensive cell death of pachytene spermatocytes and/or arrest of zygotene spermatocytes in C/C testes.

We independently validated the results of RNA-Seq by quantitative PCR (QPCR) of >50 candidate genes (full list in Supplementary Table 5). There was a strong correlation (r^2=0.8376) between RNA-Seq and QPCR results (Fig. 4d). Five genes in particular (*Lamb2*, *Ddr1*, *Apoe*, *Cpe*, *SerpinG1*) were highly upregulated, ranging from ~1.5-4 fold above WT (Fig. 4e). We also tested four genes, which were predicted by RNA-Seq to be highly downregulated. Of these, *Npy*, a gene that codes for neuropeptide Y pro-protein, was strongly downregulated, though *Npy* expression was highly variable. The other three genes (*Tppp2*, *Ppp2r2b*, and *Oaz3*) all trended downward, but the decrease was only statistically significant for *Tppp2* (Fig. 4f). Of the five steroid biosynthesis genes (*Lss*, *Msmo1*, *Dhcr7*, *Lipa*, and *Dhcr24*) that we tested, all were significantly upregulated (Fig. 4g). In addition, we tested 5 genes participating in spermatogenesis; however, although all trended downward, we observed a significant decrease only in *Gapdhs* (Fig. 4h). Many genes involved in axon guidance were affected; of these, the five most strongly upregulated were *Ephb1*, *Efnb1*, *Ntn3*, *Cxcl12*, and *Slit3* (Fig. 4i). *Ppp3r2* was the only gene in this category that was significantly downregulated (Fig. 4i). Several long non-coding RNAs (lncRNAs) had altered expression levels in C/C testes. Of the five that were validated, *Meg3*, *Neat1*, and *Malat1* were strongly upregulated (Fig. 4j). We examined the expression of all the *Gemins*, which code for proteins associated with SMN complex assembly and function. Only *Gemin4* was significantly downregulated (Fig. 4k).

Apoptotic pathways are dysregulated in C/C testes

Our RNA-Seq results suggested significant dysregulation of the apoptotic pathways in C/C testes (Fig. 4b), with much fewer changes in brain and liver (Fig. 5a). We confirmed by QPCR that a number of pro- (*Bax*, *Casp7*, *Casp8*, *Casp9*, *Capn2*, *Aifm1*) and anti- (*Bcl3*, *Akt*) apoptotic genes were significantly upregulated in C/C testis (Fig. 5b), but not in brain (Fig. 5c) or liver (Fig. 5d). Terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) staining, which detects DNA fragmentation usually associated with apoptosis, revealed a ~3-fold increase in the percentage of seminiferous tubules with at least one TUNEL-positive cell (Fig. 5e,f). We also observed a significant increase in the average number of TUNEL-positive cells per tubule (Fig. 5g). For WT testes, TUNEL-positive cells were predominantly spermatocytes and the remaining were spermatogonia (Fig. 5h). These WT TUNEL-positive cells most likely represented damaged germ cells that must be eliminated before meiotic division into abnormal spermatids (Aitken et al 2011). Most TUNEL-positive cells in C/C testes were spermatocytes, similar to WT, but we also observed TUNEL staining of a small number of Sertoli cells and round spermatids (Fig. 5h). Since depletion of Sertoli cells can have drastic consequences for germ cell health (Orth et al 1988), increased apoptosis in spermatogonia and spermatocytes may be due to dysfunction of Sertoli cells, or vice versa. Immunostaining for cleaved caspase 3, a marker for apoptosis, revealed positive spermatogonia (Fig. 5i, panel I), spermatocytes (panel II) and multinucleated bodies (panel III) in C/C testes, but a lack of cleaved caspase 3-positive cells in WT testes (panel IV). This observation further underscores that apoptosis is dysregulated in C/C testes.

SMN2 undergoes splicing switch during testicular development

The first wave of spermatogenesis in mice occurs from P10 to P35. This process is characterized by synchronous transitions from one cell type to another and marked by predictable changes in the testis transcriptome (Bellve et al 1977, Margolin et al 2014) (Fig. 6a). We employed MESDA to capture splicing of *SMN2* exons during the first wave of spermatogenesis. In particular, we compared the splicing pattern of *SMN2* in P7 testes, when spermatogonia are the predominant germ cell type, with testes undergoing meiosis (P10, P12 and P18), and near the end (P30) and after completion of the first wave of spermatogenesis (P42). We observed a dramatic shift in *SMN2* exon 7 splicing from overwhelming skipping at P7 to higher inclusion at P18 and later (Fig. 6b). We also observed increased skipping of exon 5 and decreased co-skipping of exons 3 and 7 at later time points (Fig. 6b). With regards to hybrid *Smn* (Fig. 3a), there was an increase in exon 7 inclusion starting at P18 (Fig. 6c). Our findings represent the first report of a shift in *SMN2* splicing pattern during testis development. Increased *SMN2* exon 7 inclusion coincided with the completion of initial stages of meiosis and likely occurred in spermatocytes and spermatids. *SMN2* transcripts of somatic testicular cells and spermatogonia, which are more highly represented at early time points were likely still dominated by exon 7 skipped forms, and therefore likely had a deficiency in SMN protein expression.

Previous evidence suggested that hnRNP Q and Tra2β are important for *SMN2* exon 7 inclusion in testes (Chen et al 2008, Chen et al 2015). Thus, we

examined whether changes in expression of splicing factors known to regulate *SMN2* exon 7 splicing coincided with the changes in *SMN2* splicing during testis development. hnRNP A1 and Q protein levels were highest at P7 and dropped during testis development, dramatically so in the case of hnRNP A1 (Fig. 6d). Since hnRNP Q was expressed at relatively low levels at P18 and P30, when *SMN2* exon 7 inclusion increased, it likely did not play a role in the splicing switch. hnRNP A1, however, is a known negative regulator of *SMN2* exon 7 inclusion, so its low expression in later time points is consistent with the change in exon 7 splicing. Tra2β was expressed at the highest levels at P18 and P30, when *SMN2* exon 7 inclusion is highest, followed by a decrease at P42 (Fig. 6d). hnRNP A2/B1, a negative regulator of *SMN2* exon 7 inclusion, had a nearly identical expression pattern. Overall, these data indicate that multiple splicing factors contribute to the *SMN2* splicing pattern we observe during early stages of testicular development.

In order to further verify that SMN levels are altered during testis development, we first established baseline expression during the first wave of spermatogenesis using primers targeted to mouse *Smn* mRNA in WT testes. *Smn* expression peaked at P18 and P30 and then dropped slightly at P42 (Fig. 7a). In contrast, full-length *SMN2* expression in C/C testes dramatically increased at P18 and held steady thereafter (Fig. 7b). We next examined how SMN protein expression changed during testis development. In WT testes, SMN protein (derived from *Smn*) steadily increased and peaked at P30 with a slight decrease at P42 (Fig. 7c). In C/C testes, however, SMN protein (derived from *SMN2* and hybrid *Smn*) was low until a

marked increase at P18, peak expression at P30 and a subsequent decrease at P42 (Fig. 7d). When SMN protein was compared between the genotypes, C/C SMN protein level was approximately half the WT level from P7 to P12, but by P18 there was no significant difference between the genotypes (Fig. 7e,f). We also examined expression of Gemin2, a protein that interacts with SMN. Similar to SMN, Gemin2 protein level was reduced by about half from P7 to P12 in C/C testes. Gemin2 rose to near WT level by P30 followed by a drop at P42 (Fig. 7e,g). This disparity in SMN and Gemin2 levels may suggest a deficiency of SMN functions mediated by Gemin2.

Dysregulation of alternative splicing in adult C/C testes

To confirm that critical alternative splicing events are affected during C/C testes development, we examined splicing patterns of several genes known to undergo regulated alternative splicing during spermatogenesis. *Add3* exon 14 inclusion increases during spermatogonia differentiation (Schmid et al 2013). Both WT and C/C testes exhibited a similar increase in *Add3* exon 14 inclusion, but at P42 there was a decrease in exon 14 inclusion in C/C compared to WT (Fig. 8a). *Lrrc16a* exon 38 splicing is somewhat complicated: pattern changes from predominant exclusion to inclusion during Sertoli cell maturation and spermatogonia differentiation while round spermatids exhibit predominant exclusion (Schmid et al 2013). In C/C testes, exon 38 inclusion followed a pattern similar to WT at P7 and P18, but not P10, and subtle differences emerged at P30 and P42 (Fig. 8b). In Sertoli cells and spermatogonia, there is a mix of *Picalm* exon 13 inclusion and exclusion, while spermatocytes and round spermatids exhibit predominant exon 13 inclusion[11].

Splicing of *Picalm* was similar for WT and C/C testes except at P42 where there was more exon 13 exclusion in C/C (Fig. 8c). Prevalence of a *Picalm* exon 13 alternative 3′ splice site was similar in WT and C/C testes (Fig. 8c).

In addition, RNA-Seq data revealed differential regulation of alternative splicing of several genes. We selected two such genes for validation. *Wt1* is a tumor suppressor gene and essential for proper urogenital development (Kreidberg et al 1993). In WT testes, *Wt1* exon 5 is predominantly skipped at P7, with a gradual increase in inclusion up to P30 (Fig. 8d). C/C testes showed a similar pattern, but there was a decrease in exon 5 inclusion compared to WT testes at P30 and P42 (Fig. 8d). Skipping of *Sulf1* exon 21 is predicted to cause a frameshift eliminating the stop codon and resulting in a much longer protein. In WT testes, *Sulf1* exon 21 shifted from complete inclusion at P7 to roughly 40% at P18, then subsequently increased again (Fig. 8e). In C/C testes, exon 21 skipping remained the predominant event (Fig. 8e).

Gene expression of candidate genes is not affected until after puberty

Having observed the preponderance of splicing changes between P30 and P42 in C/C testes, we wanted to know whether expression of key genes followed a similar pattern. We tested mRNA levels throughout development of several genes that were strongly upregulated or downregulated (Fig. 4e-f) in C/C testes by QPCR. In WT, *Apoe* was expressed highest at P18, followed by P30 and P42, and at similar levels for all other time points. However, there was a drastic increase in expression at P42

in C/C testes, although levels did not pass the threshold value we set for FDR (Fig. 8f). *Cpe* was expressed at the highest level from P10 to P18, with the lowest expression at P42. In C/C testes, expression was significantly higher than WT at P42 (Fig. 8g). In WT, *Ddr1* steadily decreased from P7 onwards. The pattern in C/C testes matched with WT until a drastic increase at P42 (Fig. 8h). *Lamb2* expression steadily increased until P18 and then fell again. In C/C, the pattern was the same with the exception that expression did not drop at P42. Expression in C/C was significantly lower at P7, although the absolute change in expression was small (Fig. 8i). *SerpinG1* expression in WT increased slightly at P18 and then dropped steadily until P42, whereas in C/C expression was highly increased at this time point (Fig. 8j).

Npy expression exhibited quite large variation, both between different time points and between replicates. However, the average values at each time point were quite similar between WT and C/C testes, with the exception of P42, at which C/C exhibited a striking, though not statistically significant decrease, similar to what we observed before (Fig. 8k). *Tppp2*, *Oaz3*, and *Ppp2r2b* exhibited strikingly similar expression patterns, with very low levels until P18, then drastically increased at P30 and P42, coinciding with the emergence of spermatozoa. For all three genes, there were no differences in expression between WT and C/C testes until P42, when expression of *Oaz3* and *Tppp2* was significantly reduced, and expression of *Ppp2r2b* trended downward (Fig. 8l-n).

Overall, none of the changes that we observed occurred until P42, after completion of the first wave of spermatogenesis. All the downregulated genes, with the exception of *Npy*, were only expressed at very low levels until P30, indicating that they are most likely spermatid- or spermatozoa-specific. This downregulation is consistent with the observation that C/C testes were generally lacking these cell types (Fig. 1). For genes upregulated in C/C testes, expression in WT was generally lowest at P42, except *Apoe*. Interestingly, expression of *Apoe* in C/C testes at P42 was higher than WT at any time point, indicating that there is a mechanism beyond developmental dysregulation causing increased expression of this gene.

Discussion

In this report, we employed the C/C mouse to evaluate for the first time the role of SMN in mammalian sex organ development and fertility. We observed severe impairment of male reproductive organ development characterized by smaller testes, widespread degeneration of seminiferous tubules, loss of post-meiotic cells, and a drastic reduction in male fertility in C/C mice. These surprising findings underscore a requirement for a relatively high level of SMN for the development and maintenance of male reproductive organs in mammals. A report published in 1986 implicated testis dysfunction in two mild SMA patients (Richert et al 1986), although genetic diagnoses were not available at that time. A subsequent report briefly mentioned unexplained clinical symptoms due to the influence of male sex in type 3 SMA patients (Zerres and Rudnik-Schoneborn 2003). Although the authors

did not elaborate on the exact nature of the symptoms, they conveyed the necessity for an animal model of mild SMA to examine male specific pathology. Hence, our finding of a unique testicular phenotype in C/C mice fills a critical knowledge gap with respect to SMN function.

The testis has unique RNA metabolism, including transcription, splicing and 3′-end processing (Berkovits et al 2012, Schmid et al 2013, Song et al 2014) Given the prominent role of SMN in RNA metabolism and the presence of Gemin3, an SMN-interacting protein, in germ cell-specific chromatoid bodies (Meister and Fischer 2002, Zhang et al 2008, Ginter-Matuszewska et al 2011), it is not unexpected that SMN plays a critical role in testis development and spermatogenesis. However, it was somewhat unexpected that we did not observe drastically altered gene expression until an adult age, when SMN expression in C/C testes reached high levels due to high inclusion of *SMN2* exon 7. Gemin2 has a stabilizing effect on SMN and its interaction with SMN is critical for most SMN functions (Ogawa et al 2007). Low level of Gemin2 in C/C testes at P42 is likely to impact all SMN functions dependent on SMN-Gemin2 interactions. A dramatic change in the transcriptome of C/C testis at P42 underscores perturbations in pathways critical to spermatogenesis and testis function, including steroid biosynthesis, apoptosis, and spermatogenesis itself (Fig. 4). There was an enrichment of markers of interstitial cells and somatic cells of the seminiferous tubules and general loss of expression of genes enriched in post-meiotic cell types (Fig. 4, Supplementary Fig. 4). Intergenic lncRNAs and transposons expressed during meiosis are degraded through the Piwi-interacting

RNA (piRNA) pathway in which Tudor domain-containing proteins play an essential role (Gou et al 2014). SMN being a tudor domain-containing protein and given the propensity of tudor domains to forge interactions among themselves (Ginter-Matuszewska et al 2011, Tripsianes et al 2011), it is possible that SMN directly or indirectly affects piRNA pathway.

It is known that lncRNAs are differentially regulated in testis, especially during spermatogenesis (Sun and Wu 2015, Washietl et al 2014). The most highly affected lncRNAs in C/C testes were *Meg3*, *Malat1* and *Neat1*. *Meg3* is expressed from paternally imprinted loci in mammals (Li et al 2004) and is known to act as a tumor suppressor (Zhou et al 2007). *Malat1* functions by affecting the phosphorylation of splicing factors (Tripathi et al 2010), and *Neat1* serves as a scaffold for the formation of paraspeckles in which many RNA binding proteins are sequestered (Clemson et al 2009). All types of non-coding RNAs and untranslated regions impact the availability of RNA binding proteins and control the fate of mRNAs during spermatogenesis (Idler et al 2012). Altered expression of lncRNAs reveals a novel function of SMN in controlling the levels of freely available RNA-binding proteins in testes.

Apoptosis plays an essential role in spermatogenesis as it eliminates damaged meiotic and post-meiotic cells in testes (Aitken et al 2011, Shukla et al 2012). SMN has anti-apoptotic properties (Kerr et al 2000, Vyas et al 2002, Parker et al 2008) and interacts with Bcl-2 to protect cells from Fas-mediated apoptosis

(Iwahashi et al 1997, Sato et al 2000). Very low levels of SMN trigger DNA fragmentation in muscles and motor neurons of SMA patients (Simic et al 2000, Stathas et al 2008). Our finding of testicular defects and increase in TUNEL-positive cells in C/C seminiferous tubules suggests a necessity for much higher levels of SMN to prevent the dysregulation of apoptotic pathways in this tissue. Sertoli cells regulate germ cell survival through the Fas system; Fas ligand decorating the surface of Sertoli cells interacts with Fas present on germ cells (Lee et al 1997). Sertoli cells are the major cell type in testes until P18 (Bellve et al 1977), the time during which SMN is most strongly reduced in C/C testes. Since SMN modulates Fas-mediated apoptosis, low SMN in Sertoli cells could disrupt the normal regulation of this process. Further, low SMN causes defects in snRNP assembly and expression levels (Zhang et al 2008), resulting in accumulation of the aberrantly spliced mRNA isoforms, a condition to which Sertoli cells are particularly sensitive (Lee et al 1997, Bao et al 2015). SMN has also been implicated in the resolution of double-stranded breaks requiring expression of functional H2AX (Takaku et al 2011). Hence, increased TUNEL staining in C/C testes could be due to an increase in the unresolved double-stranded breaks arising from meiotic recombination.

Substantial downregulation of mRNA encoding neuropeptide Y proprotein suggests possible denervation of C/C testes and is likely a downstream effect of the low levels of SMN at the early stages of testicular development. Supporting this argument, SMN-independent defects leading to mild SMA display abnormal male reproductive organ development and defective spermatogenesis. The most notable

among these is the mutations in *Vps54* gene leading to a motor neuron disease similar to mild SMA and defects in male reproductive organ development (Schmitt-John et al 2005). Plastin 3 coded by *PLS3* is a protective modifier of SMA caused by the loss of *SMN1* (Oprea et al 2008). Interestingly, the protective effect of Plastin 3 is not fully penetrant and appears to be specific to females. A recent study in rat testis implicates the role of Plastin 3 in spermatogenesis (Li et al 2015). We did not observe aberrant expression of *Vps54* or *Pls3* genes in C/C testis, excluding a direct role of these genes in testicular defects in C/C mice. However, our results combined with the above findings suggest an overlapping pathway of neurodegeneration and male reproductive organ development.

About 50% of male infertility cases, affecting >5% of the population worldwide, are caused by genetic abnormalities (Comhaire et al 1987, Hwang et al 2010). More than 20% of these cases represent nonobstructive azoospermia (NOA) in which semen contains little or no sperm (Huynh et al 2002, Hwang et al 2010). A genome-wide association study (GWAS) reported in 2012 linked only 6 genes to male fertility traits (Kosova et al 2012), none of which were linked to pre-mRNA splicing, which is uniquely regulated during spermatogenesis. However, a GWAS conducted on Chinese men focused on serine-arginine (SR) protein-coding genes and found a strong correlation between NOA and mutations in *SFRS9*, encoding SR protein SRp30c (Ni et al 2014). Incidentally, SRp30c regulates SMN levels by promoting inclusion of exon 7 during pre-mRNA splicing (Young et al 2002). Other studies provide additional evidence supporting the role of SMN in spermatogenesis.

For example, the transcription factor SP1 is predicted to modulate SMN levels in both mouse and humans (Rouget et al 2005). A critical role for SP1 in mouse male germ cell development has been previously described (Thomas et al 2007). Further, a recent study suggested that PRMT5, which plays an essential role in assembly of spliceosomal snRNPs along with SMN (Meister and Fischer 2002), is required for germ cell survival during spermatogenesis in mice (Wang et al 2015). Finally, upregulation of SMN upon activation of phosphatidylinositol 3-kinase (PI3K)/AKT and its positive impact on disease phenotype of severe SMA has been demonstrated (Branchu et al 2013). Abrogation of PI3K/AKT pathway has been shown to affect spermatogenesis but not oogenesis (Baker et al 2014). Our finding of a reproductive phenotype and loss of fertility in C/C males, but not females, are consistent with those observations. Incidentally, several human studies have revealed fewer females than males afflicted with the milder forms of SMA (Furukawa et al 1968, Hausmanowa-Petrusewicz et al 1979, Jedrzejowska et al 2009). Given these findings, it is rather expected that SMN plays a critical role in testis development and male germ cell survival in mammals.

Humans are unique in harboring *SMN2*, a gene whose function is still unknown. Generally, *SMN2* is considered to be dispensable for survival, since enough SMN could be generated from *SMN1*. Our findings provide a strong rationale that the preservation of *SMN2* during evolutions of humans was driven by the heightened requirement of SMN in male reproductive organ development. A very specific splicing switch of *SMN2* to elevate SMN levels in adult testis in humans

appears to be part of a fine regulatory mechanism that remains to be further understood. The presence of *SMN2* in SMA patients carries profound significance due to its ability to serve as a promising therapeutic target. Several pre-clinical studies employing *SMN2*-containing severe mouse models of SMA have shown extraordinary survival benefits[3]. However, gender-specific efficacy of drugs and reproductive phenotype of severe SMA mice that survive much longer upon treatment, remain unknown. Lack of gender-specific data in pre-clinical studies in neurodegenerative diseases has been cited as one of the major sources of misinterpretations and faulty conclusions (Perrin 2014). Our findings validate those concerns and provide a repertoire of histological and molecular markers that could be employed for the careful monitoring of male reproductive organ health as an essential component of drug development process for SMA.

Materials and Methods

All methods were performed in accordance with the approved biosafety and radiation safety guidelines of Iowa State University (ISU), adhering to the federal and state guidelines. All animal experiments were carried our in accordance with the approved protocols by the Institutional Animal Care and Use Committee (IACUC) of ISU, adhering to the guidelines of American Veterinary Medical Association (AVMA), United States Health and Human Services (US HHS), United States Department of Agriculture (USDA) and State of Iowa.

Mice

Mice heterozygous for the C allele (Osborne et al 2012) ($Smn^{C/+}$) on the C57BL/6 background were purchased from Jackson Laboratory (stock number 008714). Breeding cages with one male and one female heterozygote were set up and maintained at ISU to generate the mice used for all experiments. This cross generated WT mice ($Smn^{+/+}$) and C/C mice ($Smn^{C/C}$). Pups were genotyped and sacrificed at the appropriate ages. Unless otherwise stated, all experiments were carried out at P42.

Fertility assessment

Monogamous breeding pairs were established between three- to five-month-old mice and followed for ninety days. Once established, females were checked every twelve hours for a vaginal plug for the first five days. The number of litters and number of pups per litter were recorded. After ninety days mice were weighed, euthanized and tissues were collected. Testis mass and epididymal sperm count were determined for male mice.

Epididymal sperm counting

P60 WT and C/C mice were anesthetized with isoflurane and killed by cervical dislocation. The epididymides were dissected, minced with a fresh number 10 disposable scalpel and placed in 3 ml Dulbecco's Modified Eagles' Medium supplemented with 10% fetal bovine serum (Life Technologies) for 15 minutes at 37°C. After incubation, the cell suspension was triturated 15 times. Dilutions of the

sperm were prepared in water and incubated at room temperature for two minutes. The diluted cell suspension was loaded on a haemocytometer, allowed to settle, and only sperm with a head and tail were counted.

Tissue collection

For RNA and protein isolation, P7, P10, P12, P18 and P30 WT and C/C mice were deeply anesthetized with isoflurane and decapitated. Tissues were removed and stored at -80°C for future use. P42 mice were deeply anesthetized with 100 mg/kg ketamine and 10 mg/kg xylazine, the carotid artery was pierced and blood collected and the mouse was subsequently transcardially perfused with phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, pH 7.4). Unless otherwise stated, tissues were removed, flash frozen on dry ice, and stored at -80°C for future use. Blood was allowed to clot at room temperature for 15 minutes, centrifuged at 2,000 x *g* for 10 minutes at 4°C and the serum stored at -80°C for future use.

For histology, P42 WT and C/C mice were weighed, deeply anesthetized with 100 mg/kg ketamine and 10 mg/kg xylazine and transcardially perfused with PBS followed by 4% paraformaldehyde (pH 7.4). After perfusion, the sex organs were removed, weighed with an analytical balance (Ohaus) and post-fixed overnight in modified Davidson's fixative (Electron Microscopy Diatome) at 4°C. Organs were processed, embedded in paraffin, sectioned and stained with haemotoxylin and eosin (H&E).

Histological analysis of tissue

H&E stained sections were analyzed by a board-certified Veterinary Pathologist to determine any pathological changes. Photographs of sections were taken with a Nikon microscope with SPOT Advanced software (Diagnostic Imaging, Inc.).

We utilized a previously published 10-point scoring system to assess the progression of spermatogenesis within the seminiferous tubules (Daigle et al 2009). Twenty seminiferous tubules from all areas of each testis were scored. The pathologist was blinded as to genotype of each sample. The scoring system is as follows:

1 - No cells in tubular cross section

2 - Sertoli cells only

3 - Only spermatogonia present

4 - No spermatozoa, no spermatids, but fewer than 5 spermatocytes present

5 - No spermatozoa, no spermatids, but many spermatocytes present

6 - No spermatozoa, but fewer than 5-10 spermatids present

7 - No spermatozoa, but many spermatids present

8 - All stages of spermatogenesis present, but fewer than 5-10 spermatozoa

9 - Many spermatozoa present, but germinal epithelium disorganized with marked

sloughing or obliteration of lumen

10 - Complete spermatogenesis

Terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) staining

Three-μm formalin-fixed paraffin-embedded testis cross sections from P42 mice were subjected to TUNEL staining with the *In situ* Cell Death Detection Assay, Fluorescein (Roche) following the manufacturer's protocol. For the negative control, two slides were incubated with only Labeling Solution (no enzyme). For the positive control, one slide was incubated with 2 U/μl DNase I, recombinant (Roche) for 10 minutes at room temperature to induce DNA fragmentation and then incubated as per the kit protocol. After labeling, sections were coverslipped with VectaShield Mounting Medium with DAPI (Vectastain) and imaged. Eight to ten micrographs from all regions of the testis were captured for analysis. TUNEL positive bodies were counted in at least 50 tubules from all regions of the testis for quantification.

Immunohistochemistry

Three-μm formalin-fixed paraffin-embedded testis cross sections from P42 mice were deparaffinized in xylene and rehydrated to distilled water. The sections were subjected to antigen retrieval for 30 minutes at 95°C in antigen retrieval buffer (10 mM sodium citrate, 0.05% Tween-20, pH 6.0). The sections were then incubated in 0.3% hydrogen peroxide (Fisher) for 30 minutes to block endogenous peroxidase activity. Immunostaining was performed using the VECTASTAIN Elite ABC-Peroxidase kit (Vector Laboratories) following the manufacturer's protocol. The primary antibody, rabbit anti-cleaved caspase 3 (Cell Signaling; 9661) was diluted 1:100 and incubated with sections for 90 minutes at room temperature. Detection

was performed using the ImmPACT DAB substrate (Vector Laboratories). Slides were subsequently stained with modified Mayer's hematoxylin (Fisher), dehydrated in an ethanol gradient, coverslipped with Permount (Fisher) and imaged.

Testosterone enzyme linked immunosorbent assay (ELISA)

Total serum testosterone was measured using the Mouse/Rat Testosterone ELISA kit (CalBiotech) using the manufacturer's instructions. A standard curve was constructed using GraphPad Prism v6.0 using a sigmoidal 4-parameter logistics fit and testosterone concentration (ng/ml) were determined based on this curve.

Protein isolation

Tissue was homogenized in ten volumes radioimmunoprecipitation assay (RIPA) buffer (Boston Biotechnologies) with 1X HALT protease inhibitor (Thermo Fisher Scientific). The tissue was sonicated on ice three times for 10 seconds each with one minute between pulses at speed 2 with a Microson Ultrasonic Cell Disruptor (Misonix). The homogenate was incubated on ice for 30 minutes, centrifuged at 12,000 rpm for 15 minutes at 4°C and the supernatant transferred to a new tube. If the supernatant still appeared cloudy after centrifugation, the previous step was repeated. Protein concentration was determined with the BioRad Protein Assay. For immature (P7-P30) testes, tissue was homogenized in a glass tissue grinder (Wheaton), then processed the same as above, omitting the sonication step.

Immunoblotting, silver staining, and densitometric analysis

Protein lysates were diluted with 2X Laemmli sample buffer (BioRad) with 10% β-mercaptoethanol (Sigma-Aldrich). Samples were boiled for 10 minutes and then centrifuged at maximum speed for 5 minutes at room temperature. Different amounts of protein were loaded depending on the tissues: 12.5-15 μg for testis, 30 μg for uterus/ovaries and brain, 50 μg for liver and 75 μg for heart. For silver staining, protein was separated on 10% SDS-PAGE gels and then stained with the BioRad Silver Stain Plus kit according to the manufacturer's protocol. For western blotting, protein was separated on 10% SDS-PAGE gels and then transferred to Immobilon polyvinylidene fluoride (PVDF) membrane (Millipore) using the Turbo Transfer system (BioRad). Membranes were briefly washed in 1X Tris-buffered saline with Tween 20 (TBST; 50 mM Tris, 150 mM NaCl, 0.05% Tween-20, pH 7.4), and then incubated in 5% nonfat milk prepared in TBST. Membranes were then incubated in the appropriate primary antibody diluted in 5% milk: mouse anti-SMN (BD Biosciences; 610646) diluted 1:2,000 (for heart, liver, and uterus/ovaries) or 1:5,000 (for testes), mouse anti-Gemin2 (Clone 2E17; Sigma G6669) diluted 1:500; mouse anti-hnRNPA1 (Clone 9H19; Abcam; ab5832) diluted 1:10,000; mouse anti-hnRNPA2/B1 (Abcam; ab6102) diluted 1:1,000; mouse anti-hnRNP Q (Sigma; R6353) diluted 1:800; goat anti-Tra2β (Abcam; ab31353) diluted 1:1,000; or rabbit anti-Actin (Sigma; A2066) diluted 1:2,000. Incubations were performed overnight at 4°C with shaking, except anti-Actin, which was incubated for 1 hour at room temperature. Blots were then washed three times in TBST for at least 10 minutes with shaking and incubated in the appropriate secondary antibody: goat anti-mouse

IgG, goat anti-rabbit IgG or donkey anti-goat IgG all conjugated to horseradish peroxidase (HRP) for 1 hour at room temperature or for 30 minutes at 37˚C with shaking. Blots were then developed with either Clarity HRP substrate (BioRad) or West Femto substrate (Thermo Scientific Pierce) and visualized with the BioSpectrum AC Imaging System (UVP). For each blot, the mean intensity of each band was determined using ImageJ software. For each sample, the mean intensity of the protein of interest was divided by the mean intensity of Actin. For direct comparisons between WT and C/C, the WT average was set at 1.0 and all other samples were expressed relative to this value. For all time-course experiments within a genotype, the maximum expression was set at 1.0 and all other samples were expressed relative to this value.

Deep sequencing and analysis of C/C transcriptome

RNA was isolated from P42 testes, liver, and brain (n=4 WT, n=3 C/C for testis, n=2 males, 2 females per genotype for liver and brain) using Trizol reagent according to the manufacturer's instructions. The isolated RNA was quantified using a UV spectrophotometer and 10 μg of total RNA was subjected to digestion with RQ1 RNase-free DNase (Promega). After DNase treatment, the RNA was repurified using phenol:chloroform extraction followed by ethanol precipitation. Quality and quantity of samples was assessed on an Agilent bioanalyzer using an RNA Nano chip (Agilent). 300 ng of RNA was used as input for poly(A) + RNA enrichment using the NEBNext poly(A) mRNA Magnetic Isolation Module (New England Biolabs) and RNA-Seq libraries were generated using the NEBNext Ultra Directional RNA Library

Prep Kit (New England Biolabs). During library preparation, barcode sequences were introduced for identifying sequences from each biological replicate and 6-7 libraries were pooled and run per lane on an Illumina HiSeq 2500 using a 100-cycle single-end program.

After sequencing, contaminating adapter sequences and low-quality ends were removed using the cutadapt program (Martin 2011). After trimming, reads were mapped to the mouse genome (version GRm38/mm10) using Tophat (Trapnell et al 2009). For transcript identification, the Gencode annotation was used (http://www.gencodegenes.org/). In order to identify differentially regulated transcripts, read counts were obtained for all Gencode annotated genes using HTSeq-count (Anders et al 2015) and altered expression levels were tested using the DESeq package implemented in R (Anders and Huber 2010). GO term and KEGG pathway analysis were performed using the WebGESTALT web server (http://bioinfo.vanderbilt.edu/webgestalt/). Candidate alternatively spliced exons were identified using MATS (Shen et al 2012).

Quantitative PCR and semi-quantitative RT-PCR

RNA was isolated from testes using the same protocol as total RNA extraction used for deep sequencing, except 30 μg of RNA was used for DNase treatment and cleanup was performed on RNeasy columns (Qiagen). 2.5 μg of total RNA was converted to cDNA using SuperScript III reverse transcriptase (Life Technologies) using random primers (Promega). For quantitative PCR, 1.5 μl of a 1:20 dilution of cDNA (equivalent to cDNA produced from 9.375 ng of RNA) was used as template in

a 20 μl reaction containing 300 nM forward and reverse primers and 1X FastStart Universal SYBR Green Master containing ROX reference dye (Roche). QPCR was performed using a Stratagene MX3005P thermocycler (Agilent) and relative expression levels were determined using the ΔΔCt method using β-Actin as a reference gene. For comparisons in adult tissues, values are expressed as relative to WT average. For time-course experiments, values are expressed as relative to WT average at P7. MESDA was performed as described previously[14]. For other semi-quantitative PCR reactions, 0.5 μl of cDNA was used per 20 μl PCR reaction using Taq polymerase (New England Biolabs) in the presence of 0.8-1.2 μCi of [α-^{32}P] dATP. 6 μl of PCR product was separated on a 5% (for MESDA) or 8% (for other semi-quantitative PCR) TBE polyacrylamide gel, dried, and exposed to a phosphorimager screen. Gel images were scanned using a Fujifilm FLA-5100 and densitometric quantification was carried out using Multi Gauge software (Fujifilm). For all reactions labeled with [α-^{32}P] dATP, values were corrected by dividing by the number of A/T base pairs in the product. All primer sequences appear in Supplementary Table 6.

Statistical analysis

Statistical analysis was performed with GraphPad software. All data are presented as mean ± standard error of the mean (S.E.M.) unless otherwise stated. Data on sex organ weights, sperm count, fertility, TUNEL staining, and protein densitometry were analyzed with two-tailed unpaired Student's *t* test. QPCR results were analyzed with unpaired two-tailed unpaired Student's *t* test with multiple

hypothesis correction using the Benjamimi and Hochberg method. Seminiferous tubule score means were analyzed with the Mann-Whitney U test. For all statistical analyses, $p<0.05$ was considered significant.

References

Aitken, R.J., Findlay, J.K., Hutt, K.J., and Kerr, J.B. (2011). Apoptosis in the germ line. Reproduction *141*, 139-150.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology *11*, 12.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Awano, T., Kim, J.K., and Monani, U. (2014). Spinal Muscular Atrophy: Journeying From Bench to Bedside. Neurotherapeutics *11*, 786-795.

Baker, M.D., Ezzati, M., Aloisio, G.M., Tarnawa, E.D., Cuevas, I., Nakada, Y., and Castrillon, D.H. (2014). The small GTPase Rheb is required for spermatogenesis but not oogenesis. Reproduction *147*, 615-625.

Bao, J.Q., Tang, C., Yuan, S.Q., Porse, B.T., and Yan, W. (2015). UPF2, a nonsense-mediated mRNA decay factor, is required for prepubertal Sertoli cell development and male fertility by ensuring fidelity of the transcriptome. Development *142*, 352-362.

Bebee, T.W., Dominguez, C.E., and Chandler, D.S. (2012). Mouse models of SMA: tools for disease characterization and therapeutic development. Human Genetics *131*, 1277-1293.

Bellve, A.R., Cavicchia, J.C., Millette, C.F., Obrien, D.A., Bhatnagar, Y.M., and Dym, M. (1977). SPERMATOGENIC CELLS OF PREPUBERAL MOUSE - ISOLATION AND MORPHOLOGICAL CHARACTERIZATION. Journal of Cell Biology *74*, 68-85.

Berkovits, B.D., Wang, L., Guarnieri, P., and Wolgemuth, D.J. (2012). The testis-specific double bromodomain-containing protein BRDT forms a complex with multiple spliceosome components and is required for mRNA splicing and 3 '-UTR truncation in round spermatids. Nucleic Acids Research *40*, 7162-7175.

Boyer, J.G., Ferrier, A., and Kothary, R. (2013). More than a bystander: the contributions of intrinsic skeletal muscle defects in motor neuron diseases. Frontiers in physiology *4*, 356.

Branchu, J., Biondi, O., Chali, F., Collin, T., Leroy, F., Mamchaoui, K., Makoukji, J., Pariset, C., Lopes, P., Massaad, C.*, et al.* (2013). Shift from Extracellular Signal-Regulated Kinase to AKT/cAMP Response Element-Binding Protein Pathway Increases Survival-Motor-Neuron Expression in Spinal-Muscular-Atrophy-Like Mice and Patient Cells. Journal of Neuroscience *33*, 4280-+.

Chen, H.H., Chang, J.G., Lu, R.M., Peng, T.Y., and Tarn, W.Y. (2008). The RNA Binding Protein hnRNP Q Modulates the Utilization of Exon 7 in the Survival Motor Neuron 2 (SMN2) Gene. Molecular and Cellular Biology *28*, 6929-6938.

Chen, Y.C., Chang, J.G., Jong, Y.J., Liu, T.Y., and Yuo, C.Y. (2015). High Expression Level of

Tra2-beta 1 Is Responsible for Increased SMN2 Exon 7 Inclusion in the Testis of SMA Mice. Plos One *10*, 14.

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

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. Molecular Cell *33*, 717-726.

Comhaire, F.H., Dekretser, D., Farley, T.M.M., and Rowe, P.J. (1987). TOWARDS MORE OBJECTIVITY IN DIAGNOSIS AND MANAGEMENT OF MALE-INFERTILITY. International Journal of Andrology, R3-53.

Daigle Jr, H., Cole, D., Carlson, A., Lee, W., and Wilson, V. (2009). Ethylene dichloride disruption of fertility in male mice. The Open Toxicology Journal *3*, 39-46.

Furukawa, T., Nakao, K., Sugita, H., and Tsukagos.H (1968). KUGELBERG-WELANDER DISEASE - WITH PARTICULAR REFERENCE TO SEX-INFLUENCED MANIFESTATIONS. Archives of Neurology *19*, 156-&.

Ginter-Matuszewska, B., Kusz, K., Spik, A., Grzeszkowiak, D., Rembiszewska, A., Kupryjanczyk, J., and Jaruzelska, J. (2011). NANOS1 and PUMILIO2 bind microRNA biogenesis factor GEMIN3, within chromatoid body in human germ cells. Histochemistry and Cell Biology *136*, 279-287.

Gou, L.T., Dai, P., and Liu, M.F. (2014). Small noncoding RNAs and male infertility. Wiley Interdisciplinary Reviews-Rna *5*, 733-745.

Grice, S.J., and Liu, J.L. (2011). Survival Motor Neuron Protein Regulates Stem Cell Division, Proliferation, and Differentiation in Drosophila. Plos Genetics *7*, 12.

Hausmanowa-Petrusewicz, I., Zaremba, J., and Borkowska, J. (1979). Chronic form of childhood spinal muscular atrophy. Are the problems of its genetics really solved? Journal of the neurological sciences *43*, 317-327.

Heier, C.R., Satta, R., Lutz, C., and DiDonato, C.J. (2010). Arrhythmia and cardiac defects are a feature of spinal muscular atrophy model mice. Human Molecular Genetics *19*, 3906-3918.

Huynh, T., Mollard, R., and Trounson, A. (2002). Selected genetic factors associated with male infertility. Human Reproduction Update *8*, 183-198.

Hwang, K., Yatsenko, A.N., Jorgez, C.J., Mukherjee, S., Nalam, R.L., Matzuk, M.M., and Lamb, D.J. (2010). Mendelian genetics of male infertility. Annals of the New York Academy of Sciences *1214*, E1-E17.

Idler, R.K., and Yan, W. (2012). Control of Messenger RNA Fate by RNA-Binding Proteins: An Emphasis on Mammalian Spermatogenesis. Journal of Andrology *33*, 309-337.

Iwahashi, H., Eguchi, Y., Yasuhara, N., Hanafusa, T., Matsuzawa, Y., and Tsujimoto, Y. (1997). Synergistic anti-apoptotic activity between Bcl-2 and SMN implicated in spinal muscular atrophy. Nature *390*, 413-417.

Jedrzejowska, M., Milewski, M., Zimowski, J., Borkowska, J., Kostera-Pruszczyk, A., Sielska, D., Jurek, M., and Hausmanowa-Petrusewicz, I. (2009). Phenotype modifiers of spinal muscular atrophy: the number of SMN2 gene copies, deletion in the NAIP gene and probably gender influence the course of the disease. Acta Biochimica Polonica *56*, 103-108.

Kerr, D.A., Nery, J.P., Traystman, R.J., Chau, B.N., and Hardwick, J.M. (2000). Survival motor neuron protein modulates neuron-specific apoptosis. Proceedings of the National Academy of Sciences of the United States of America *97*, 13312-13317.

Kosova, G., Scott, N.M., Niederberger, C., Prins, G.S., and Ober, C. (2012). Genome-wide Association Study Identifies Candidate Genes for Male Fertility Traits in Humans. American Journal of Human Genetics *90*, 950-961.

Kreidberg, J.A., Sariola, H., Loring, J.M., Maeda, M., Pelletier, J., Housman, D., and Jaenisch, R. (1993). WT-1 IS REQUIRED FOR EARLY KIDNEY DEVELOPMENT. Cell *74*, 679-691.

Lee, J.W., Richburg, J.H., Younkin, S.C., and Boekelheide, K. (1997). The Fas system is a key regulator of germ cell apoptosis in the testis. Endocrinology *138*, 2081-2088.

Li, J.Y., Lees-Murdock, D.J., Xu, G.L., and Walsh, C.P. (2004). Timing of establishment of paternal methylation imprints in the mouse. Genomics *84*, 952-960.

Li, N., Mruk, D.D., Wong, C.K.C., Lee, W.M., Han, D.S., and Cheng, C.Y. (2015). Actin-bundling protein plastin 3 is a regulator of ectoplasmic specialization dynamics during spermatogenesis in the rat testis. Faseb Journal *29*, 3788-3805.

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

Margolin, G., Khil, P.P., Kim, J., Bellani, M.A., and Camerini-Otero, R.D. (2014). Integrated transcriptome analysis of mouse spermatogenesis. Bmc Genomics *15*, 19.

Meister, G., and Fischer, U. (2002). Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. Embo Journal *21*, 5853-5863.

Ni, B.X., Ma, H.X., Lin, Y., Dai, J.C., Guo, X.J., Xia, Y.K., Sha, J.H., and Hu, Z.B. (2014). Genetic variants in Ser-Arg protein-coding genes are associated with the risk of nonobstructive azoospermia in Chinese men. Fertility and Sterility *101*, 1711-U1573.

Ogawa, C., Usui, K., Aoki, M., Ito, F., Itoh, M., Kai, C., Kanamori-Katayama, M., Hayashizaki, Y., and Suzuki, H. (2007). Gemin2 plays an important role in stabilizing the survival of motor neuron complex. Journal of Biological Chemistry *282*, 11122-11134.

Oprea, G.E., Kroeber, S., McWhorter, M.L., Rossoll, W., Mueller, S., Krawczak, M., Bassell,

G.J., Beattie, C.E., and Wirth, B. (2008). Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy. Science *320*, 524-527.

Orth, J.M., Gunsalus, G.L., and Lamperti, A.A. (1988). EVIDENCE FROM SERTOLI CELL-DEPLETED RATS INDICATES THAT SPERMATID NUMBER IN ADULTS DEPENDS ON NUMBERS OF SERTOLI CELLS PRODUCED DURING PERINATAL-DEVELOPMENT. Endocrinology *122*, 787-794.

Osborne, M., Gomez, D., Feng, Z.H., McEwen, C., Beltran, J., Cirillo, K., El-Khodor, B., Lin, M.Y., Li, Y., Knowlton, W.M.*, et al.* (2012). Characterization of behavioral and neuromuscular junction phenotypes in a novel allelic series of SMA mouse models. Human Molecular Genetics *21*, 4431-4447.

Parker, G.C., Li, X.L., Anguelov, R.A., Toth, G., Cristescu, A., and Acsadi, G. (2008). Survival motor neuron protein regulates apoptosis in an in vitro model of spinal muscular atrophy. Neurotoxicity Research *13*, 39-48.

Perrin, S. (2014). Make mouse studies work. Nature *507*, 423-425.

Richert, J.R., Antel, J.P., Canary, J.J., Maxted, W.C., and Groothuis, D. (1986). ADULT ONSET SPINAL MUSCULAR-ATROPHY WITH ATROPHIC TESTES - REPORT OF 2 CASES. Journal of Neurology Neurosurgery and Psychiatry *49*, 606-608.

Rouget, R.L., Vigneault, F., Codio, C., Rochette, C., Paradis, I., Drouin, G., and Simard, L.R. (2005). Characterization of the survival motor neuron (SMN) promoter provides evidence for complex combinatorial regulation in undifferentiated and differentiated P19 cells. Biochemical Journal *385*, 433-443.

Sato, K., Eguchi, Y., Kodama, T.S., and Tsujimoto, Y. (2000). Regions essential for the interaction between Bcl-2 and SMN, the spinal muscular atrophy disease gene product. Cell Death and Differentiation *7*, 374-383.

Schmid, R., Grellscheid, S.N., Ehrmann, I., Dalgliesh, C., Danilenko, M., Paronetto, M.P., Pedrotti, S., Grellscheid, D., Dixon, R.J., Sette, C.*, et al.* (2013). The splicing landscape is globally reprogrammed during male meiosis. Nucleic Acids Research *41*, 10170-10184.

Schmitt-John, T., Drepper, C., Mussmann, A., Hahn, P., Kuhlmann, M., Thiel, C., Hafner, M., Lengeling, A., Heimann, P., Jones, J.M.*, et al.* (2005). Mutation of Vps54 causes motor neuron disease and defective spermiogenesis in the wobbler mouse. Nature Genetics *37*, 1213-1215.

Schreml, J., Riessland, M., Paterno, M., Garbes, L., Rossbach, K., Ackermann, B., Kramer, J., Somers, E., Parson, S.H., Heller, R.*, et al.* (2013). Severe SMA mice show organ impairment that cannot be rescued by therapy with the HDACi JNJ-26481585. European Journal of Human Genetics *21*, 643-652.

Seo, J., Howell, M.D., Singh, N.N., and Singh, R.N. (2013). Spinal muscular atrophy: An update on therapeutic progress. Biochimica Et Biophysica Acta-Molecular Basis of Disease *1832*, 2180-2190.

Shen, S.H., Park, J.W., Huang, J., Dittmar, K.A., Lu, Z.X., Zhou, Q., Carstens, R.P., and Xing, Y. (2012). MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucleic Acids Research *40*, 13.

Shukla, K.K., Mahdi, A.A., and Rajender, S. (2012). Apoptosis, spermatogenesis and male infertility. Frontiers in bioscience (Elite edition) *4*, 746-754.

Simic, G., Seso-Simic, D., Lucassen, P.J., Islam, A., Krsnik, Z., Cviko, A., Jelasic, D., Barisic, N., Winblad, B., Kostovic, I.*, et al.* (2000). Ultrastructural analysis and TUNEL demonstrate motor neuron apoptosis in Werdnig-Hoffmann disease. Journal of Neuropathology and Experimental Neurology *59*, 398-407.

Singh, N.N., Seo, J., Rahn, S.J., and Singh, R.N. (2012). A Multi-Exon-Skipping Detection Assay Reveals Surprising Diversity of Splice Isoforms of Spinal Muscular Atrophy Genes. Plos One *7*, 17.

Song, H.W., and Wilkinson, M.F. (2014). Transcriptional control of spermatogonial maintenance and differentiation. Seminars in Cell & Developmental Biology *30*, 14-26.

Stathas, D., Kalfakis, N., Kararizou, E., and Manta, P. (2008). Spinal muscular atrophy: DNA fragmentation and immaturity of muscle fibers. Acta Histochemica *110*, 53-58.

Sun, J., and Wu, J. (2015). Expression profiling of long noncoding RNAs in neonatal and adult mouse testis. Data in brief *4*, 322-327.

Takaku, M., Tsujita, T., Horikoshi, N., Takizawa, Y., Qing, Y., Hirota, K., Ikura, M., Ikura, T., Takeda, S., and Kurumizaka, H. (2011). Purification of the Human SMN-GEMIN2 Complex and Assessment of Its Stimulation of RAD51-Mediated DNA Recombination Reactions. Biochemistry *50*, 6797-6805.

Thomas, K., Wu, J., Sung, D.Y., Thompson, W., Powell, M., McCarrey, J., Gibbs, R., and Walker, W. (2007). SP1 transcription factors in male germ cell development and differentiation. Molecular and Cellular Endocrinology *270*, 1-7.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A.*, et al.* (2010). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. Molecular Cell *39*, 925-938.

Tripsianes, K., Madl, T., Machyna, M., Fessas, D., Englbrecht, C., Fischer, U., Neugebauer, K.M., and Sattler, M. (2011). Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. Nature Structural & Molecular Biology *18*, 1414-U1136.

Vyas, S., Bechade, C., Riveau, B., Downward, J., and Triller, A. (2002). Involvement of survival

motor neuron (SMN) protein in cell death. Human Molecular Genetics *11*, 2751-2764.

Wang, Y.B., Zhu, T.X., Li, Q.L., Liu, C.Y., Han, F., Chen, M., Zhang, L.J., Cui, X.H., Qin, Y., Bao, S.L.*, et al.* (2015). Prmt5 is required for germ cell survival during spermatogenesis in mice. Scientific Reports *5*, 10.

Washietl, S., Kellis, M., and Garber, M. (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. Genome Research *24*, 616-628.

Young, P.J., DiDonato, C.J., Hu, D., Kothary, R., Androphy, E.J., and Lorson, C.L. (2002). SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1. Human Molecular Genetics *11*, 577-587.

Zerres, K., and Rudnik-Schoneborn, S. (2003). 93rd ENMC international workshop: non-5q-spinal muscular atrophies (SMA) - clinical picture (6-8 April 2001, Naarden, The Netherlands). Neuromuscular disorders : NMD *13*, 179-183.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. Cell *133*, 585-600.

Zhou, Y., Zhong, Y., Wang, Y., Zhang, X., Batista, D.L., Gejman, R., Ansell, P.J., Zhao, J., Weng, C., and Klibanski, A. (2007). Activation of p53 by MEG3 non-coding RNA. Journal of Biological Chemistry *282*, 24731-24742.

Acknowledgements

We thank Rachel Anderson for technical support. Authors acknowledge help of Biotechnology Center at Iowa State University (IA, USA) for help with RNA-seq experiments.

Funding:

This work was supported by grants from National Institutes of Health (NS055925 and NS080294) and Salsbury Endowment (Iowa State University, IA, USA) to RNS.

Author contributions:

E.W.O. and M.D.H. designed the research, performed experiments and wrote the manuscript. N.N.S., E.M.W. and J.S. designed the research and performed experiments. R.N.S. designed the experiments and wrote the manuscript.

Additional information:

Authors declare no competing financial interests.

Figure Legends

Figure 1: Male reproductive organs and functions are abnormal in P42 C/C mice. (a) Representative photograph of WT (left) and C/C (right) male reproductive organs, including testis, epididymis and epididymal fat pad. Scale bar represents 5 mm. (b) Gross testis mass. (c) Relative testis mass was determined by dividing gross testis mass by total body weight (n=10 WT and 6 C/C mice). (d) Serum testosterone level determined by ELISA (n=5 mice per genotype). (e) Representative H&E stained testis cross-sections. Circled numbers represent score of each tubule based upon a 10 point system that assesses morphology[13]. Damage and degeneration, including vacuolization (black arrowheads), multinucleated bodies (red arrowheads) and sloughed tissue (blue arrowheads) are indicated. Scale bars represent 50 μm. (f) Average seminiferous tubule scores for WT and C/C testes. Each point represents a single mouse (n=10 WT and 6 C/C mice). Pie charts below data points represent the distribution of all seminiferous tubule scores for WT and C/C males. The legend underneath indicates the color for each score. (g) Representative H&E stained cross sections of epididymal regions. The cartoon to the right indicates the location of each region. Scale bars represent 50 μm. (h) Total epididymal sperm count from P60 males (n=8 WT and 6 C/C mice). (Statistical significance: ***p<0.001)

Figure 2: P42 female C/C reproductive organs exhibit minimal abnormalities. (a) Representative photograph of WT (top) and C/C (bottom) uterus and ovaries. Scale bar represents 5 mm. (b) Gross uterus and ovaries mass. (c) Relative uterus

and ovaries mass was determined by dividing gross uterus and ovaries mass by total body weight (n=10 WT and 7 C/C mice). (d) Representative H&E stained cross-sections of ovary and the endometrium. Follicle (F) and corpus luteum (CL) are indicated on ovary micrographs. Scale bars represent 100 μm. (e-f) Fertility of heterozygous (n=11) and C/C (n=3) female mice. Females were paired in monogamous breeding cages with either a heterozygote or WT male. The average number of litters born over ninety days (e) and the average litter size for heterozygote or C/C mothers (f) were recorded. (Statistical significance *p<0.05)

Figure 3: High levels of SMN are expressed in C/C testis (a) Diagram of the *Smn*^{C} locus. Endogenous mouse *Smn* is interrupted 2,195 bp into intron 6 and replaced with the equivalent human *SMN2* sequence through the 3′- UTR, followed by 42 kb human *SMN2* genomic sequence. Exon lengths appear above colored boxes and intron lengths appear below lines. This transgene results in two genes, hybrid *Smn* and WT *SMN2*. MESDA[14] was used to determine all splice variants for *SMN2* (b) and *Smn* hybrid (c) in various tissues. Band identities are given to the left of the lanes, where Δ indicates the lack of specific exon(s). Tissue abbreviations are indicated in panel (d). Additional abbreviations: E, exon; In, intron; UTR, untranslated region; FL, full-length. (d) Western blots of SMN, Gemin2 and Actin (top three panels) expression in adult WT tissue, and accompanying silver stain to confirm approximately equal loading. Samples were prepared by dissecting tissues from 3 WT males and immediately grinding them together in liquid nitrogen. 5 μg of protein was loaded for silver staining and 7.5 μg was loaded for western blotting.

Tissue abbreviations appear to the right. (e-h) Representative SMN and Actin western blots from male (M) and female (F) brain (e), heart (f), liver (g), and testis and uterus/ovaries (U/O) (h). For (e-h), n=3 or 4 mice per genotype. (Statistical significance: *p<0.05)

Figure 4. RNA-Seq reveals drastic alterations in C/C testis transcriptome. (a) MA plots for testis (top), brain (middle) and liver (bottom). The y-axis depicts $\log_2$ fold change (LFC) in gene expression in C/C compared to WT, and the x-axis depicts the mean read count for each gene between all samples. Each dot represents one gene, with red dots representing genes with significantly altered expression values (Benjamini and Hochberg (B+H) adjusted p value < 0.05). (b) KEGG pathways enriched for genes with altered expression levels in C/C testis. The length of each bar represents the B+H adjusted p value of enrichment for each pathway. The number at the top of each bar represents the number of genes participating in each pathway with significantly altered expression. (c) Percentage of genes upregulated (white bars) or downregulated (black bars) in C/C testes that are classified as being primarily expressed in certain testicular cell types. Abbreviations used: spermatogonia (SG), spermatocytes (SC). (d) LFC in C/C testes of several candidate genes as measured by RNA-Seq (y values) and QPCR (x values). (e-f) QPCR showing aberrant testis expression of genes that are strongly upregulated (e) or downregulated (f). (g-i) QPCR for select genes involved in steroid biosynthesis (g), spermatogenesis (h) and axon guidance (i). (j) QPCR for expression of lncRNAs. (k)

QPCR for expression of genes that encode members of the SMN complex. For (e-k), n=8 mice per genotype (Statistical significance: *p<0.05; **p<0.01; ***p<0.001)

Figure 5. Expression of apoptosis-related genes is disproportionally affected in C/C testes. (a) Heat map depicting the relative expression of every gene involved in apoptosis, as annotated by the Kyoto Encyclopedia of Genes and Genomes. Each colored bar indicates the LFC of expression in C/C compared to WT. The color scale is indicated below, with red colored bars indicating genes with lower expression in C/C and green colored bars indicating genes with increased expression in C/C. (b-d) Relative expression of several candidate genes involved in apoptosis, as measured by QPCR in testes (b), brain (c), and liver (d); n=4 mice per genotype. (e) Representative micrographs from TUNEL assay showing nuclei (DAPI, blue), TUNEL signal (green) and merged image. Individual seminiferous tubules are noted with dashed white outline. Scale bars represent 50 μm. (f) Percentage of seminiferous tubules with at least one TUNEL-positive cell. (g) Average number of TUNEL positive cells per seminiferous tubules. (h) Percentage of TUNEL-positive cells classified by cell type, including Sertoli cells, spermatogonia (SG), spermatocytes (SC) and round spermatids (RS). For (f-h), n=4 mice per genotype. (i) Representative images of cleaved caspse 3 immunostaining. C/C testes exhibited cleaved caspase 3 positive spermatogonia (panel I), spermatocytes (panel II) and multinucleated body (panel III) indicated by arrowheads. WT testes showed absence of cleaved caspase 3 (panel IV). Scale bars represent 50 μm. (Statistical significance: *p<0.05; **p<0.01; ***p<0.001)

Figure 6. *SMN2* and hybrid *Smn* mRNA splicing changes during testis development. (a) Timeline of the ages and estimated WT cell type distribution of seminiferous tubules of mice used for this study. The height of each designated region represents the estimated proportion of total number of cells in an average seminiferous tubule. (b,c) MESDA for *SMN2* (b) and hybrid *Smn* (c); n=2 mice per genotype. Isoform identities are given on the right side of each gel image where Δ indicates a skipped exon. Each lane is a representative example taken from two replicates. Lower panels show results of densitometric quantification of the most prominent isoforms of each gene. (d) Western blots for Tra2β, hnRNP Q, hnRNP A1, hnRNP A2/B1, and Actin during testis development. Equal amounts of protein from each of 3-4 C/C samples were pooled together and 15 μg of protein loaded for each time point. Densitometric quantifications are given on the right. Abbreviation: SZ, spermatozoa

Figure 7. SMN mRNA and protein expression during testis development. (a-b) QPCR measurement of relative quantities of *Smn* transcript in WT testes (a) and FL (including both exons 5 and 7) *SMN2* transcript in C/C testes (b) during development. Expression is calculated relative to the average expression level at P7. (c-d) Time course of SMN protein expression in WT (c) and C/C (d) mice. Blots shown are representative examples of each time point. (e) SMN, Gemin2 and Actin expression in WT and C/C testes during development for each age. (f) Densitometric analysis for SMN western blots shown in (e). (g) Densitometric analysis for Gemin2

western blots shown in (e). N=2-4 mice per genotype and age. (Statistical significance: *p<0.05; **p<0.01; ***p<0.001)

Figure 8. Alternative splicing and gene expression in C/C testes during spermatogenesis. Splicing patterns for *Add3* exon 14 (a), *Lrrc16a* exon 38 (b), *Picalm* exon 13 (c), *Wt1* exon 5 (d) and *Sulf1* exon 21 (e) during testis development. The top panels are representative autoradiograms of semi-quantitative PCR. Isoforms are labeled to the right of each gel image. The bottom panels portray autoradiogram quantification. For (a-e), n=2 mice per genotype. (f-n) QPCR expression of candidate genes during testis development, including *Apoe* (f), *Cpe* (g), *Ddr1* (h), *Lamb2* (i), *SerpinG1* (j), *Npy* (k), *Oaz3* (l), *Ppp2r2b* (m) and *Tppp2* (n). *Npy* (k) is charted on a $\log_{10}$ scale due to large expression changes. For (f-n), n=3-4 mice per genotype. (Statistical significance: *p<0.05; **p<0.01; ***p<0.001)

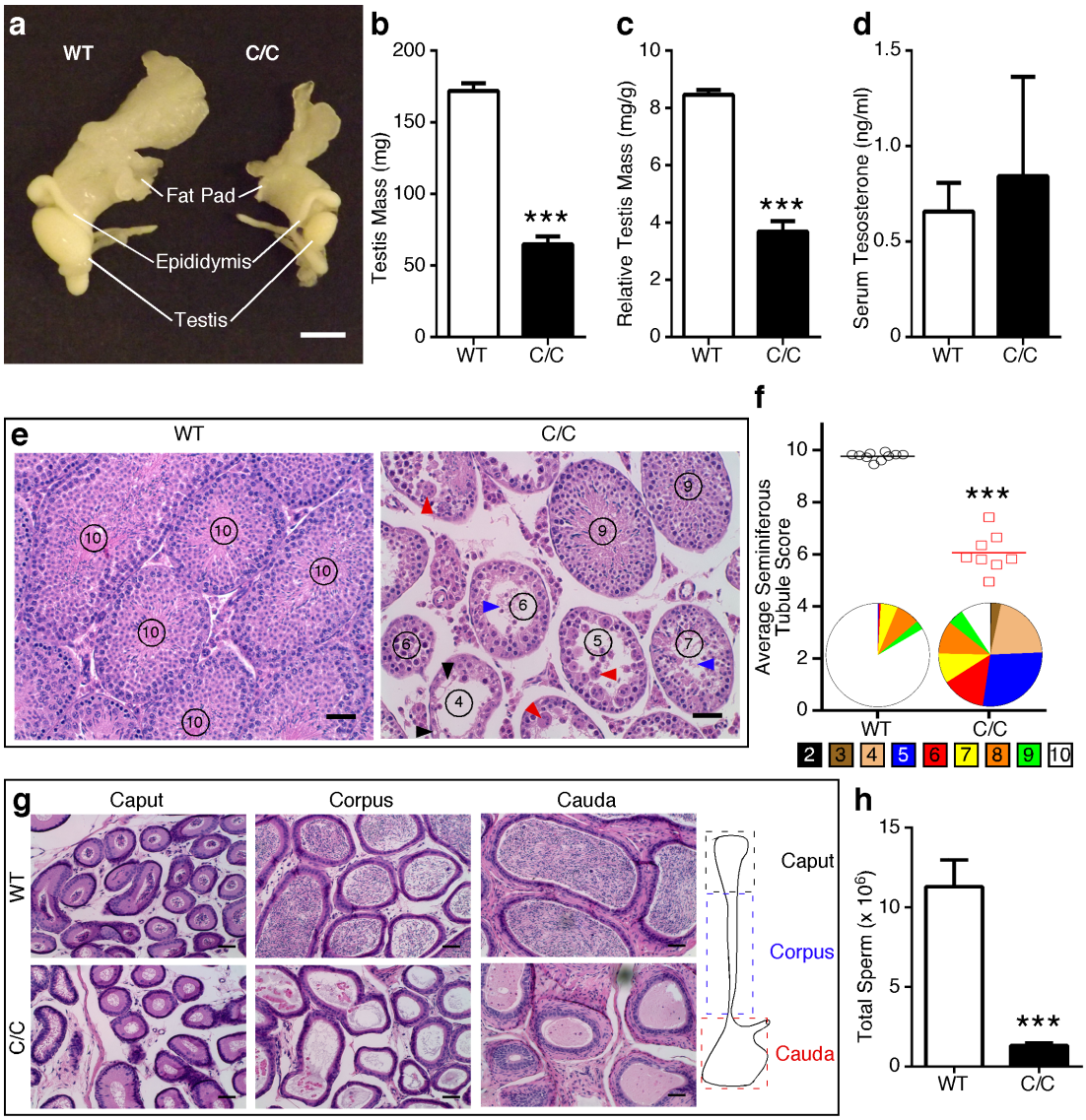


Figure 1. Male reproductive organs and functions are abnormal in P42 C/C mice.

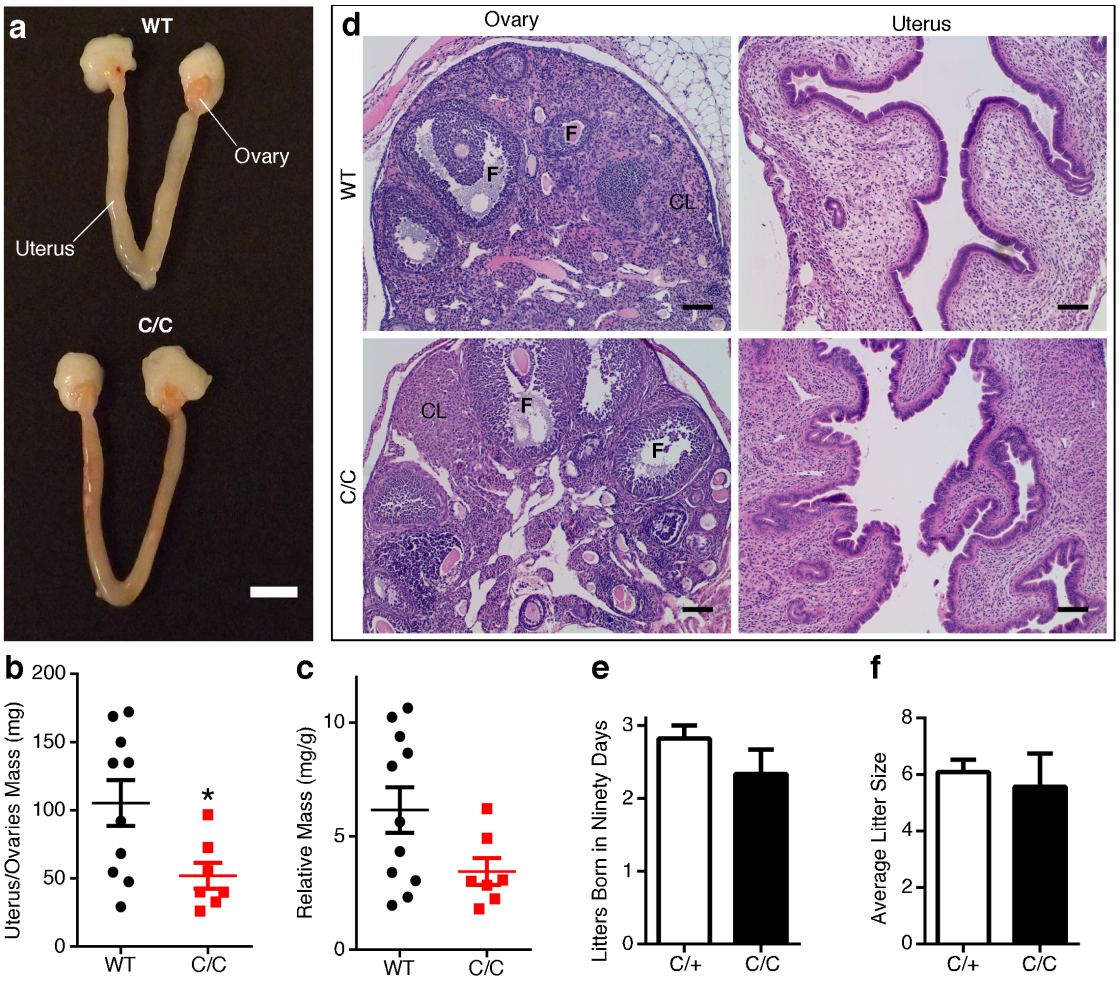


Figure 2. P42 female C/C reproductive organs exhibit minimal abnormalities

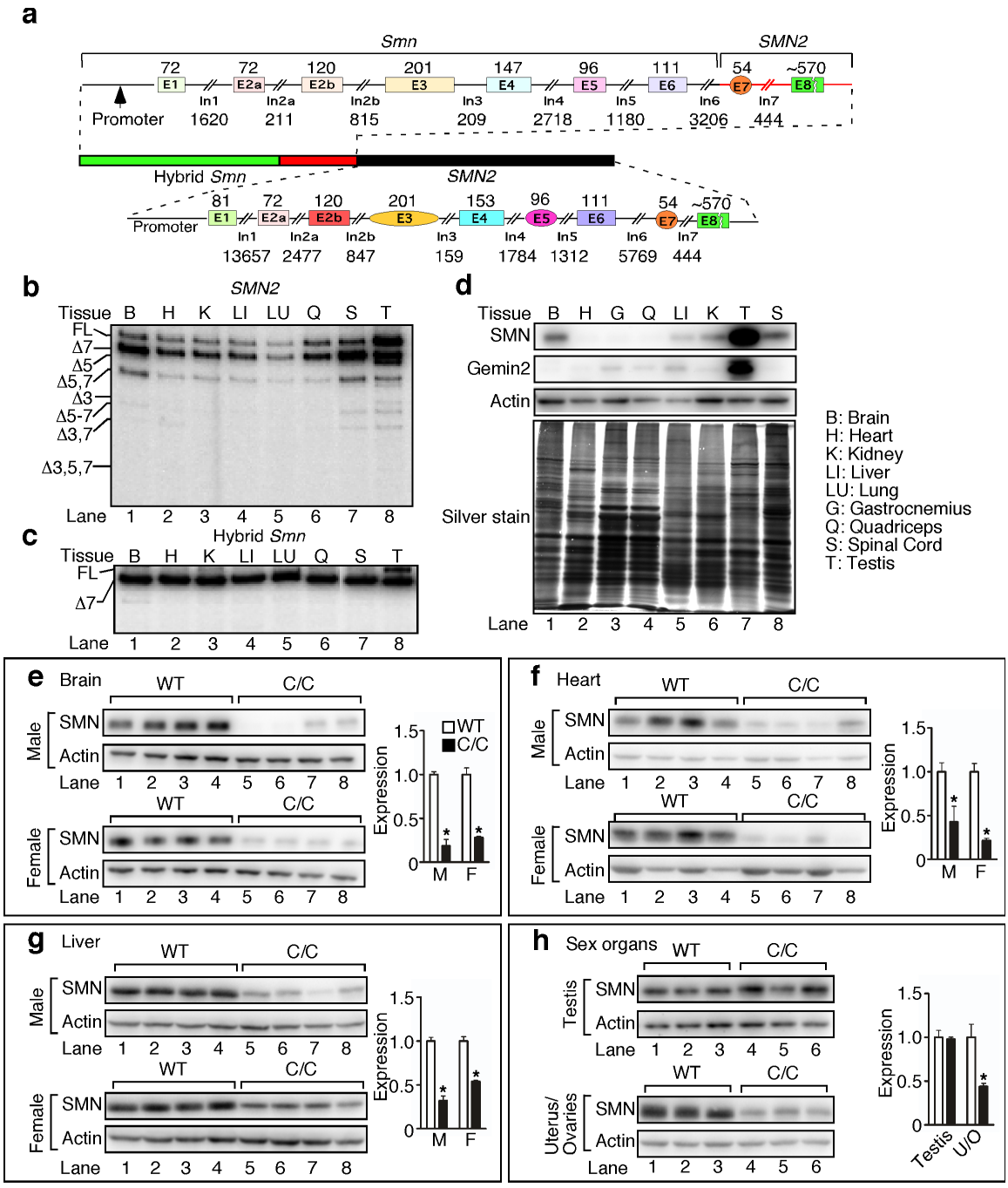


Figure 3. High levels of SMN are expressed in C/C testis

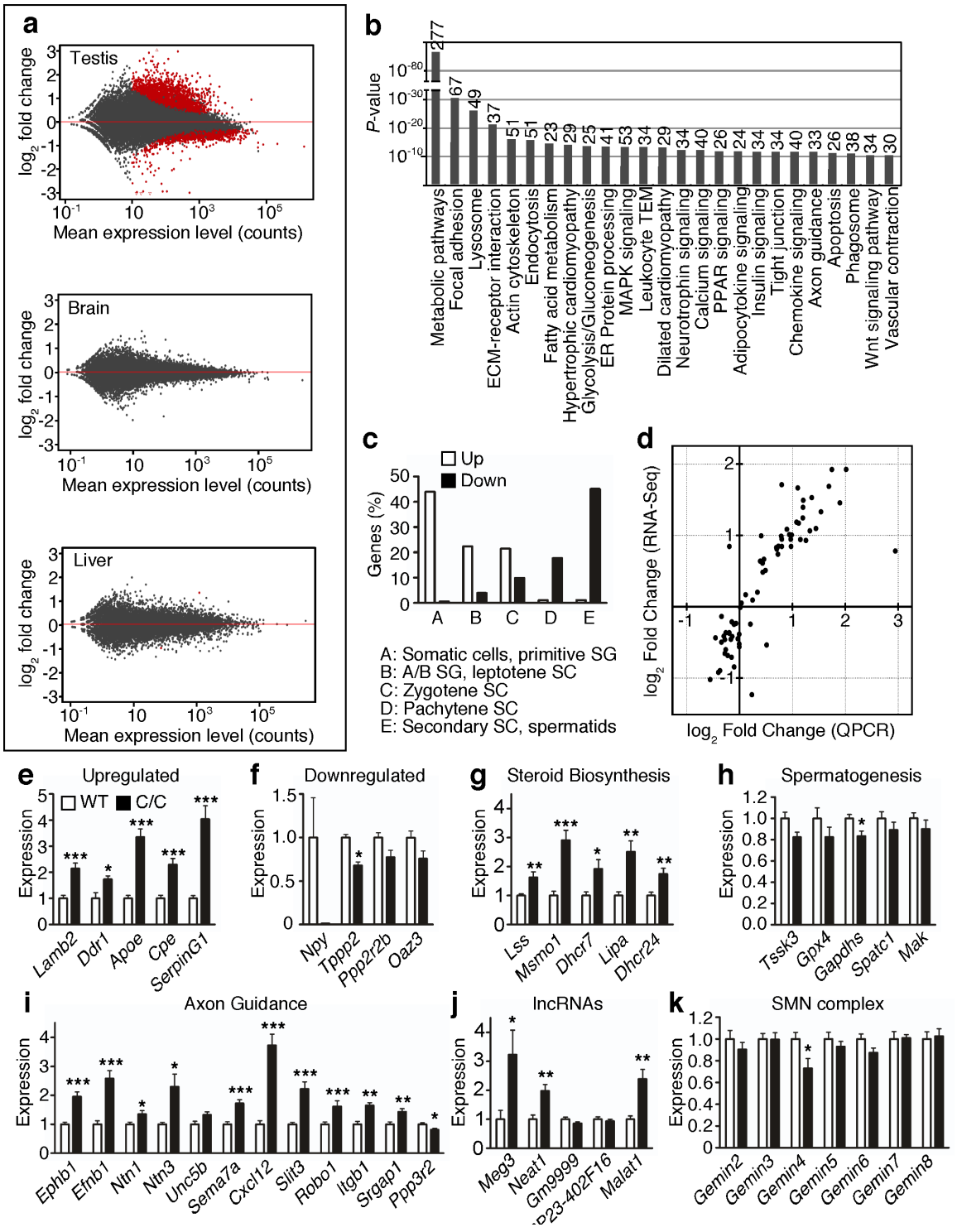


Figure 4. RNA-Seq reveals drastic alterations in C/C testis transcriptome

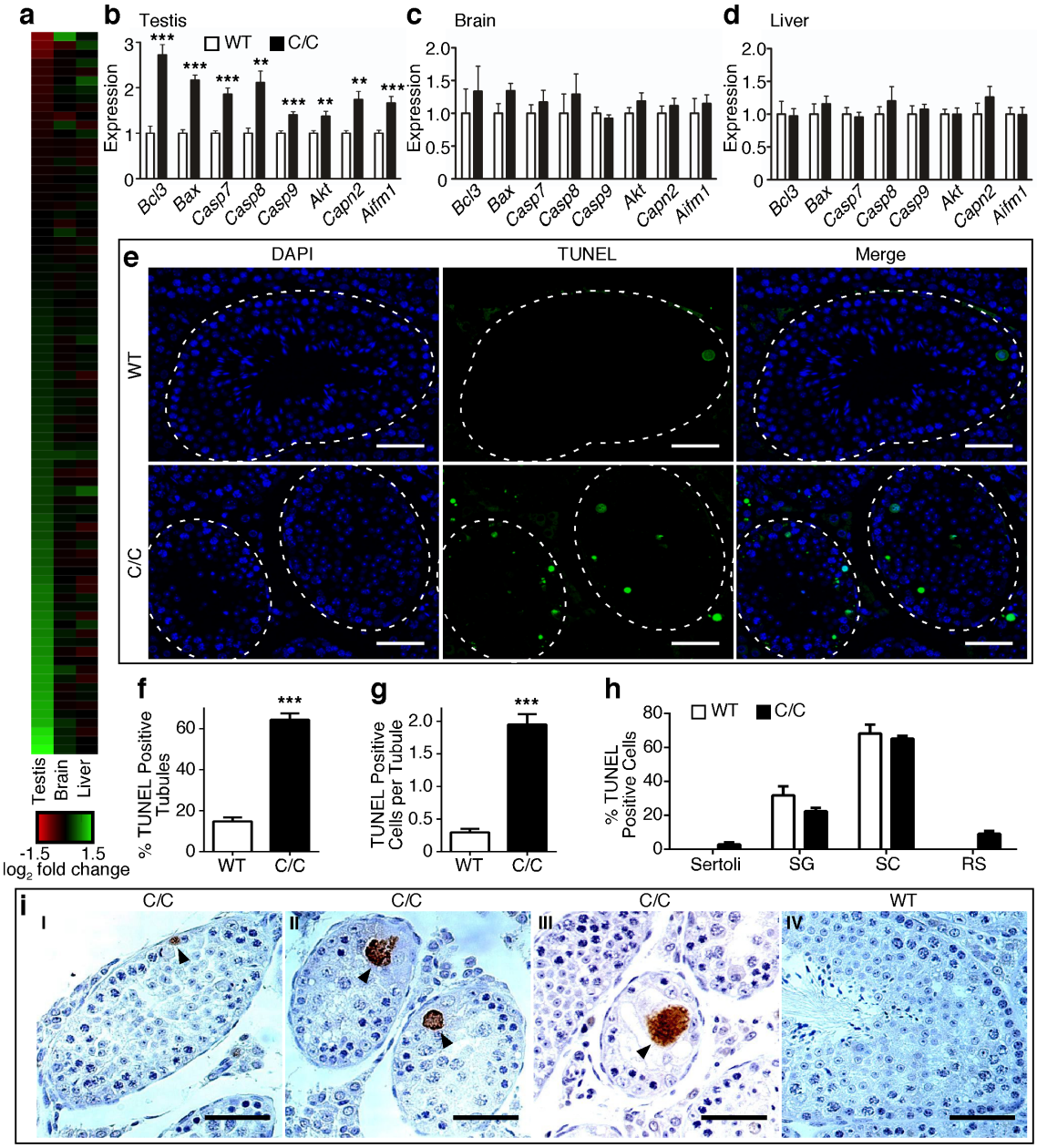


Figure 5. Expression of apoptosis-related genes is disproportionally affected in C/C testes

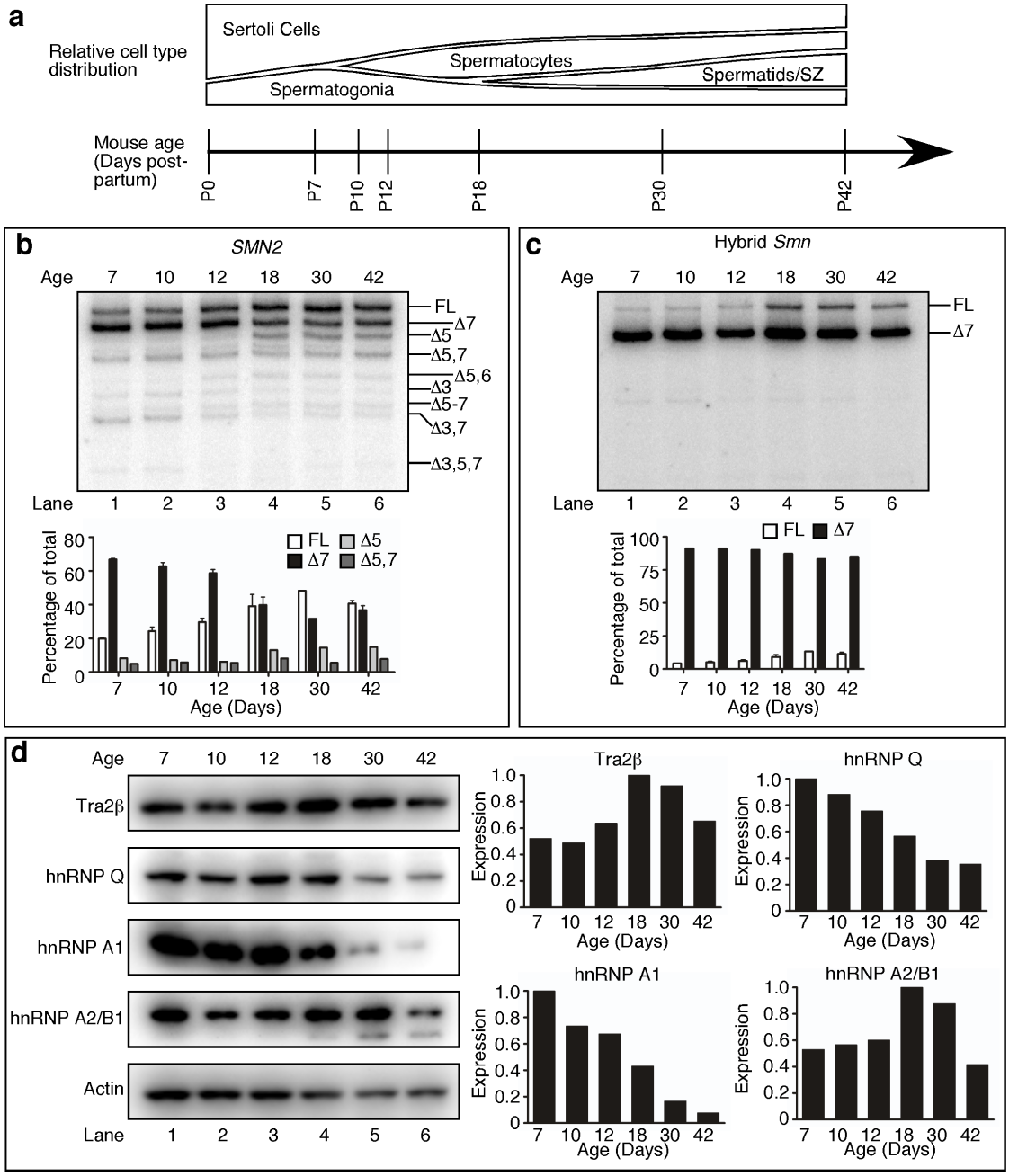


Figure 6. *SMN2* and hybrid *Smn* mRNA splicing changes during testis development

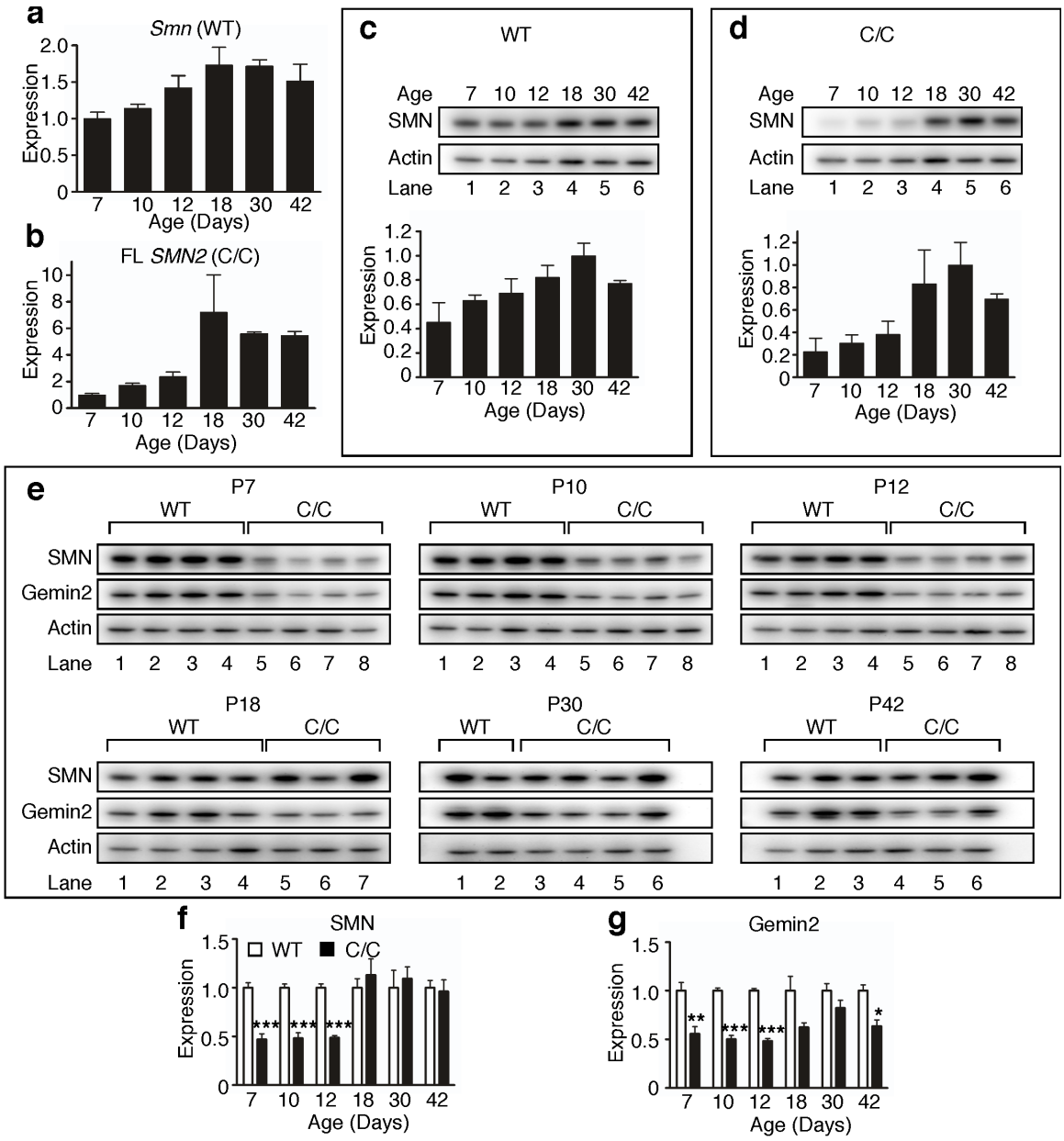


Figure 7. SMN mRNA and protein expression during testis development

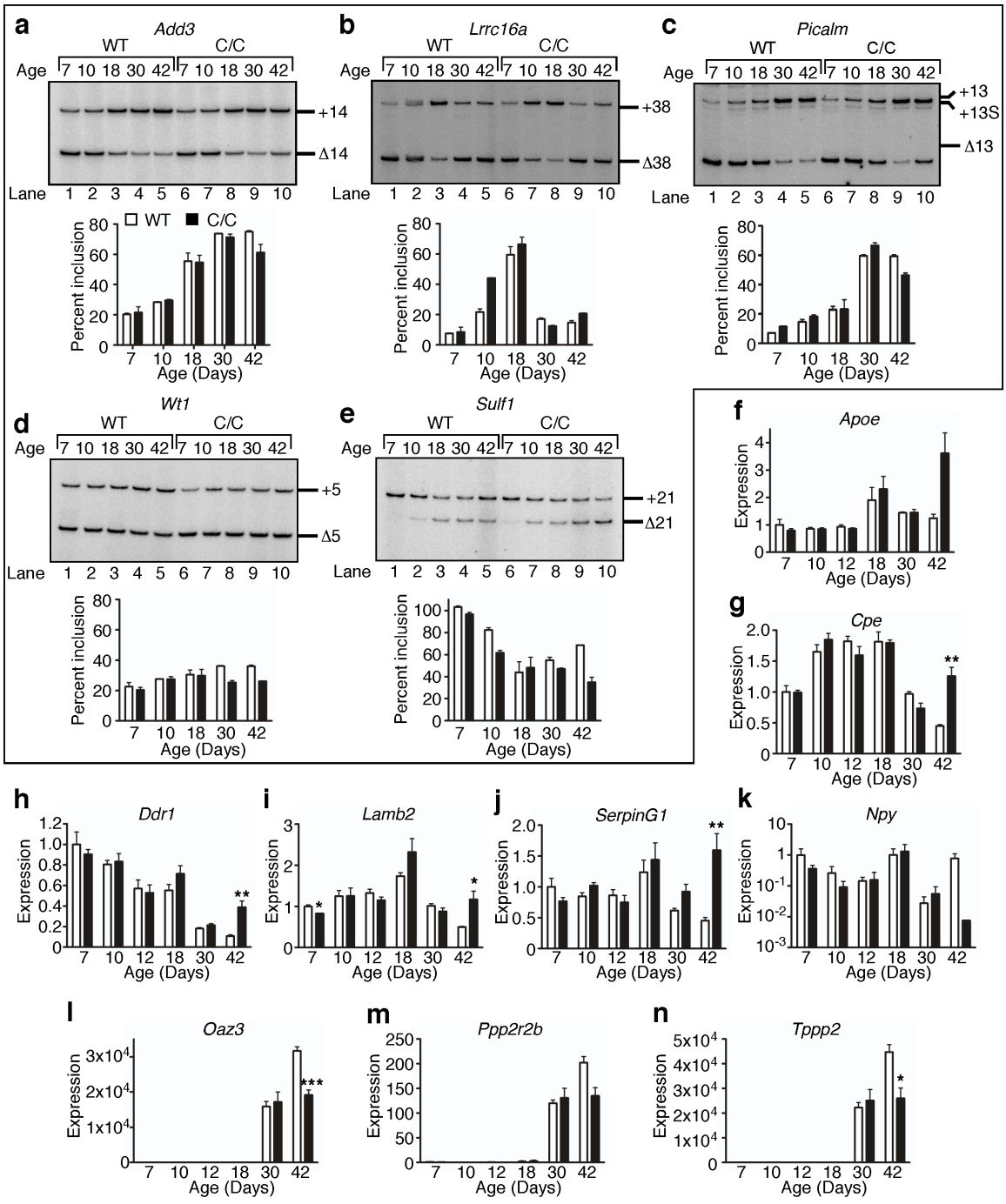


Figure 8. Alternative splicing and gene expression in C/C testes during spermatogenesis

APPENDIX A

CHAPTER II SUPPLEMENTARY FIGURES AND TABLES

Supplementary Figure Legends

Supplementary Figure 1. GO terms enriched in all significantly altered genes. The length of each bar represents the B+H adjusted *p* value of enrichment for each pathway. GO terms are characterized as representing biological processes (a), molecular functions (b), or cellular components (c)

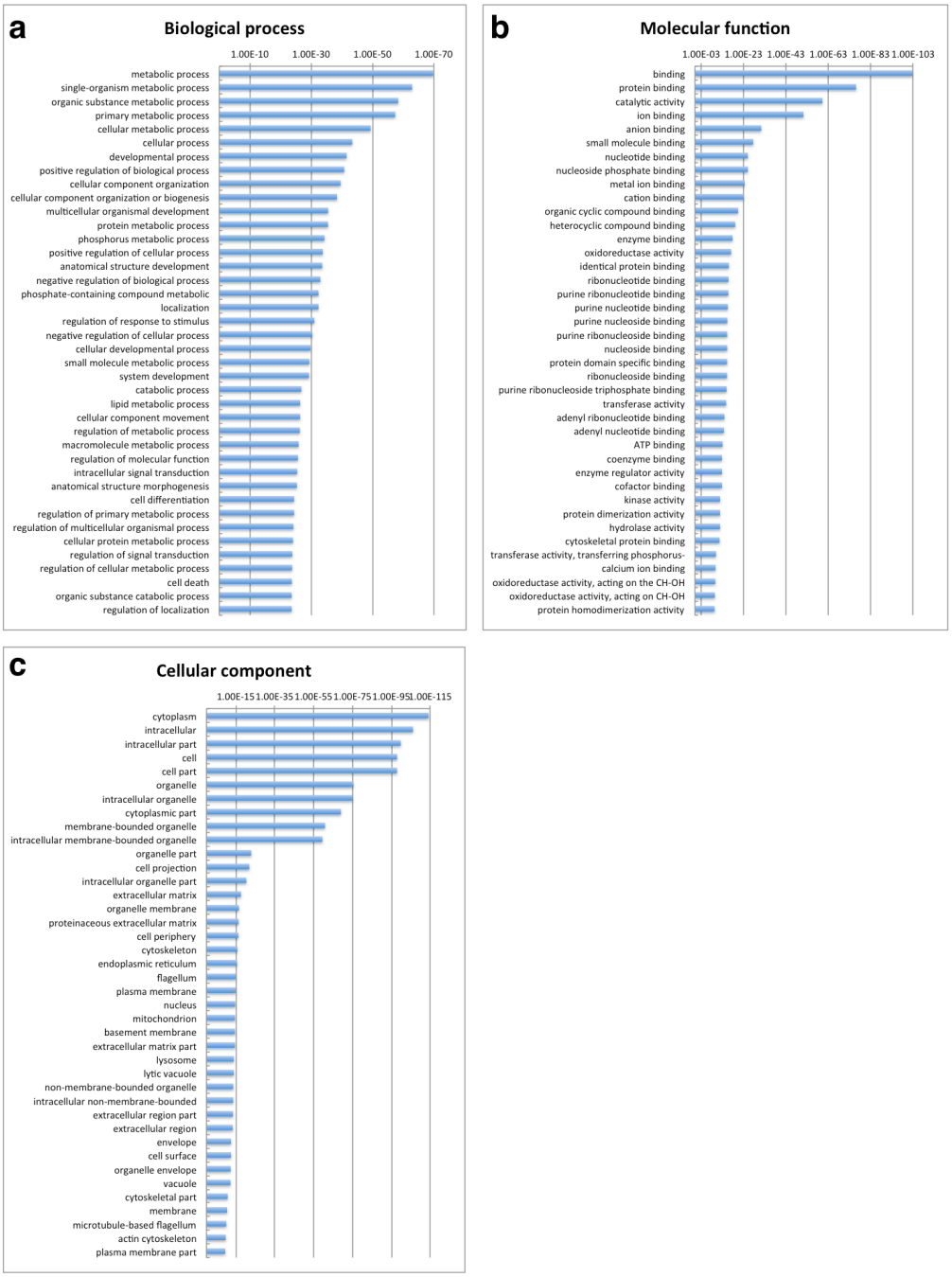
Supplementary Figure 2. GO terms enriched in significantly upregulated genes. The length of each bar represents the B+H adjusted *p* value of enrichment for each pathway. GO terms are characterized as representing biological processes (a), molecular functions (b), or cellular components (c)

Supplementary Figure 3. GO terms enriched in significantly downregulated genes. The length of each bar represents the B+H adjusted *p* value of enrichment for each pathway. GO terms are characterized as representing biological processes (a) or cellular components (b). No GO terms corresponding to specific molecular functions were significantly enriched.

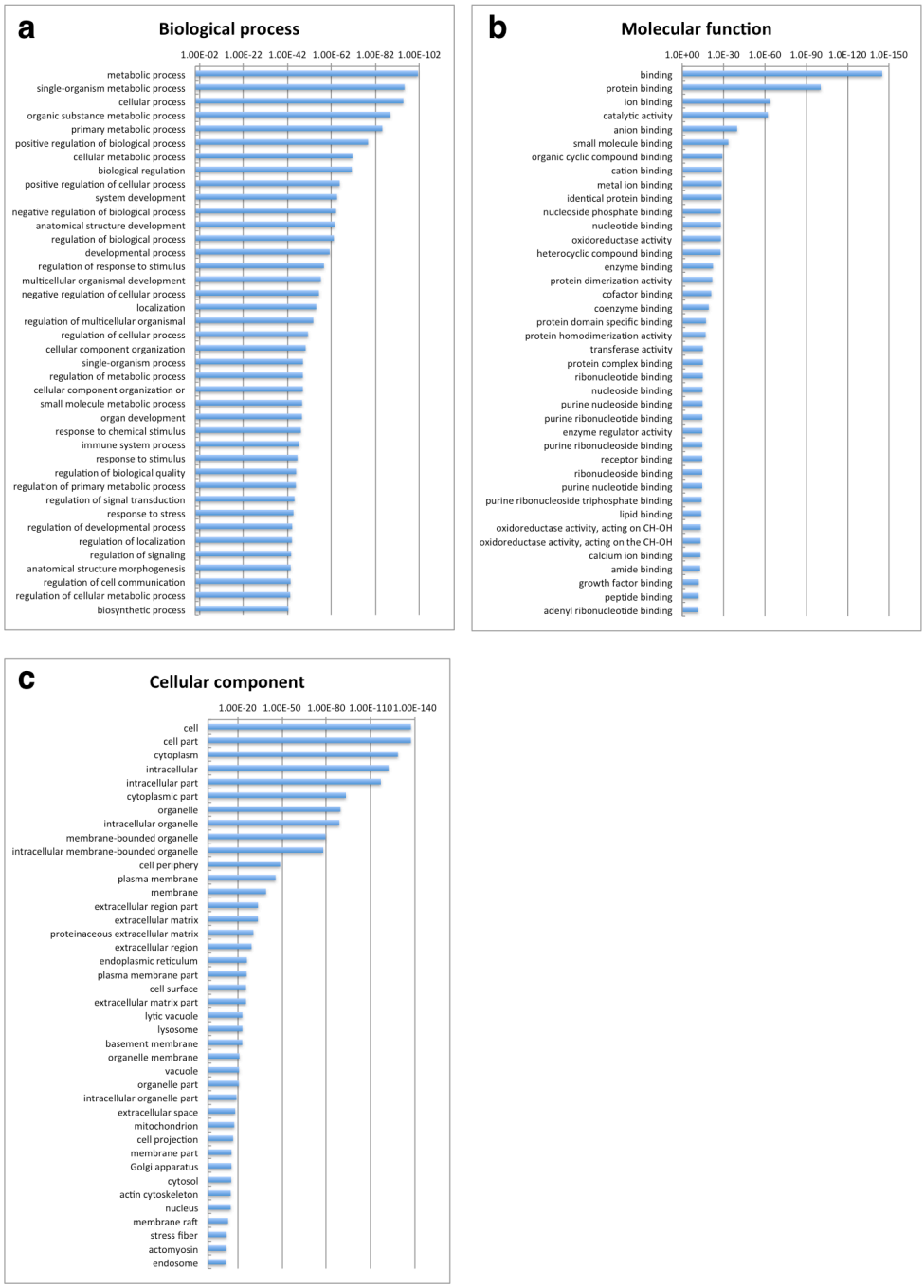
Supplementary Figure 4. Expression of markers of particular cell types. mRNA expression level of cell-type markers in C/C testes as measured by RNA-Seq. Value is expressed as expression relative to WT. (Statistical significance: *$p<0.5$; **$p<0.01$; ***$p<0.001$)

Supplementary Figure 5. Expression of genes mutated in oligospermia. mRNA expression level in allele C testes of genes whose mutation causes oligospermia, as

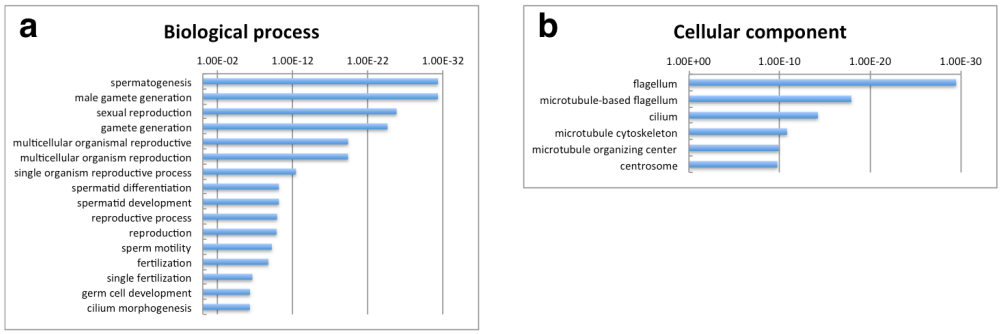
measured by RNA-Seq. Value is expressed as expression relative to WT. (Statistical significance: *p<0.5; **p<0.01; ***p<0.001)



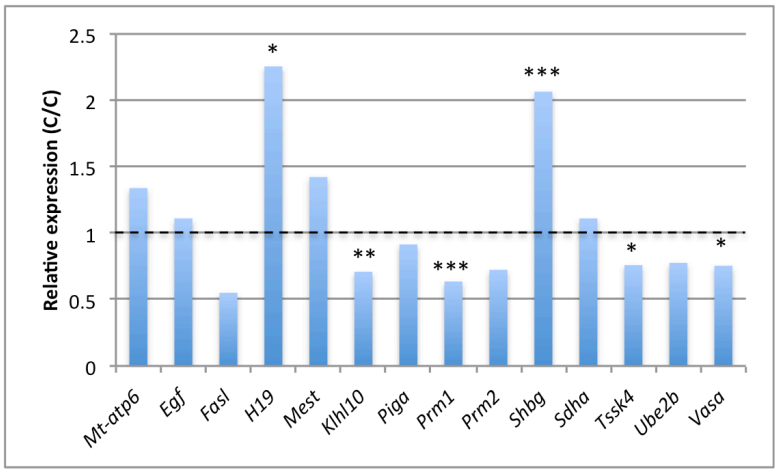
Supplementary Figure 1. GO terms enriched in all significantly altered genes.



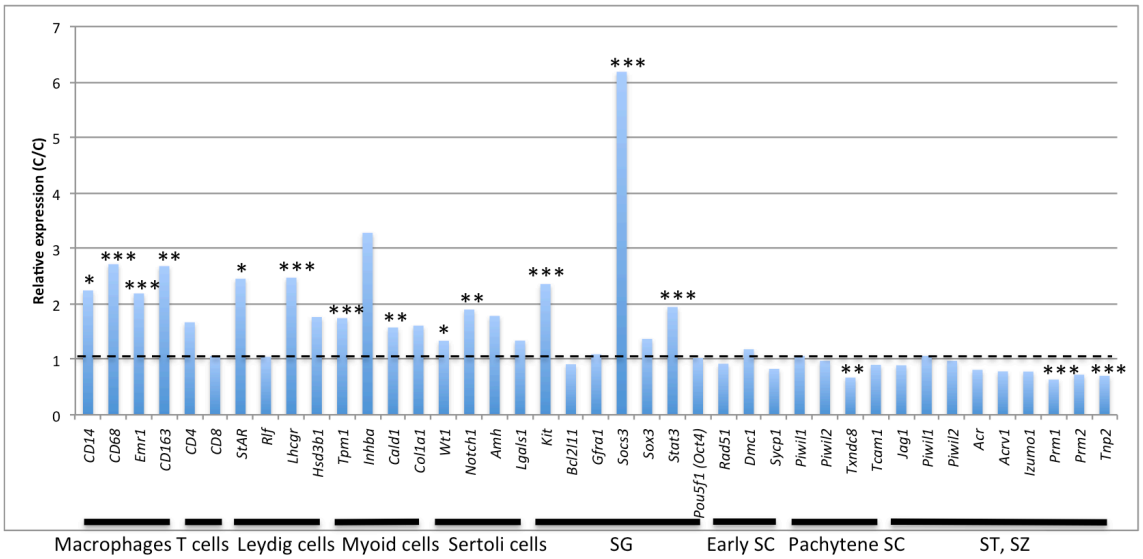
Supplementary Figure 2. GO terms enriched in significantly upregulated genes



Supplementary Figure 3. GO terms enriched in significantly downregulated genes



Supplementary Figure 4. Expression of markers of particular cell types



Supplementary Figure 5

Supplementary Table 1: Details on Male Mice used for Fertility Study					
Male	**Male *Smn* Genotype**	**Female *Smn* Genotype**	**Total Pups Sired in 90 days (Total Litters)**	**Relative Testis Mass***	**Sperm Count (x10^6)**
1	WT	C/C	11 (3)	6.4	9.15
2	WT	C/C	15 (2)	6.31	16.5
3	WT	C/C	13 (2)	5.62	20
4	C/C	WT	3 (1)	1.73	0.165
5	C/C	WT	0 (0)	1.5	1.59
6	C/C	WT	0 (0)	1.72	0
7	C/C	WT	0 (0)	1.62	0.99
8	C/C	WT	0 (0)	1.14	1.05
9	C/C	WT	0 (0)	1.43	1.71
10	C/C	C/C	8 (1)	2.25	2.4
11	C/C	C/C	0 (0)	2.02	3.96

*Relative testis mass determined by dividing the testis mass by total body weight.

Supplementary Table 2: Sequencing and mapping of WT and C/C Transcriptome

Organ	Sex	Mouse Identifier	Genotype	Barcode	Millions of reads	% Mapped	% Multiply mapped
Testis	M	817	WT	ACAGTG	30.54	89.56%	6.34%
	M	841	WT	ATCACG	26.66	87.67%	6.59%
	M	1218	WT	GCCAAT	32.05	86.60%	6.75%
	M	1222	WT	CGATGT	23.75	86.29%	6.81%
	M	818*	C/C	GATCAG	20.92	84.22%	7.70%
	M	839	C/C	ACAGTG	23.49	91.80%	5.88%
	M	1395	C/C	TAGCTT	30.25	88.66%	5.87%
Liver	M	841	WT	ATCACG	25.29	79.89%	9.70%
	M	1218	WT	GATCAG	28.85	79.86%	12.43%
	F	826	WT	CGATGT	25.29	80.48%	11.40%
	F	1241	WT	GGCTAC	20.85	81.26%	9.98%
	M	818	C/C	ACAGTG	26.70	79.54%	9.88%
	M	839	C/C	ATCACG	27.08	81.18%	11.70%
	F	827	C/C	GCCAAT	28.78	80.02%	12.95%
	F	1097	C/C	CGATGT	25.27	80.78%	10.03%
Brain	M	817	WT	ATCACG	37.43	86.74%	6.65%
	M	841	WT	GATCAG	32.32	88.05%	6.78%
	F	1241	WT	CGATGT	31.57	86.85%	6.94%
	F	1244	WT	TAGCTT	36.17	92.56%	7.06%
	M	1395	C/C	ACAGTG	19.12	88.65%	6.98%
	M	818	C/C	ATCACG	31.98	94.70%	6.25%
	F	827	C/C	GCCAAT	27.11	87.40%	7.25%
	F	833	C/C	CGATGT	29.09	95.34%	6.48%

* - Outlier sample, which was removed from downstream analyses

Supplementary Table 3: Most highly upregulated genes identified by RNA-Seq				
Gene symbol	Ensembl ID	Log2 Fold Change	Standard error	Adjusted P value
Hist1h1c	ENSMUSG00000036181	2.03	0.20	8.71E-21
Lamb2	ENSMUSG00000052911	1.67	0.17	1.67E-19
Ddr1	ENSMUSG00000003534	1.71	0.17	1.67E-19
Apoe	ENSMUSG00000002985	1.92	0.19	2.26E-19
Cpe	ENSMUSG00000037852	1.39	0.14	1.42E-18
Serping1	ENSMUSG00000023224	1.93	0.20	1.42E-18
Aldh1a1	ENSMUSG00000053279	1.47	0.16	1.14E-17
Sparcl1	ENSMUSG00000029309	1.70	0.18	2.28E-17
Aebp1	ENSMUSG00000020473	1.82	0.20	8.94E-16
Timp2	ENSMUSG00000017466	1.12	0.13	8.45E-15
Wnt6	ENSMUSG00000033227	1.55	0.18	1.63E-14
Igfbp3	ENSMUSG00000020427	2.45	0.29	1.63E-14
Rcn3	ENSMUSG00000019539	1.97	0.23	2.14E-14
Lamb1	ENSMUSG00000002900	1.43	0.17	3.76E-14
C1ra	ENSMUSG00000055172	2.02	0.25	2.07E-13
Arpc1b	ENSMUSG00000029622	1.22	0.15	2.36E-13
Rgs11	ENSMUSG00000024186	2.16	0.26	2.43E-13
Clec3b	ENSMUSG00000025784	2.05	0.26	8.20E-13
Ildr2	ENSMUSG00000040612	2.63	0.33	9.31E-13
Pdgfra	ENSMUSG00000029231	1.55	0.19	1.24E-12
Prss35	ENSMUSG00000033491	1.96	0.25	1.64E-12
Kit	ENSMUSG00000005672	1.24	0.16	2.78E-12
Col6a6	ENSMUSG00000043719	2.38	0.30	3.29E-12
Lrp1	ENSMUSG00000040249	1.39	0.18	6.16E-12
Lgmn	ENSMUSG00000021190	1.11	0.14	1.00E-11
Ndrg2	ENSMUSG00000004558	1.42	0.19	1.00E-11
Dcn	ENSMUSG00000019929	1.87	0.24	1.02E-11
Tspan4	ENSMUSG00000025511	1.70	0.22	1.27E-11
Lipg	ENSMUSG00000053846	1.29	0.17	1.29E-11
Phlda3	ENSMUSG00000041801	2.44	0.32	1.34E-11
Mpeg1	ENSMUSG00000046805	1.53	0.20	1.41E-11
9030617O03Rik	ENSMUSG00000021185	2.73	0.36	1.59E-11
Sulf2	ENSMUSG00000006800	1.29	0.17	2.82E-11
Aox1	ENSMUSG00000063558	1.53	0.20	3.04E-11
Kif1c	ENSMUSG00000020821	1.20	0.16	3.26E-11
Tshz2	ENSMUSG00000047907	1.26	0.17	3.69E-11
Ctsz	ENSMUSG00000016256	1.77	0.24	4.75E-11

Supplementary Table 3 Continued				
Gene symbol	Ensembl ID	Log2 Fold Change	Standard error	Adjusted P value
Meg3	ENSMUSG00000021268	1.69	0.23	5.23E-11
Osr2	ENSMUSG00000022330	1.73	0.23	5.23E-11
Lamc3	ENSMUSG00000026840	1.34	0.18	5.65E-11
Myh9	ENSMUSG00000022443	1.18	0.16	6.22E-11
Vim	ENSMUSG00000026728	1.08	0.15	7.12E-11
Aplp1	ENSMUSG00000006651	1.39	0.19	9.74E-11
C1qa	ENSMUSG00000036887	1.89	0.26	9.81E-11
Timp1	ENSMUSG00000001131	3.08	0.42	1.12E-10
Aldh2	ENSMUSG00000029455	1.24	0.17	1.13E-10
Cyp27a1	ENSMUSG00000026170	1.56	0.21	1.13E-10
Akr1cl	ENSMUSG00000025955	1.58	0.22	1.15E-10
Ccdc8	ENSMUSG00000041117	1.45	0.20	1.51E-10
Cdkn1a	ENSMUSG00000023067	1.43	0.20	1.53E-10

Supplementary Table 4: Most highly downregulated genes identified by RNA-Seq				
Gene symbol	Ensembl ID	Log2 fold change	Standard error	Adjusted P
Tmem56	ENSMUSG00000028132	-1.06	0.13	1.58E-12
Tmsb15a	ENSMUSG00000060726	-1.23	0.17	6.05E-11
Hypm	ENSMUSG00000040456	-1.03	0.14	1.43E-10
Tppp2	ENSMUSG00000008813	-1.02	0.14	3.04E-10
Oaz3	ENSMUSG00000028141	-0.89	0.13	3.82E-10
Hist1h1t	ENSMUSG00000036211	-1.55	0.22	6.97E-10
1700012A03Rik	ENSMUSG00000029766	-1.00	0.15	2.07E-09
1700027A15Rik	ENSMUSG00000101968	-0.95	0.15	2.50E-08
Ppp2r2b	ENSMUSG00000024500	-0.92	0.15	3.62E-08
1700009N14Rik	ENSMUSG00000028287	-0.83	0.13	3.95E-08
1700031F10Rik	ENSMUSG00000047046	-1.31	0.21	7.50E-08
1700015G11Rik	ENSMUSG00000094445	-0.81	0.13	8.62E-08
Spata18	ENSMUSG00000029155	-0.84	0.14	1.00E-07
Lrrc8b	ENSMUSG00000070639	-0.72	0.12	2.07E-07
Abcb5	ENSMUSG00000072791	-3.26	0.54	2.23E-07
1700016H13Rik	ENSMUSG00000029320	-0.75	0.13	4.34E-07
Yam1	ENSMUSG00000098178	-1.04	0.18	5.05E-07
Hist1h2ba	ENSMUSG00000050799	-1.06	0.18	5.44E-07
Clmn	ENSMUSG00000021097	-0.74	0.13	8.70E-07
Gm23935	ENSMUSG00000076258	-0.95	0.17	1.14E-06
Pdzd8	ENSMUSG00000074746	-0.72	0.13	1.37E-06
Nsun4	ENSMUSG00000028706	-0.79	0.14	1.71E-06
Tnp1	ENSMUSG00000026182	-0.62	0.11	1.84E-06
H2afb1	ENSMUSG00000062651	-0.72	0.13	2.16E-06
Lars2	ENSMUSG00000035202	-0.92	0.17	2.52E-06
Npy	ENSMUSG00000029819	-3.29	0.60	2.67E-06
1700020G17Rik	ENSMUSG00000101801	-0.86	0.16	2.79E-06
Gm11780	ENSMUSG00000085155	-0.83	0.15	2.98E-06
Klk1b8	ENSMUSG00000063089	-0.87	0.16	3.37E-06
Prom1	ENSMUSG00000029086	-0.68	0.13	3.79E-06
Prkar2a	ENSMUSG00000032601	-0.72	0.13	3.79E-06
Cdrt4	ENSMUSG00000042200	-0.73	0.14	4.09E-06
Gm498	ENSMUSG00000031085	-0.77	0.14	4.16E-06
BB014433	ENSMUSG00000049008	-0.71	0.13	4.17E-06
Rpl39l	ENSMUSG00000039209	-0.68	0.13	4.24E-06
1700019O17Rik	ENSMUSG00000036574	-0.69	0.13	4.35E-06
Lrrc74a	ENSMUSG00000059114	-0.75	0.14	4.51E-06

Supplementary Table 4 Continued				
Gene symbol	Ensembl ID	Log2 fold change	Standard error	Adjusted P
Gm5893	ENSMUSG00000011350	-0.66	0.12	5.20E-06
1700001C19Rik	ENSMUSG00000047150	-0.71	0.13	5.26E-06
Dbil5	ENSMUSG00000038057	-0.79	0.15	5.66E-06
St6galnac2	ENSMUSG00000057286	-0.76	0.14	5.66E-06
Spem1	ENSMUSG00000041165	-0.77	0.14	6.10E-06
Bpifa3	ENSMUSG00000027482	-0.80	0.15	7.41E-06
Lrrc57	ENSMUSG00000027286	-0.74	0.14	7.52E-06
1700080E11Rik	ENSMUSG00000032566	-0.84	0.16	8.56E-06
Akap4	ENSMUSG00000050089	-0.72	0.14	9.03E-06
Ubqlnl	ENSMUSG00000051437	-0.68	0.13	9.06E-06
Mir6236	ENSMUSG00000098973	-1.07	0.21	9.06E-06
Acsl1	ENSMUSG00000018796	-0.71	0.14	9.07E-06
1700001K19Rik	ENSMUSG00000056508	-0.74	0.14	9.60E-06

Supplementary Table 5: QPCR validation of RNA-Seq

Gene	Selected GO terms/KEGG pathways	Allele C relative expression: QPCR	Allele C relative expression: RNA-Seq
Unc5b	Netrin receptor, apoptotic process, axon guidance	1.327	1.992
Ntn3	Receptor binding, axonogenesis	2.292	2.371
Gpx4	Response to oxidative stress, spermatogenesis	0.824	0.727
Casp8	Apoptotic process, cysteine-type endopeptidase activity	2.112	2.277
Bcl3	DNA damage response, negative regulation of apoptotis	2.721	2.141
Spatc1	Gamma-tubulin binding, centrosome	0.823	0.635
Cdkl2	Cyclin-dependent protein serine/threonine kinase activity	0.995	0.697
Hdac7	B cell activation, histone deacetylation, vasculogenesis	1.964	1.798
Bax	Sertoli cell proliferation, apoptotic process	2.171	2.254
Ntn1	Axon guidance, proteinaceous extracellular matrix	1.352	1.529
EphB1	Axon guidance, angiogenesis, cell adhesion	1.956	1.927
HoxD10	Regulation of transcription, embryonic limb morphogenesis	2.304	2.813
Mak	Cyclin-dependent protein serine/threonine kinase activity	0.892	0.730
Tssk3	Protein serine/threonine kinase activity, spermatogenesis	0.900	0.612
Gapdhs	Glycolytic process, sperm motility	0.834	0.659
Sema7a	Axon guidance, integrin binding	1.725	1.987
Itgb1	Axon guidance, cell-matrix adhesion	1.619	1.778
Npy	GPCR binding, neuropeptide signaling pathway	0.012	0.103
Lamb2	Integrin binding, axon guidance, basement membrane	2.145	3.174
Lss	Cholesterol biosyntheic process	1.628	1.752
Meg3	DNA methylation, genetic imprinting	3.225	3.224
Cxcl12	Chemokine activity, germ cell development, axon guidance	3.723	2.747
Ddr1	Collagen binding, protein tyrosine kinase activity	1.737	3.274
Msmo1	Cholesterol biosyntheic process	2.906	2.517
Neat1	Protein binding, nuclear body organization	1.977	2.009
Slit3	Axon guidance, cell differentiation	2.217	1.922
Tppp2	Tubulin binding	0.679	0.493
Apoe	Cholesterol homeostatis, lipid homeostasis	3.352	3.795
Dhcr7	Cholesterol biosynthetic process	1.920	2.015
Gm9999	Unknown function	0.855	0.624

Supplementary Table 5 Continued

Gene	Selected GO terms/KEGG pathways	Allele C relative expression: QPCR	Allele C relative expression: RNA-Seq
Ppp3r2	Calcium-dependent protein SR phosphatase activity	0.813	0.752
Oaz3	Polyamine biosynthetic process	0.773	0.538
Lipa	Cholesterol biosynthetic process	2.511	2.092
RP23-402F16	Unknown function	0.931	0.558
Fus	RNA binding	1.356	1.400
Robo1	Axon guidance	1.653	1.667
Srgap1	Axon guidance, Rho protein signal transduction	1.425	0.690
Ppp2r2b	Apoptotic process, signal transduction	0.757	0.528
Cpe	Peptidase activity, insulin processing	2.298	2.628
Serping1	Negative regulation of complement activation	4.043	3.798
Dhcr24	Cholesterol biosynthetic process	1.740	1.928
Malat1	Regulation of alternative mRNA splicing	2.385	1.906
Efnb1	Axon guidance, cell differentiation	2.576	2.889
Srpk1	Protein serine/threonine kinase activity, RNA splicing	0.955	0.852
Lsm2	RNA splicing	0.874	0.773
Rbmxl2	RNA splicing	0.790	0.910
Rbmx	RNA splicing, DNA-templated transcription	1.310	1.558
SFPQ	RNA splicing, DNA repair, DNA-templated transcription	1.186	1.068
Gemin2	Spliceosomal snRNP assembly	0.903	0.863
Gemin3	Helicase activity, Spliceosomal snRNP assembly	0.996	0.678
Gemin4	Spliceosomal snRNP assembly	0.729	0.739
Gemin5	snRNA binding, spliceosomal snRNP assembly	0.929	0.843
Gemin6	Spliceosomal snRNP assembly	0.873	0.777
Gemin7	Spliceosomal snRNP assembly	1.008	0.994
Gemin8	Spliceosomal snRNP assembly	1.025	1.037
Ptbp1	RNA splicing	1.082	1.124
Ptbp2	RNA splicing	1.152	0.845
Hnrnpa1	RNA splicing	1.264	1.152

Supplementary Table 6: Primers used for QPCR		
Gene	Forward Primer	Reverse Primer
Gemin2	GGAGAGTGCAGATCGAAGCA	GCCACTTGTTGTTGTTGCCA
Gemin3	AGTCGGGCACTGGAAAAACT	ACCAGGAGATCCAACAGCAA
Gemin4	GGAAGACCTCAACACGACGT	GCAGGGAAGGCAGTGAGAAT
Gemin5	ACAGCATAACCCCTGGAAGC	AGCCAGGACTTTGCCATCTC
Gemin6	GCAACTGGTAACCATGAGCG	GGCCCATGATTCCAGTCACA
Gemin7	ATCCCCCAGAGTCACAGGAA	AGGTCTGTGGCTCCAAACTG
Gemin8	AGAGGGAATGAAGCACGGTG	GGTGGTTGAAGTACAGGCCA
Ptbp1	AGTCTGGAAACCTGGCCTTG	GATCTTCAGGACGGTGCCAA
Ptbp2	GGCAACAGAGGAAGCAGCTA	GAACTACTTGCGCACGTTGG
Hnrnpa1	ACGAGAGTCTGAGGAGCCAT	TCCACTTCTTCCACAGTGGC
Srpk1	ACAAAGCCCAAAGGAAGCCT	TCCCCAGCCCAATTTTCGAA
Lsm2	GGCAAGGATGTAGTCGTGGA	AGGGTATTTCTCAGGGTCTGTG
Rbmxl2	TCTTCATCGGCGGTCTCAAC	TTCTCGAAGGTGACGAAGGC
Rbmx	GTTCGTAGCAGCAGTGGACT	TCCATCATCTCTCGGGGACA
Sfpq	GCCAGCAGCAAGAAAAGCAT	TGTGCCATGCTGAGCAAAAC
HisH1C	GGCATCCTGGTGCAAACCAA	CCTTTTTGGGTGTGGCAGCA
Unc5b	ACGGCCAACTACACCTGTGT	ACAGAAGGCGCCTCCATTGA
Netrin 1	TGTCTCAACTGCCGCCACAA	TTGCAGGGACATTGGCCAGT
Netrin 3	ATGCCGCTTCAACATGGAGC	TGGCAGTCACAAGCTCTGCA
Casp7	AACCGTCCACAATGACTGCTCT	TGTCACGCCATCTTTCCCGTAA
Casp8	AGCTGCGGGATCCAGACAAT	TGCCAGCATGGTCCTCTTCT
Casp9	GGACCGTGACAAACTTGAGCAC	ATCTCCATCAAAGCCGTGACCA
Akt1	ACACCTTTATCATCCGCTGCCT	CTCTTCAGCCCCTGAGTTGTCA
Apaf1	AGAGATCCACACAGGCCATCAC	ATCACACCGTGAACCCAACTCA
Capn2	TTGACGCCAATGAGGAGGACAT	CCACTCCCATCTTCATCCAGCA
Aifm1	CCGTTCGGAGAGTGAGACAGAG	TAGCACAATCCCCACCACAACT
Bcl3	ACAGCGGCCTCAAGAACTGT	TGGCACTTTGGTCTGGGGAT
Bax	CCCACCAGCTCTGAACAGATCA	TGTCCACGTCAGCAATCATCCT
Fus	AGTTGAAGGGTGAGGCAACAGT	TTCCACCACCCCGATTGAAGTC
HoxD10	CGAAGTGCAGGAGAAGGAAAGC	AGCGTTTGGTGCTTGGTGTAAG
EphB1	GCAGCAGGAAACGAGCTTACAG	CGGCTTCATTGGGGTCCTCATA
Hdac7	CTCTTGAGCCCTTGGACACAGA	GGGTTTGTAGCGCAACTTCAGG
Spatc1	ACGCCCCACCGAAAGTCTAA	TGGCACAGGTCCTCATCCAA
Cdkl2	AGTCTCCCAGTCTGGCGTTG	CGTCACCGACCAGCAGTTCT
Mak	CGAGGCCATCCAGCTCATGA	CCCGGCAGGCTGATCAAGTA
Tssk3	CTTTGCCAAGGTGCTACCCAAG	ATACAGGACCACACCCATGCTC
H1fnt	GCTAAGGACCTAGTGCGTTCCA	GTACTGCTGCTCTTTGGCACTG
Gpx4	CGCTCCATGCACGAATTCTCAG	GGCCAGGATTCGTAAACCACAC

Supplementary Table 6 Continued		
Gene	Forward Primer	Reverse Primer
Hist1h1a	TCTACTGCCACGGAGAAACCTG	CCACTGCGCTCTTTGGAAGAAG
Hils1	AAATGGTGGCTGGAGACCAAGA	TTATGTACCCGCTTGTCAGCCA
Dnahc8	AGGCTGAGCTGGATAAAGTGCA	GGTAGGAGAGAAAGCCTGTGCA
Gapdhs	GTGTCAGTTGTGGACCTGACCT	GGGATTGCCGTTAAAGTCCGTG
Efnb1	AACGTCCAATGGGAGCTTGGAG	TAGTCAACTGCTCGGGTGTCAC
Sema7a	CGGAAGCTCTATGTGACCTCCC	GTGTGGCTCCGCTGGATTAATG
Cxcl12	GGTAAACCAGTCAGCCTGAGCT	TTGTTGTTCTTCAGCCGTGCAA
Slit3	GTGGCGACTGACAAGGACAATG	CCACACTGTGAAACTGGCCATC
Itgb1	GTGAGACATGTCAGACCTGCCT	TCCTTGCAATGGGTCACAGGAT
Robo1	CTCCCCCACATAGCAACAGTGA	GTAGGACACAGCAGCTGGAGAA
Ppp3r2	GCTGGTGGACAAGAGCATCTT	AAGTCTTCTTGACCGTGTTCTACA
Ppp3cc	AGGAAGAACCAAGCAACTGGCT	TGAAGTTTGGGAGCCAGTAGGG
Srgap1	CTGTACTTCCGAGGGCTGGAAA	CGTCACTGTACTGGGAGAGGTG
Npy	CTGACCCTCGCTCTATCTCTGC	CAGTGTCTCAGGGCTGGATCTC
Tmsb15a	ACCAGACTTGTCGGAAGTGGAG	ACTCCTTCTCTTGCTCGATGGT
Tppp2	GGCCAAGAATGCCAGAACCATC	TGGTGTCTGTCAACCGGTCTAC
Oaz3	GGATCAAGGCAACCGAGAAAGC	TCTAACCACCTCGAAGCCCATG
Ppp2r2b	ACCGGAAGATCCAAGCAACAGA	AACTTGCTGCGGAGGTAGTCAT
Lamb2	GGGTGAGAGACAGAAGGCAGAG	ACCGGCAGAGTTCAGAGACTTC
Ddr1	GACTCCTCTGACACCTTCCCAC	TGAGAAGCAGCAGCAGGATGAT
Apoe	GTCACATTGCTGACAGGATGCC	CTTCCTGGACCTGGTCAGACAG
Cpe	TCTCTGTGGACGGGATAGACCA	CAAGCTCAAAGTCCACCCCAAC
Serping1	CCACTTACCTGACGATGCCTCA	TGGCAGACACCTGAAGATCTGG
Lss	AGGAGCACGTTTCTCGGATCAA	AGGGCAGAAACTCAGGTCTGTG
Msmo1	GCTGTGCAGTCATTGAGGACAC	GGGATGTGCGTATTCTGCTTCG
Dhcr7	GCCTGGACCCTCATTAACCTGT	AAGTGGTCATGGCAGATGTCGA
Lipa	AATTTTGCTTCAGGCCCGCTAC	GCGCAAAGCTCCTTCATGATGA
Dhcr24	GATGGACATCCTGGAGGTGGAC	GGTCATCAAGCTCAGGCAACAC
Meg3	CCATCTCCACAGAAGAGCAGCT	CCCACGCAGGATTCCAGATGAT
Neat1	TGTTCACCATGGCAAGCAGATG	ACAAGCTGACTTCAACGATGGC
Gm9999	GGGACCCCTCTGGAAGTAAACA	TGTGCGATTACTTTATGCGAAACCT
RP23-402F16.1	ATGAAGACTTCTCCACCCAGCC	CCACCCTACCTACTGTAGTACGC
Malat1	TTCATGGAGCTGCTCAGGACTT	ACCTGAAGTCAAGACACCTGCA
Smn	AGAACAGAACACTCAGGAGAATGA	CTTTCCTGGTCCTAATCCTGA
FL SMN2	TTCCTTCTGGACCACCAATAA	TCTATGCCAGCATTTCTCCTTAATTTAAG

Supplementary Table 7: Primers used for semiquantitative PCR		
Gene	Forward Primer	Reverse Primer
Add3	CTAACCCATTCAGCCACCTC	CATCATGCATCTCGTCCTTG
Lrrc16a	CTTCCACCCATCCTCAAAAA	CTCCCTTGGTGACCATCTTT
Picalm	GCCAAATGACCTGCTTGACT	CCAGAATCTACAGCAACATTCG
Wt1	CCAGCTTGAATGCATGACCTGG	CTCACTCTCATACCCTGTGCCG
Sulf1	CACACAGTAGAACGGAGCATCT	ACTAACCTTCCCATCCATCCCA
SMN2	CGCGGGTTTGCTATGGCGAT	TTAGTGCTGCTCTATGCCAGCATTT
Hybrid Smn	GAAGAATGCCACAACTCCCTTGA	TTAGTGCTGCTCTATGCCAGCATTT

CHAPTER III

IN VITRO SELECTION REVEALS HIGH AFFINITY RNA TARGETS OF THE SURVIVAL MOTOR NEURON PROTEIN

A manuscript currently under preparation

Eric W Ottesen, Natalia N Singh, Ravindra N Singh

Abstract

The Survival Motor Neuron (SMN) protein is an essential protein required for survival of all animal cells. Low levels of SMN result in spinal muscular atrophy (SMA), a leading genetic disease of children and infants. Consistent with the multiple functions associated with SMN, several proteins interact with SMN. There is also evidence to suggest that SMN directly interacts with RNA. However, there is no study demonstrating the diversity of RNA motifs with which SMN interacts with high specificity and affinity. Here we employ Systematic Evolution of Ligands by Exponential Enrichment (SELEX) to identify RNA sequence and structural motif(s) of SMN. Our results reveal a combination of sequence motifs and structural contexts that drive the specificity of SMN-RNA interactions. Among sequence motifs that were frequently selected include but not limited to UGG and UGC. Mutations within these motifs caused a noticeable decrease in binding. All of the selected sequences had at least two terminal stem-loop (TSL) structures. We show that the strengthening of the stem of the 3′ TSL in one of the high-

affinity sequences substantially improved the RNA-SMN interaction. Our results of truncation and substitution experiments suggest a requirement for multiple contacts between SMN and RNA to maintain the high affinity. We demonstrate that both affinity and specificity of SMN-RNA interaction are influenced by salt concentrations. Our findings bring a new perspective to SMN-RNA interaction and suggest a wider role of SMN in RNA metabolism.

Introduction

Spinal muscular atrophy (SMA) is a genetic disorder which leads to the progressive deterioration of motor neurons. It is the most prevalent genetic cause of infant mortality, affecting approximately 1 in 10,000 infants (Wirth et al 2006). SMA is caused by homozygous deletion or mutation of the *Survival Motor Neuron 1* gene (*SMN1*) (Lefebvre et al 1995). Humans have a second gene, *SMN2*, which is only partially functional due to a C-to-T transition at the sixth position of exon 7. Although translationally silent, the C-to-T transition (C-to-U in the mRNA) interferes with efficient splicing of *SMN2* exon 7, causing predominant skipping of this exon (Lorson et al 1999). The major protein produced by *SMN2*, SMNΔ7, is extremely unstable due to the creation of a C-terminal degradation signal, or degron (Cho and Dreyfuss 2010).

The full extent of the function(s) of the SMN protein remains an active topic of research. The most well-studied function of SMN is the assembly of small nuclear ribonucleoproteins (snRNPs) (Meister et al 2001, Zhang et al 2008). The U1, U2, U4, U5, and U6 snRNPs form the catalytic core of the major spliceosome, which is required for

the removal of most eukaryotic introns (Wahl et al 2009). The snRNPs consist of a RNA component, known as the small nuclear RNA (snRNA) and a septameric ring of proteins known as the Smith (Sm) proteins (Lsm for U6), as well as proteins specific to each snRNA (Will and Lührman 2001, Wahl et al 2009). SMN enhances the association of snRNAs with the Sm ring in conjunction with a number of other proteins, which form a large macromolecular complex known as the SMN complex (Pellizzoni et al 2002). Components of the SMN complex include Gemins 2-8 and Unrip (Cauchi et al 2010). In the current model of snRNP assembly, Gemin5 is the sole component of the SMN complex that directly interacts with snRNAs (Yong et al 2010). Recently, however, it has been shown that U1-70K can directly recruit SMN to enhance Sm recruitment on U1 snRNA independently of Gemin5 (So et al 2016). This is likely in order to meet an increased demand for U1 snRNP assembly, as U1 is more abundant than all other snRNPs combined (Kaida et al 2010). Currently, it is not known whether SMN in SMN complex directly contacts RNA in a sequence and/or structure-specific manner.

Several implicated functions of SMN involve interaction with RNAs directly or indirectly. These functions include but are not limited to mRNA transport (Rossoll et al 2003, Akten et al 2011, Fallini et al 2011, Fallini et al 2014), stress granule formation (Hua et al 2004, Zou et al 2011), biogenesis of the signal recognition particle (SRP) (Piazzon et al 2013), transcription (Strasswimmer et al 1999, Zhao et al 2016), and translation (Sanchez et al 2013). The SMN protein has been shown to bind RNA *in vitro* (Lorson and Androphy 1998, Bertrandy et al. 1999). This activity was localized to the N terminal region of the protein, specifically the region coded by exons 2A and 2B (Lorson and Androphy 1998). However, not much is known about SMN's binding specificity,

aside from a preference for poly(G) oligonucleotides (Lorson and Androphy 1998, Bertrandy et al. 1999). A recent study employed RNA immunoprecipitation followed by microarray analysis (RIP-Chip) to capture SMN-interacting RNAs (Rage et al., 2013). However, sequence specificity was not investigated.

Systematic Evolution of Ligands by Exponential Enrichment (SELEX) is a powerful technique that uses iterative selection to identify high-affinity nucleic acid sequences (ligands) that bind a target molecule of interest including proteins, nucleic acids and small compounds (Tuerk and Gold 1990, Ellington and Szostak 1990, Jayasena 2001, Stoltenburg et al 2007). SELEX has been frequently employed as a decisive tool to identify the sequence specificity of RNA binding proteins (RBPs) (Tacke and Manley 1995, Dember et al 1996, Dubey et al 2005).

Here, we use SELEX to identify the sequence specificity of RNA molecules (ligands) that bind to purified SMN with high affinity. We performed nine rounds of selection to enrich RNA pool showing about ~35 fold higher affinity for SMN than the initial pool used for the selection experiment. We observed that the preference of the selected motifs included but was not limited to UGG and UGC. Selected sequences folded into a variety of the predicted secondary structures. Mutations and/or deletions abrogating the predicted structures impacted the affinity of RNA-SMN interactions. We show that the RNA-SMN interaction is strongly influenced by binding conditions, specifically the concentration of salt in the binding reaction. Our results bring new perspective to the mode of action of SMN and may suggest new functions of SMN in RNA metabolism.

Results

Selection of high-affinity sequences by SELEX

In order to determine whether SMN binds RNA in a sequence-specific manner, we performed SELEX. We began by purifying recombinant SMN from E. Coli expressing SMN via an IMPACT system vector (Supplementary Figure 1). The IMPACT system generates tag-free protein due to its self-cleaving intein tag (Chong et al 1997). The initial pool comprised of template oligonucleotides that contained a T7 promoter, 5′ and 3′ constant regions for PCR amplification, and a 25 nucleotide (nt) randomized region (Figure 1A). Of note, randomized regions of this length or lower have been used in a number of SELEX experiments (Tuerk and Gold 1990, Ghisolfi-Nieto et al 1996, Takagaki and Manley 1997, Dubey et al 2005, Skrisovska et al 2007). We sequenced 50 clones from the initial pool to test for sequence bias. We did not observe any duplicate sequences or biased base composition in the randomized region (Table 1). We used 10 randomly selected clones from pool 0 to establish a baseline binding affinity of SMN for unselected sequences by the nitrocellulose filter paper binding assay (Wong and Lohman 1993) (Figure 2A).

Selection was performed using nitrocellulose filter paper to capture RNA-protein complexes (Figure 1B; Singh et al., 2002). The initial RNA input for selection was an estimated 2.4 x 10^{14} molecules, generated from 1.2 x 10^{14} molecules of DNA template, resulting in an estimated 2 copies of each unique molecule. In order to capture rare sequences with high affinity, we started with an excess of SMN protein over RNA and gradually increased the selection stringency over the first 4 rounds of selection by

reducing the amount of protein in the reaction (Table 2). After 6 rounds, we performed a binding assay with a limited number of sequences and discovered a mix of sequences with low and high affinity for SMN (Figure 2B). In order to remove the majority of low-affinity sequences, we performed a highly stringent round of selection with more than 10-fold less protein than RNA, followed by two more rounds with near equimolar concentrations (Table 1). The final pool (P9) had roughly 35-fold higher increase in binding than the initial pool (Table 1).

Characterization of selected sequences

We next cloned the final pool and sequenced 99 randomly selected clones (Table 3). We observed 10 pairs of identical sequences, suggesting that the selection was nearing completion. In order to reduce bias coming from identical sequences, we counted duplicate clones as one clone for analysis. The selected sequences tended to be G- and U-rich with following occurrence: 11.9% A, 31.9% U, 16.5% C, and 39.8% G (Table 3). We examined the selected sequences for enrichment of specific trinucleotides (Table 4). In particular, we noticed high occurrences of UGC, UGG, GUG, and UUG. Since UGG was frequently represented in the initial pool as well (Table 4B), we infer that the specific enrichment of UGG did not occur during selection. We searched the selected sequences for additional enriched sequence motifs using MEME (Bailey et al 2009). We observed enrichment for a single motif, GUGCG (*E*-value=4.6×10^{-9}). This motif frequently occurred at the 5′ end of the randomized region, although location varied somewhat and some sequences had multiple occurrences of the GUGCG motif (Table 3). We also

obtained high-affinity clones without GUGCG motif suggesting a pentameric motif is not required for the high affinity and/or specificity of RNA-SMN interaction.

We analyzed the binding of 78 sequences from the final pool (Table 3, Supplementary Figure 2). We observed a wide range of binding affinities for individual sequences (Supplementary Figure 2). For further analysis, we separated the sequences based on their binding affinity into two categories: non-binders, with <20% of the binding compared to pooled P9 RNA, and binders, with >20% of the binding compared to the pool. The base composition of the two groups of clones was similar; however, the SMN-binding sequences had a higher prevalence of the GUGCG motif, with a 55.6% (25/45) occurrence, compared to 12.1% (4/33) of the non-binders (Table 3). However, due to the presence of low binders with GUGCG and high binders without, we conclude that presence of a GUGCG motif is not the sole determinant of the strength of SMN-RNA interaction.

We chose the 10 highest-affinity binders for further characterization (Supplementary Figure 3). Sequence-wise, they resembled the general population of selected clones, with high frequency of Gs and Us (Supplementary Figure 3A). We performed binding with these 10 clones (Supplementary Figure 3B) and selected the top 5 for further characterization. For the 5 highest binders, we generated binding curves and determined the apparent dissociation constants (K_d^{app}) (Figure 3B-3G). For an estimate of nonspecific binding to RNA we used P0-44 from the initial unselected pool. The calculated dissociation constants ranged from 20 nM in the case of P9-74 to 46 nM for P9-71. (Figure 3B-3F). In contrast, for the unselected RNA, we were unable to determine

an exact K_d^{app} in this assay as binding was not saturated even at the highest concentration of protein tested (Figure 3G).

We predicted the RNA secondary structures of all of the top 5 binders (Figure 3H). In all of the Mfold (Zuker et al 2003) predicted structures, at least two terminal stem-loops (TSLs) were formed. The 5′ TSL generally contained the GUGCG motif, although its position relative to the loop was not conserved. In both TSLs, the stems and loops were of variable length, although in general the 5′TSL had a smaller loop, ranging from 4-6 bases. The 3′TSL was variable both in loop size and stem length. Of note, all of the TSLs comprised of bases from both the constant and selected regions (Figure 3H).

Effect of RNA Sequence on Binding of SMN

We developed a comparative binding assay based on the nitrocellulose binding assay to assess the relative affinity of sequences with respect to a reference RNA that is common in all reactions. This approach has been successfully employed for an unbiased determination of the relative affinities of RNAs against a given protein (Singh et al 2002, Singh et al 2011). We chose P9-10 as the lead candidate for the examination of RNA-SMN interactions in further detail. We constructed P9-10E, a reference RNA, by adding 21 bases to the 3′ end of P9-10 (Figure 4A). We used P0-44, a sequence from the initial pool, as a negative control. As an additional negative control, we used P9-10RV that has sequence reverse to that of P9-10. Both P0-44 and P9-10RV showed background level of binding. Within P9-10, we tested the relative contribution of G residues within the first (positions 16, 18, 20, 21, 22, and 26) or second (positions 30, 32, 33, 38, and 39) half of the selected region. Our choice of G residue was based on the prevalence of G-rich motifs

in the selected region. Supporting that GUGCG motif is not critical for binding, G-to A substitutions within GUGCG motif or outside of this motif showed similar reductions in binding (Figure 4A, lanes 7-10). We next created 4 new mutants with less extensive base changes (Figure 4B, top panel). The first mutant, G1-3A, had a triple substitution at positions 16, 18, and 20 that changed the GUGCG sequence to AUACA. In addition, we made other G-A substitutions: G4-6A with mutations at positions 21, 22, and 26, G7-9A with mutations at positions 30, 32, and 33, and finally G10-11A with mutations at positions 38 and 39. G1-3A had the strongest effect on SMN binding, reducing the activity to ~30% of WT (Figure 4B, lanes 5-6), compared to ~50-65% for the other 3 mutants (Figure 4B, lanes 7-12).

Role of Flanking Sequences and Secondary Structure in SMN-RNA Interaction

All of the structures predicted to be formed by the top binding RNAs require extensive base pairing between the constant and selected regions of each RNA molecule (Figure 3H). In order to investigate the effect of RNA structure on binding of SMN, we generated a set of mutants with alterations of the constant region designed to disrupt particular structural features of the P9-10 RNA. P9-10 forms 3 predicted stem-loops: stem loop 1, which is formed by 2 base pairs within the 5′ constant region, stem loop 2, which is formed by base pairing between the 5′ constant region and 3 bases within the selected region, and stem loop 3, which is formed by extensive base pairing between the selected region and the 3′ constant region. We designed 2 mutants within the 5′ constant region that disrupted stem loop 2, 2 designed to disrupt stem loop 3, and 2 that reinforce stem loop 3, by either adding a new base-pairing partner for the bulging base or by

extending the stem into the loop by 4 bases (Figure 5). Both of the mutants within the 5′ constant region moderately reduced binding affinity (Figure 5, lanes 5-8). Surprisingly, eliminating the bulging base from the 3′ stem had a strong stimulatory effect on binding of SMN (Figure 5, lanes 9-10). Of all the mutants, the largest reduction in activity was achieved by modifying the 3′-most bases, which form the base of the 3′ stem (Figure 5, lanes 11-12). Finally, reducing or increasing base-pairing at the distal end of the 3′ stem had a moderate impact on SMN binding (Figure 5, lanes 13-16).

Effect of RNA Size on SMN-RNA Interaction

We also generated a series of truncated mutants in P9-10 RNA to identify the minimal binding region for SMN. All 3′ truncations reduced binding to background levels, even Δ48-57 which retains the entire selected region (Figure 6A, lanes 5-12). This may be due to the loss of base pairing to the selected region. Consistent with these results, mutation of the 3′-most bases had a similar reduction in binding (Figure 5, lanes 11-12). There is another noticeable drop in activity when 30 bases are deleted at the 3′ end, which could be due to either reduction in total size of the RNA molecule or loss of a secondary binding contact with SMN (Figure 6A, lanes 9-10). Yet more activity is lost with deletion of another 10 bases from the 3′ end (Figure 6A, lanes 11-12). This final deletion has lost almost all of the selected region, including the 5′ GUGCG motif (Figure 6A, top panel). Of the 5′ deletions, removal of the first 10 bases after GGG (required for efficient T7 transcription) has a moderate effect on SMN binding (Figure 6B, lanes 5-6), whereas further deletion abrogates binding almost completely (Figure 6B, lanes 7-10). The loss of binding coincides with loss of the GUGCG motif (Figure 6B, top panel).

Effect of Salt Concentrations on SMN Binding to RNA

Previous reports indicate that the RNA binding activity of SMN is largely resistant to salt, with strong binding at up to 250 mM and some residual binding at concentrations as high as 1 M (Lorson and Androphy 1998, Betrandy et al 1999). In order to test the binding of SMN to our selected RNA in various conditions of ionic strength, we performed our comparative binding assay in a range of salt concentrations (Figure 7). Notably, binding of P9-10 was almost completely lost at 200 mM NaCl (Figure 8A, lanes 1-7). These results are not totally unexpected, since we performed selection experiment at 150 nM. At 250 mM NaCl, binding is indistinguishable from reactions with SMN protein completely omitted (Figure 8A, lanes 6-7). We tested binding of two negative controls, P9-10 RV and P0-44 (Figure 7A, lanes 8-21). Both have weaker binding than the reference RNA at all concentrations of NaCl tested. Our results also supported that a lower salt concentrations promoted non-specific RNA-SMN interaction (Figure 7B). However, as expected, the specificity was highest a 150 nM, the concentration used for selection experiment.

Discussion

In this study, we employed SELEX to identify sequence motifs required for a strong interaction between RNA and the SMN protein. We performed 9 rounds of selection, obtaining a ~35-fold enrichment in relative binding (Table 2). The selected sequences

with the highest binding affinity have apparent dissociation constants ranging from 20 to 46 nM (Figure 3H). This is similar to the binding affinity of Gemin5 for snRNA, which binds U4 with a K_d of 30.5 nM (Lau et al 2009), and is typical for a specific interaction of an RBP with RNA, which can range from below 8 nM up to 200 nM or more (Ma et al 2002, Cammas et al 2007, Bhardwaj et al 2013, Schwartz et al 2013, Workman et al 2014).

Analysis of the sequences selected for interaction with SMN yielded a high enrichment for a GUGCG motif in several but not all clones. Interestingly, GUGCG is also present in the minor spliceosomal U11 snRNA and the 7SL RNA, which forms the RNA component of the SRP (Keenan et al 2001), both of which are known targets of the SMN complex (Lotti et al 2012, Piazzon et al 2013). However, several lines of evidence support that GUGCG motif was not sole criterion for RNA-SMN interaction. Consistently, several mutations outside the GUGCG motif reduced affinity for SMN interaction (Figure 5, Figure 6). Our results suggest that SMN contacts the RNA substrate at two or more sites. It is likely that one of the sites serves as the primary sites whereas the other sites serve as the secondary sites. Such a mode of binding has been observed for the bacterial protein CsrA (Dubey et al 2005) and for the mammalian testes-specific protein RBMY (Skrisovska et al 2007).

Because the motif occurred at a conserved location relative to the flanking regions and because 3 of the top 5 binders formed strikingly similar structures, we hypothesized that selection favored particular structures formed in the context of the 5′ and 3′ constant sequences. In fact, it is common for SELEX-derived sequences to recruit portions of the constant flanking sequences to form structures required for binding (Dubey et al 2005).

Supporting this conclusion, both mutation and truncation of the 3′-most portion of the flanking sequence of clone P9-10 had a significant effect on binding to SMN (Figure 5, Figure 6A). In fact, all of the mutations in the flanking sequences that we tested had some impact on binding, indicating the strong pressure that context exerted on the type of sequences that were selected.

Of all the mutants that we tested in this study, only one had a positive impact on SMN binding. This mutant, P9-10 B+1D1, reinforced the 3′ stem within P9-10 by providing a base-pairing partner to the U residue at position 28, which is normally unpaired and forms a bulge within the predicted stem. This indicates two things. First, it reinforces the critical role that RNA secondary structure plays in the binding of SMN, especially the role of the 3′ stem. Second, it suggests that SELEX was not able to identify fully idealized binding sites for SMN, possibly due to limited complexity of the initial RNA pool used for selection, or due to other constraints introduced during selection itself. These results support the possibility of even better RNA ligands for RNA-SMN interaction that we have generated in this study.

Binding of SMN to RNA was strongly impacted by the amount of salt in the binding buffer, with binding almost completely lost at concentrations greater than 200 mM NaCl. This is in direct contrast with previous studies, which found reduced but still measurable binding at concentrations as high as 500 mM-1 M (Lorson and Androphy 1998, Bertrandy et al 1999). Interestingly, the protein-protein interactions which drive the formation of the SMN complex appear to be much more stable in high-salt conditions, as indicated by the ability to isolate core SMN complexes in concentrations as high as 500 mM NaCl (Pellizzoni et al 2002). Our results support the likelihood of direct RNA-SMN

interactions at physiological conditions particularly in those cell types where SMN is excessively expressed compared to its interacting partners. Testis is one such tissue in which SMN is disproportionately expressed at high levels and role of SMN has been recently implicated in the development of testis (Ottesen et al 2016).

In this study, we establish that SMN preferentially interacts with sequence and structural RNA motifs with high specificity. The diversity of high affinity ligands isolated by SELEX support a broad spectrum of RNA-SMN interactions with implications to novel SMN functions in RNA metabolism. The extreme sensitivity of RNA-SMN interactions to salt concentrations suggests that the interface of RNA-SMN interaction is dynamic and may be amenable to manipulations by interacting proteins. Given the fact that multiple contacts within RNA is required for a tight RNA-SMN interaction, it is likely that the folding of the entire SMN determines the interface of RNA-SMN interaction. The nucleic acid-binding domain of SMN is located towards the N-terminus. Several SMN variants including a-SMN, SMNΔ7 and SMN6B retain the nucleic acid-binding domain of SMN. The overall folding of all other SMN variants are likely to be different due to the absence/difference of C-terminal residues. It will be interesting to see if the high affinity RNA ligands of SMN also interact with other SMN isoforms with similar affinity and specificity. Future studies will determine if the loss of the specific RNA-SMN interaction is also a contributory factor to SMA pathogenesis.

Materials and Methods

Chemicals, reagents, and enzymes

The manufacturer and product catalog numbers of all chemicals, reagents, and enzymes used in this study are provided in Supplementary Table 1. All oligonucleotides were obtained from IDT and are listed in Supplementary Table 2.

SMN Protein Purification

SMN protein was expressed and purified using the IMACT protein purification kit (NEB). Briefly, the coding region of *SMN* was cloned into pTXB3 and transformed into ER2566 cells. A small-scale culture was inoculated overnight from a single colony and grown overnight at 37˚C, then diluted 1:100 into 1 L of LB media. After 2.5-3 hours after OD_{600} reached 0.6-0.8, expression of SMN-Intein was induced by addition of IPTG (Sigma) to a final concentration of 0.4 mM. Induction was carried out for 5 hours at 30 ˚C. After induction, cells were placed on ice to cool and all steps from this point to the end of purification were carried out at 4 ˚C or on ice. Cells were collected by centrifugation at 6,000xg for 10 minutes, then resuspended in 20 mL of Column Buffer (20 mM Na-HEPES pH 8.5, 500 mM NaCl, 1 mM EDTA) supplemented with 0.1% Igepal CA-630 (Sigma) and 1X HALT protease inhibitor (Thermo Scientific). Cells were lysed by sonication on ice using a Microson XL ultrasonic cell disruptor (Misonix); 5-10 pulses of 20-25 seconds each. After sonication, lysates were subjected to centrifugation at 12,000xg for 15 minutes at 4 ˚C to remove debris, and the supernatant collected. Chitin columns were prepared by packing 10 mL of slurry (NEB) into a Poly-Prep

chromatography column (Biorad) and allowing the storage buffer to drain by gravity flow. Columns were then washed with 50 mL of Column Buffer. Clarified lysate was then applied to the washed column and washed with 100 mL of Column Buffer, followed by another 50 mL of Column Buffer with the NaCl concentration increased to 700 mM. Cleavage of the Intein tag was then induced by washing the column with 15 mL of Column Buffer supplemented with 50 mM DTT (Sigma) and 0.1% Igepal CA-630. The column was then capped to stop buffer flow and cleavage was allowed to proceed for ~32 hours. After cleavage, protein was eluted in 14 mL of Column Buffer supplemented with 0.1% Igepal CA-630 and concentrated using two Pierce Concentrators (20,000 KDa cutoff) (Thermo Scientific) to a final volume of ~1 mL. An equal volume of glycerol was added and protein stored in 80 µl aliquots at -80 ˚C.

SELEX initial pool generation and in vitro transcription

The double-stranded DNA template for the initial pool (P0) was generated by PCR using Taq polymerase (NEB) in a total volume of 1 mL containing 0.5 nMol of Temp-0 oligonucleotide as template and 5′Pool and 3′Pool primers to amplify. Template oligonucleotide and primer sequences are given in Supplementary Table 4. Limited amplification (4 cycles) was carried out to reduce bias. After amplification, the PCR product was concentrated by ethanol precipitation, separated on an 8% acrylamide gel, and purified by the crush and soak method (see below) and resuspended in 40 µl nuclease-free water. A 1 µl aliquot of DNA was then run on an acrylamide gel and quantified by visual comparison to a DNA ladder of known concentration. For transcription, 10.5 µg of PCR product was used as a template in a 100 µl reaction using

Megashortscript T7 (Ambion) in the presence of 7.5 mM each of ATP, CTP, GTP, and UTP and 1X T7 reaction buffer. Transcription was carried out overnight at room temperature. The reaction was then run on a denaturing 8% Urea-PAGE gel, visualized by UV shadowing, and purified using the crush and soak method (see below). After gel purification, RNA was resuspended in a final volume of 50 µl and further purified using an RNase-free Micro Bio-spin column (Biorad).

Gel extraction (crush and soak)

Bands corresponding to DNA templates or RNA transcripts were excised, transferred to a 1.5 ml microfuge tube, and crushed using a pipet tip. To elute nucleic acids from the gel, 700 µl of elution buffer (500 mM ammonium acetate, 1 mM EDTA pH 8.0, 0.1% SDS) was added and tubes were placed on a rotating mixer at 55 ˚C for 1 hour. Samples were then centrifuged at maximum speed on a tabletop microcentrifuge for 5 minutes and supernatant transferred to a new tube. An equal volume of phenol:chloroform was added, then samples were vortexed and centrifuged at maximum speed on a tabletop microcentrifuge for 5 minutes and supernatants were transferred to a new tube. Nucleic acids were precipitated by addition of 2.5 volumes of 100% ethanol, followed by incubation on dry ice for 5 minutes or at -20 ˚C overnight. Samples were then centrifuged at 14,000 rpm at 4 ˚C for 20 minutes and supernatant discarded. Pellets were washed with 1 mL of 70% ethanol and dried.

In Vitro Selection

Variable amounts of RNA and protein (See Table 1) were combined together in Binding Buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT). Selection was carried out in a total of 1 mL except for the first round, which was carried out in 4 reactions of 1 mL each. Prior to addition of protein and DTT, RNA was refolded by briefly heating to 95 ˚C, followed by incubation at 37 ˚C for 1 hour. Binding was carried out for 20 minutes at room temperature. Protein-RNA complexes were then captured by passing through a Protran BA-85 nitrocellulose filter (Whatman) pre-soaked in Wash Buffer 1 (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$) and washed once with 800 µl of Wash Buffer 1, followed by two washes with 800 µl of Wash Buffer 2 (20 mM Tris-HCl pH 7.5, 450 mM NaCl, 5 mM $MgCl_2$). Filters were then collected and placed in a 1.5 mL microfuge tube for RNA.

To extract selected RNA from nitrocellulose filters, 400 µl of Tris-EDTA buffer (pH 7.5) and 600 µl of phenol:chloroform (OmniPur) was added to the filters and the samples were vortexed for 3 pulses of 30 seconds each. Tubes were then centrifuged at maximum speed for 5 minutes in a tabletop microcentrifuge. The aqueous phase containing RNA was transferred to a new tube and the organic phase and filter were discarded. RNA was then precipitated by addition of 1 µl of 20 mg/mL glycogen (Roche), 1/10 volume of 3 M potassium acetate (pH 5.2), and 2.5 volumes of 100% ethanol, followed by incubation on dry ice for 5 minutes or at -20 ˚C overnight. Samples were then centrifuged at 14,000 rpm at 4 ˚C for 20 minutes and supernatant discarded. Pellets were washed with 1 mL of 70% ethanol and dried. RNA was then resuspended in 10 µl of RNase-free water. Half of the RNA was stored at -80 ˚C and the other half used

as a template for reverse transcription in a 20 µl reaction as follows: 5 µl of RNA was combined with 6 µl of nuclease-free water and 1 µl each of 10 mM dNTP mix and 50 µM 3′Pool primer, then heated at 65˚C for 5 minutes and cooled on ice. Then 4 µl 5X FS buffer and 1 µl each of 100 mM DTT, RNaseOUT (Invitrogen), and SuperScript III (Invitrogen) were added and the reaction was incubated at 50˚C for 50 minutes followed by heat inactivation at 70˚C for 15 minutes. Following reverse transcription, half of the cDNA was stored at -20 ˚C and half used as template for PCR. The subsequent double-stranded DNA pool was generated by PCR in a final volume of 1 mL. For each round, the number of PCR cycles ranged from 8-10 and was optimized by small-scale amplification before amplifying the entire pool. The pool was then purified and used for transcription as above. For cloning and sequencing, pools were amplified using the 5′Pool primer and 3′Hind primer, digested with EcoRI and HindIII restriction enzymes (NEB), and cloned into the pUC19 vector.

In vitro transcription for binding assays

For standard binding assays containing a single RNA, radiolabeled RNA was generated as follows: Either 1 µg of HindIII-digested plasmid or 1 µl of PCR-amplified DNA was transcribed in a 20 µl reaction containing 7.5 mM each of ATP, CTP, and GTP, 0.75 mM of cold UTP, 50 µCi of [α-^{32}P]UTP (Perkin-Elmer), 2 µl 10X T7 reaction buffer, and 2 µl MegaShortScript T7 enzyme mix (Ambion). Reactions were incubated overnight at room temperature. The reaction was then run on a denaturing 8% Urea-PAGE gel, visualized by UV shadowing, and gel purified using the crush-and-soak method. After gel

purification, RNA was resuspended in a final volume of 50 µl and further purified using an RNase-free Micro Bio-spin column (Biorad).

For transcription of high specific activity RNA for comparative binding assays or assays for determining K_d^{app}, 1 µg of HindIII-digested plasmid or 1 µl of PCR-amplified DNA was transcribed in a 20 µl reaction containing 0.5 mM each of ATP, CTP, and GTP, 14 µM of cold UTP, 60 µCi (1 µM) of [α-^{32}P]UTP (Perkin-Elmer), 2 µl 10X RNA polymerase reaction buffer (NEB), and 1 µl each of 100 mM DTT (Sigma), RNaseOUT (Invitrogen), and T7 RNA polymerase (NEB). Reactions were incubated overnight at 37 ˚C. The reaction was then run on a denaturing 8% Urea-PAGE gel and bands were visualized by exposing the gel to a phosphor imager screen for 2 minutes and screens were scanned using a Fujifilm FLA-5100 phosphorimager. Gel printouts were aligned to the gel by lining up the wells, and bands corresponding to full-length RNA were excised using a razor blade and transferred to a 1.5 mL microfuge tube. RNA was extracted using the crush and soak method similar to previous extractions, except that 0.5 µl of glycogen was added to each tube of eluted RNA immediately before addition of ethanol for precipitation.

Nitrocellulose filter paper binding assays

For direct measurement of SMN-RNA binding, 10 pmol each of Megashortscript-transcribed RNA and protein were combined together in 100 µl Binding Buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT) for a final concentration of 100 nM of each. Prior to addition of protein and DTT, RNA was refolded by briefly heating to 95 ˚C, followed by incubation at 37 ˚C for 1 hour. Binding was carried out for

20 minutes at room temperature. Protein-RNA complexes were then captured by passing through a Protran BA-85 nitrocellulose filter (Whatman) pre-soaked in Wash Buffer 1 (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$) and washed once with 800 µl of Wash Buffer 1, followed by two washes with 800 µl of Wash Buffer 2 (20 mM Tris-HCl pH 7.5, 450 mM NaCl, 5 mM $MgCl_2$). Filters were then collected and placed in a 7 mL liquid scintillation vial and Cerenkov emissions were measured in a Packard Tri-Carb 3170TR/SL liquid scintillation counter alongside a portion of input and a negative control reaction where protein was omitted.

For determination of the dissociation constant, the same procedure was carried out as above, except that RNA was transcribed using the high specific activity procedure and 30,000 cpm of RNA (an estimated 0.1 nM final concentration) as determined by Cerenkov counting was used as input for each assay. Protein input started at 100 pmol (1 µM final concentration) and followed a 2-fold dilution series until a minimum final concentration of 0.244 nM was reached. Final relative binding values were determined by subtracting the negative control value from each bound sample, dividing by input, then dividing again by the maximum value obtained for each set of assays.

For the comparative binding assay, RNA input was 30,000 cpm of high specific activity transcript for both the experimental RNA and the P9-10E control. Protein input was 5 pmol (50 nM final concentration). After binding and washing, filters were then collected and placed in a 1.5 mL microfuge tube. RNA extraction was carried out following the same procedure used during selection. After precipitation, RNA pellets were resuspended in 20 µl of Gel Loading Buffer II (Ambion). Input samples were prepared by setting aside 1 µl of reaction mixture prior to addition of protein and diluting

1:20 into loading buffer. 4 µl each of input and bound RNA was loaded in each well of an 8% Urea-PAGE denaturing gel. After running, gels were dried and exposed to phosphor image screens for 1 hour to overnight, depending on the strength of signal. Images were scanned using a Fujifilm FLA-5100 phosphorimager. Bands were quantified using Fujifilm MultiGauge software. Relative binding was determined by dividing the experimental band in each lane by the P9-10E control, then dividing the value of each 'bound' lane by the value of input. Final corrections were made by dividing by the value of the positive control parent clone.

References

Akten, B., Kye, M.J., Hao, L.T., Wertz, M.H., Singh, S., Nie, D.Y., Huang, J., Merianda, T.T., Twiss, J.L., Beattie, C.E., *et al.* (2011). Interaction of survival of motor neuron (SMN) and HuD proteins with mRNA cpg15 rescues motor neuron axonal deficits. Proceedings of the National Academy of Sciences of the United States of America *108*, 10337-10342.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J.Y., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research *37*, W202-W208.

Bertrandy, S., Burlet, P., Clermont, O., Huber, C., Fondrat, C., Thierry-Mieg, D., Munnich, A., and Lefebvre, S. (1999). The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. Human Molecular Genetics *8*, 775-782.

Bhardwaj, A., Myers, M.P., Buratti, E., and Baralle, F.E. (2013). Characterizing TDP-43 interaction with its RNA targets. Nucleic Acids Research *41*, 5062-5074.

Cammas, A., Pileur, F., Bonnal, S., Lewis, S.M., Leveque, N., Holcik, M., and Vagner, S. (2007). Cytoplasmic relocalization of heterogeneous nuclear ribonucleoprotein A1 controls translation initiation of specific mRNAs. Molecular Biology of the Cell *18*, 5048-5059.

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*, 1077-1089.

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

Chong, S.R., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., *et al.* (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene *192*, 271-281.

Dember, L.M., Kim, N.D., Liu, K.Q., and Anderson, P. (1996). Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities. Journal of Biological Chemistry *271*, 2783-2788.

Dubey, A.K., Baker, C.S., Romeo, T., and Babitzke, P. (2005). RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. Rna-a Publication of the Rna Society *11*, 1579-1587.

Ellington, A.D., and Szostak, J.W. (1990). INVITRO SELECTION OF RNA MOLECULES THAT BIND SPECIFIC LIGANDS. Nature *346*, 818-822.

Fallini, C., Rouanet, J.P., Donlin-Asp, P.G., Guo, P., Zhang, H.L., Singer, R.H., Rossoll, W., and Bassell, G.J. (2014). Dynamics of Survival of Motor Neuron (SMN) Protein Interaction with the

mRNA-Binding Protein IMP1 Facilitates Its Trafficking into Motor Neuron Axons. Developmental Neurobiology *74*, 319-332.

Fallini, C., Zhang, H.L., Su, Y.H., Silani, V., Singer, R.H., Rossoll, W., and Bassell, G.J. (2011). The Survival of Motor Neuron (SMN) Protein Interacts with the mRNA-Binding Protein HuD and Regulates Localization of Poly(A) mRNA in Primary Motor Neuron Axons. Journal of Neuroscience *31*, 3914-3925.

GhisolfiNieto, L., Joseph, G., PuvionDutilleul, F., Amalric, F., and Bouvet, P. (1996). Nucleolin is a sequence-specific RNA-binding protein: Characterization of targets on pre-ribosomal RNA. Journal of Molecular Biology *260*, 34-53.

Hua, Y.M., and Zhou, J.H. (2004). Survival motor neuron protein facilitates assembly of stress granules. Febs Letters *572*, 69-74.

Jayasena, S.D. (1999). Aptamers: An emerging class of molecules that rival antibodies in diagnostics. Clinical Chemistry *45*, 1628-1650.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664-U681.

Keenan, R.J., Freymann, D.M., Stroud, R.M., and Walter, P. (2001). The signal recognition particle. Annual Review of Biochemistry *70*, 755-775.

Lau, C.K., Bachorik, J.L., and Dreyfuss, G. (2009). Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains. Nature Structural & Molecular Biology *16*, 486-491.

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., and Zeviani, M. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80*, 155-165.

Lorson, C.L., and Androphy, E.J. (1998). The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. Human Molecular Genetics *7*, 1269-1275.

Lorson, C.L., Hahnen, E., Androphy, E.J., and 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*, 6307-6311.

Lotti, F., Imlach, W.L., Saieva, L., Beck, E.S., Hao, L.T., Li, D.K., Jiao, W., Mentis, G.Z., Beattie, C.E., McCabe, B.D.*, et al.* (2012). An SMN-Dependent U12 Splicing Event Essential for Motor Circuit Function. Cell *151*, 440-454.

Ma, A.S.W., Moran-Jones, K., Shan, J., Munro, T.P., Snee, M.J., Hoek, K.S., and Smith, R. (2002). Heterogeneous nuclear ribonucleoprotein A3, a novel RNA trafficking response element-binding protein. Journal of Biological Chemistry *277*, 18010-18020.

Meister, G., Buhler, D., Pillai, R., Lottspeich, F., and Fischer, U. (2001). A multiprotein complex

mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nature Cell Biology *3*, 945-949.

Mitchell, S.F., and Parker, R. (2014). Principles and Properties of Eukaryotic mRNPs. Molecular Cell *54*, 547-558.

Pellizzoni, L., Baccon, J., Charroux, B., and Dreyfuss, G. (2001). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. Current Biology *11*, 1079-1088.

Pellizzoni, L., Baccon, J., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002a). Purification of native survival of motor neurons complexes and identification of Gemin6 as a novel component. Journal of Biological Chemistry *277*, 7540-7545.

Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002b). Essential role for the SMN complex in the specificity of snRNP assembly. Science *298*, 1775-1779.

Piazzon, N., Schlotter, F., Lefebvre, S., Dodre, M., Mereau, A., Soret, J., Besse, A., Barkats, M., Bordonne, R., Branlant, C.*, et al.* (2013). Implication of the SMN complex in the biogenesis and steady state level of the Signal Recognition Particle. Nucleic Acids Research *41*, 1255-1272.

Rage, F., Boulisfane, N., Rihan, K., Neel, H., Gostan, T., Bertrand, E., Bordonne, R., and Soret, J. (2013). Genome-wide identification of mRNAs associated with the protein SMN whose depletion decreases their axonal localization. Rna-a Publication of the Rna Society *19*, 1755-1766.

Rossoll, W., Jablonka, S., Andreassi, C., Kroning, A.K., Karle, K., Monani, U.R., and 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*, 801-812.

Rossoll, W., Kroning, A.K., Ohndorf, U.M., Steegborn, C., Jablonka, S., and Sendtner, M. (2002). Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons? Human Molecular Genetics *11*, 93-105.

Sanchez, G., Dury, A.Y., Murray, L.M., Biondi, O., Tadesse, H., El Fatimy, R., Kothary, R., Charbonnier, F., Khandjian, E.W., and Cote, J. (2013). A novel function for the survival motoneuron protein as a translational regulator. Human Molecular Genetics *22*, 668-684.

Schwartz, J.C., Wang, X.Y., Podell, E.R., and Cech, T.R. (2013). RNA Seeds Higher-Order Assembly of FUS Protein. Cell Reports *5*, 918-925.

Singh, N.N., Seo, J.B., Ottesen, E.W., Shishimorova, M., Bhattacharya, D., and Singh, R.N. (2011). TIA1 Prevents Skipping of a Critical Exon Associated with Spinal Muscular Atrophy. Molecular and Cellular Biology *31*, 935-954.

Singh, R.N., Saldanha, R.J., D'Souza, L.M., and Lambowitz, A.M. (2002). Binding of a group II

intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. Journal of Molecular Biology *318*, 287-303.

Skrisovska, L., Bourgeois, C.F., Stefl, R., Grellscheid, S.N., Kister, L., Wenter, P., Elliott, D.J., Stevenin, J., and Allain, F.H.T. (2007). The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. Embo Reports *8*, 372-379.

So, B.R., Wan, L.L., Zhang, Z.X., Li, P.L., Babiash, E., Duan, J.Q., Younis, I., and Dreyfuss, G. (2016). A U1 snRNP-specific assembly pathway reveals the SMN complex as a versatile hub for RNP exchange. Nature Structural & Molecular Biology *23*, 225-230.

Stoltenburg, R., Reinemann, C., and Strehlitz, B. (2007). SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands. Biomolecular Engineering *24*, 381-403.

Strasswimmer, J., Lorson, C.L., Breiding, D.E., Chen, J.J., Le, T., Burghes, A.H.M., and Androphy, E.J. (1999). Identification of survival motor neuron as a transcriptional activator-binding protein. Human Molecular Genetics *8*, 1219-1226.

Tacke, R., and Manley, J.L. (1995). THE HUMAN SPLICING FACTORS ASF/SF2 AND SC35 POSSESS DISTINCT, FUNCTIONALLY SIGNIFICANT RNA-BINDING SPECIFICITIES. Embo Journal *14*, 3540-3551.

Takagaki, Y., and Manley, J.L. (1997). RNA recognition by the human polyadenylation factor CstF. Molecular and Cellular Biology *17*, 3907-3914.

Tuerk, C., and Gold, L. (1990). SYSTEMATIC EVOLUTION OF LIGANDS BY EXPONENTIAL ENRICHMENT - RNA LIGANDS TO BACTERIOPHAGE-T4 DNA-POLYMERASE. Science *249*, 505-510.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell *136*, 701-718.

Will, C.L., and Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Current Opinion in Cell Biology *13*, 290-301.

Wirth, B., Brichta, L., and Hahnen, E. (2006). Spinal muscular atrophy: from gene to therapy. Seminars in pediatric neurology *13*, 121-131.

Wong, I., and Lohman, T.M. (1993). A DOUBLE-FILTER METHOD FOR NITROCELLULOSE-FILTER BINDING - APPLICATION TO PROTEIN NUCLEIC-ACID INTERACTIONS. Proceedings of the National Academy of Sciences of the United States of America *90*, 5428-5432.

Workman, E., Veith, A., and Battle, D.J. (2014). U1A Regulates 3 ' Processing of the Survival Motor Neuron mRNA. Journal of Biological Chemistry *289*, 3703-3712.

Yong, J., Kasim, M., Bachorik, J.L., Wan, L.L., and Dreyfuss, G. (2010). Gemin5 Delivers

snRNA Precursors to the SMN Complex for snRNP Biogenesis. Molecular Cell *38*, 551-562.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. Cell *133*, 585-600.

Zhao, D.Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z.Y., Schmitges, F.W., Zhong, G.Q., Liu, K., Li, W.G., Moffat, J.*, et al.* (2016). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature *529*, 48-+.

Zou, T., Yang, X.M., Pan, D.M., Huang, J., Sahin, M., and Zhou, J.H. (2011). SMN Deficiency Reduces Cellular Ability to Form Stress Granules, Sensitizing Cells to Stress. Cellular and Molecular Neurobiology *31*, 541-550.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research *31*, 3406-3415.

Figure Legends

Figure 1. Overview of SELEX. (A) Oligonucleotide used for selection. Numbering is relative to the first transcribed base. T7 promoter is highlighted in red, variable region is highlighted in blue. Constant regions for primer binding are indicated. **(B)** Overview of the selection process. Briefly, RNA is transcribed from double-stranded PCR fragments, then mixed with SMN protein and passed through a nitrocellulose filter. Unbound RNA is washed away and bound RNA is then extracted, converted to cDNA, and amplified to provide a template for the next round of selection.

Figure 2. Binding of SMN to individual clones from unselected and intermediate pools. (A) Top panel: Sequences of 10 transcripts randomly selected from unselected pool are given in the top panel. Boxed region indicates randomized region. Base positions are given above. Bottom panel: Quantification of binding of SMN to transcripts shown in the top panel. Y axis indicates percentage of RNA bound by SMN as measured by the nitrocellulose filter paper binding assay. **(B)** Top panel: Sequences of 9 transcripts randomly selected from pool 6. Bottom panel: Quantification of binding of SMN to transcripts from 10 randomly selected clones generated from pool 6. Labeling is the same as in **(A)**.

Figure 3. Properties of 5 selected clones with high binding affinities. (A) Sequences of the 5 highest-affinity molecules. Clone number is given on the left, base position is indicated on top. The selected region is indicated with a green box. Yellow highlighted

bases are exact matches for the GUGCG motif identified by MEME. **(B-G)** Binding curves to determine K_d of the top 5 binders as well as an unselected control. Clone designation is indicated at the top left of each graph. Y axis indicates relative binding, corrected for the highest value observed during each set of experiments. X axis indicates the SMN concentration for each reaction, on a log10 scale. Each point represents one binding reaction, lines are drawn by Kaleidagraph software by fitting the individual points to the Michaelis-Menten equation. Apparent K_d and r^2 values are given **(H)** Predicted structures of top 5 selected sequences. Structures were predicted by Mfold. Clone numbers are indicated at the top left of each box. 5′ and 3′ ends are indicated, as are base positions. The selected region is indicated by blue bases, constant region is in black. Bases in bold indicate GUGCG motif. Base pairs are indicated by lines drawn between bases.

Figure 4. Comparative binding of sequence mutants of clone P9-10. (A) Effect of extensive sequence mutations within the selected region of P9-10. Top panel gives names and sequence for each clone tested. Base position is indicated at the top. Black sequences indicate constant regions, blue sequences indicate bases that match the WT selected sequence, and red sequences indicate bases that differ from WT. Dashes indicate deleted bases. Lower panel shows autoradiogram of bound RNA. Sequence number is given at the top. Upper band is an extended version of the WT sequence used as a control. Lower band is the WT, unselected, or mutated sequence. Each lane is loaded to obtain equal band intensity of the upper band. I: RNA Input. B: Extracted and purified RNA after binding reaction. Lane numbers and percent activity relative to WT are given below each

autoradiogram. **(B)** Effect of smaller sequence mutations within the selected region of P9-10. Labeling is the same as in **(A)**.

Figure 5. Effect of mutations within constant regions of P9-10. Top panel gives names and sequence for each clone tested. Base position is indicated at the top. Black sequences indicate constant regions, blue sequences indicate bases that match the WT selected sequence, and red sequences indicate bases that differ from WT. Dashes indicate deleted bases. Predicted structures of P9-10 and all constant region mutants are shown. Each structure indicates the location within the P9-10 predicted structure of individual mutations made in the flanking regions of P9-10. The selected region is indicated by blue bases, constant region is in black. Mutated bases are shown in red. WT base pairs are shown as black lines black and novel base pairs formed by mutated bases as red lines. Lower panel shows autoradiogram of bound RNA. Sequence number is given at the top. Upper band is an extended version of the WT sequence used as a control. Lower band is the WT, unselected, or mutated sequence. Each lane is loaded to obtain equal band intensity of the upper band. I: RNA Input. B: Extracted and purified RNA after binding reaction. Lane numbers and percent activity relative to WT are given below each autoradiogram.

Figure 6. Comparative binding of 5′ and 3′ truncations of clone P9-10. (A) 3′ truncations of P9-10. Top panel gives names and sequence for each clone tested. Base position is indicated at the top. Black sequences indicate constant regions, blue sequences indicate bases that match the WT selected sequence, and red sequences indicate bases

that differ from WT. Dashes indicate deleted bases. Lower panel shows autoradiogram of bound RNA. Sequence number is given at the top. Upper band is an extended version of the WT sequence used as a control. Lower band is the WT, unselected, or mutated sequence. Each lane is loaded to obtain equal band intensity of the upper band. I: RNA Input. B: Extracted and purified RNA after binding reaction. Lane numbers and percent activity relative to WT are given below each autoradiogram. **(B)** 5′ truncations of P9-10. Labeling and abbreviations are the same as in **(A)**.

Figure 7. Effect of salt conditions on binding of SMN to RNA. (A) Comparative binding of WT and mutated P9-10 RNA, as well as an unselected clone. Top panel gives names and sequence for each clone tested. Base position is indicated at the top. Black sequences indicate constant regions, blue sequences indicate bases that match the WT selected sequence, and red sequences indicate bases that differ from WT. Dashes indicate deleted bases. Middle panel and lower panel are autoradiograms of bound sequences. Sequence number is given at the top. The presence (+) or absence (-) of SMN protein in the reaction is indicated. The NaCl concentration of each reaction, in millimoles per liter, is indicated. Upper band is an extended version of the WT sequence used as a control. Lower band is the WT, unselected, or mutated sequence. Each lane is loaded to obtain equal band intensity of the upper band. I: RNA Input. B: Extracted and purified RNA after binding reaction. Lane numbers are given below each autoradiogram. **(B)** Quantification of bands from **(A)** Top row of graphs represents the total proportion bound relative to input. The X axis indicates the binding condition. The bottom row of graphs represents the relative binding of the lower band from **(A)** compared to the binding of the

constant control sequence. X axis indicates the NaCl concentration in the binding reactions. Since binding at 200 mM of NaCl and above is near background levels for all bands, only values for 150 mM and below are given.

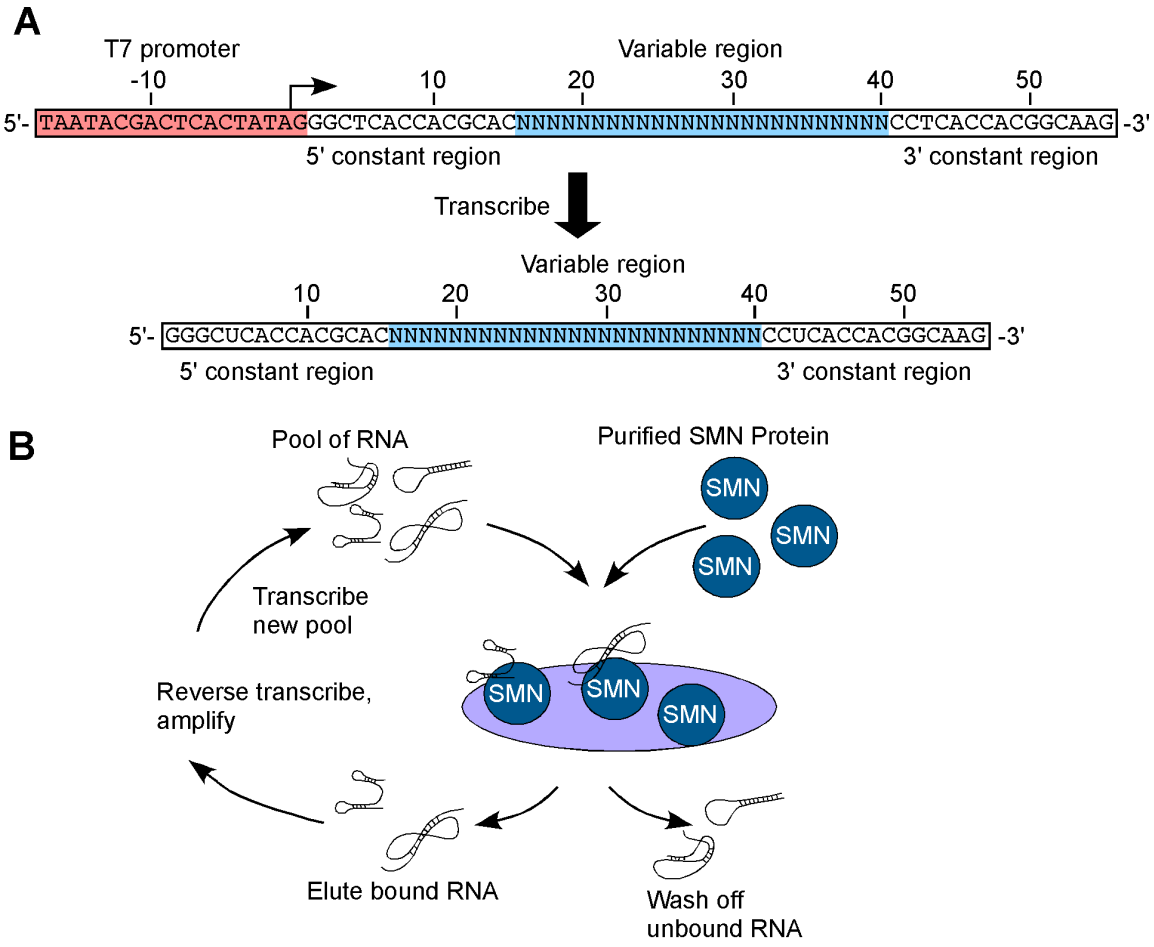


Figure 1. Overview of SELEX

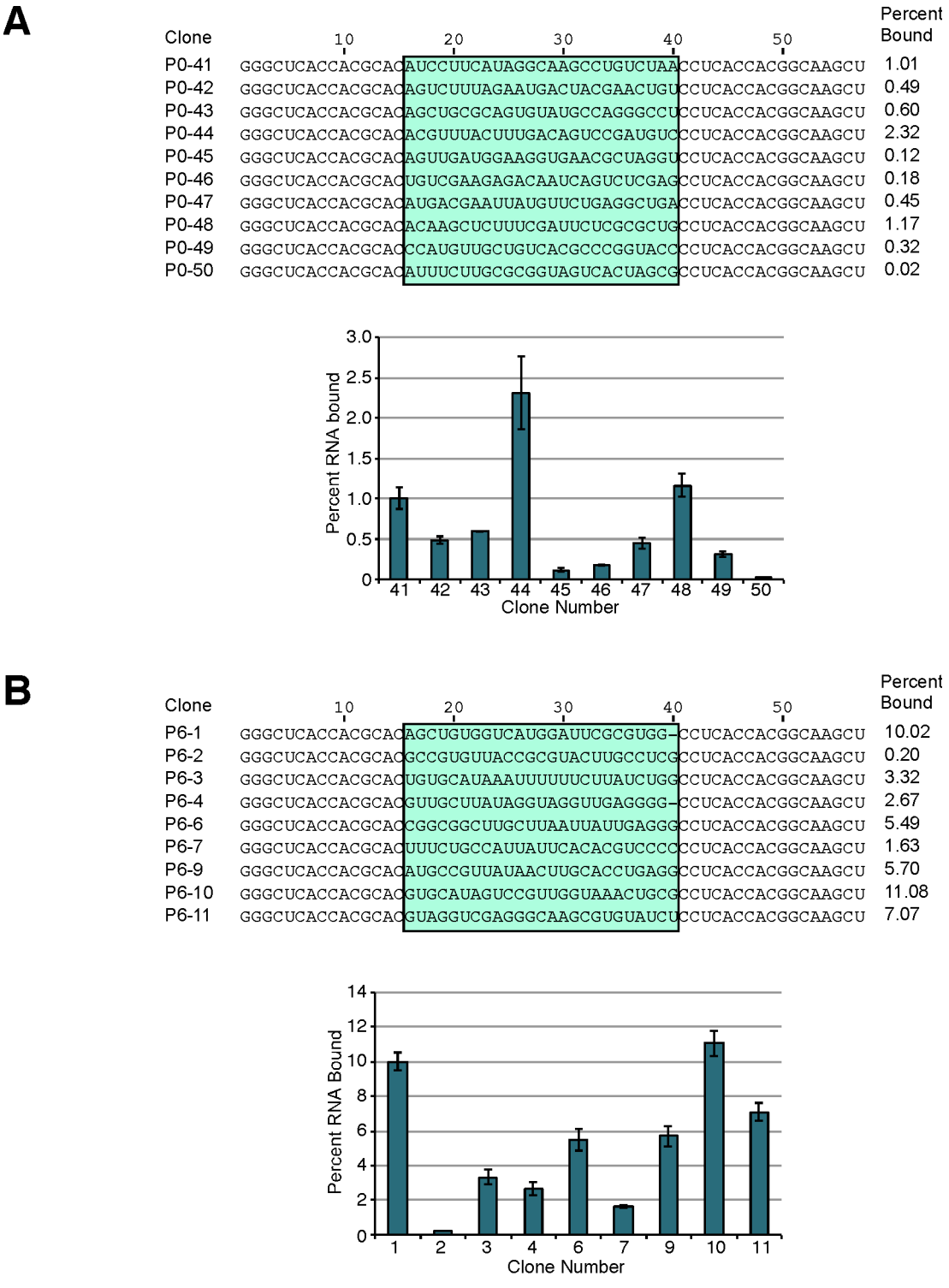


Figure 2. Binding of SMN to individual clones from unselected and intermediate pools

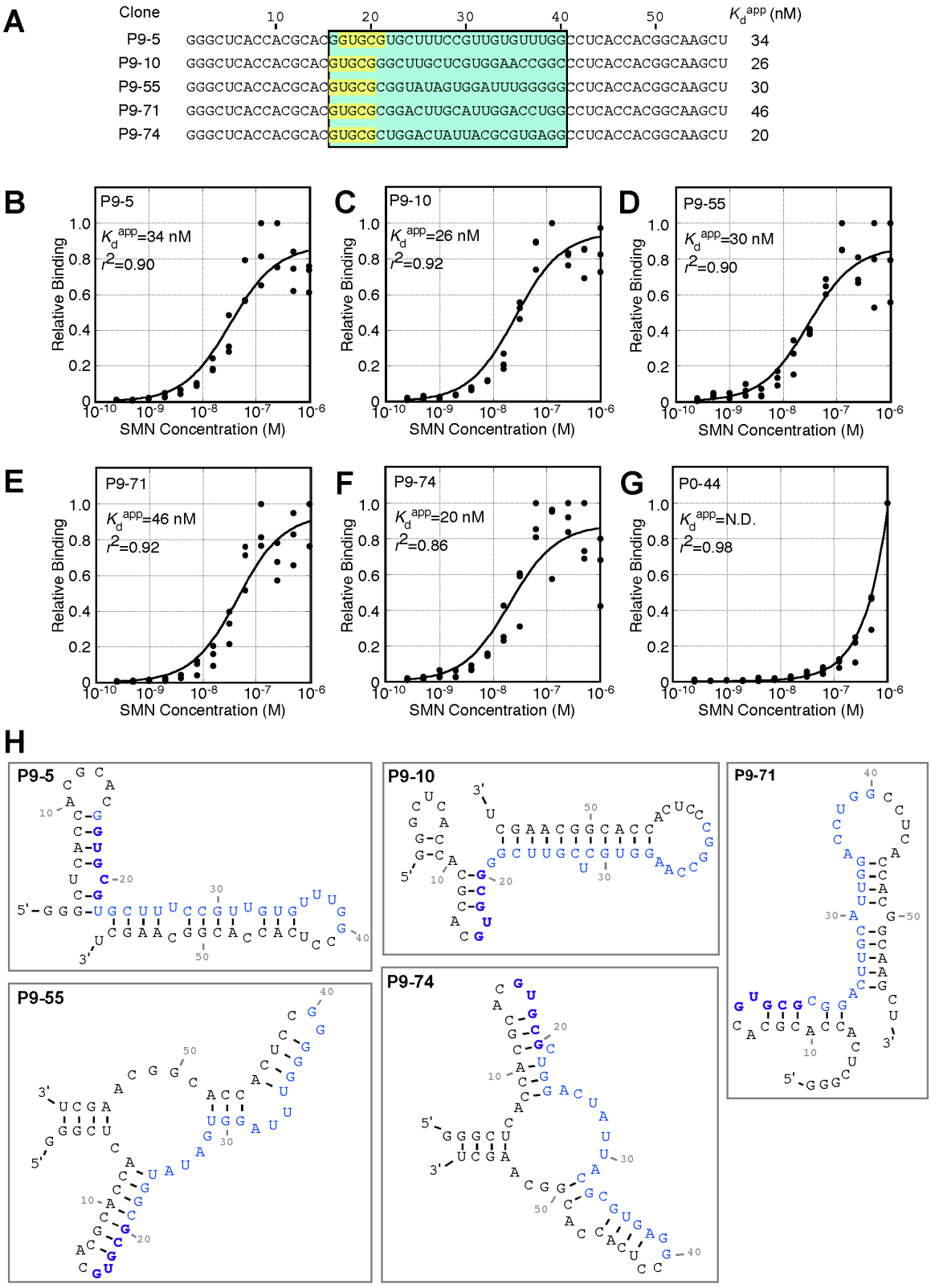


Figure 3. Properties of 5 selected clones with high binding affinities

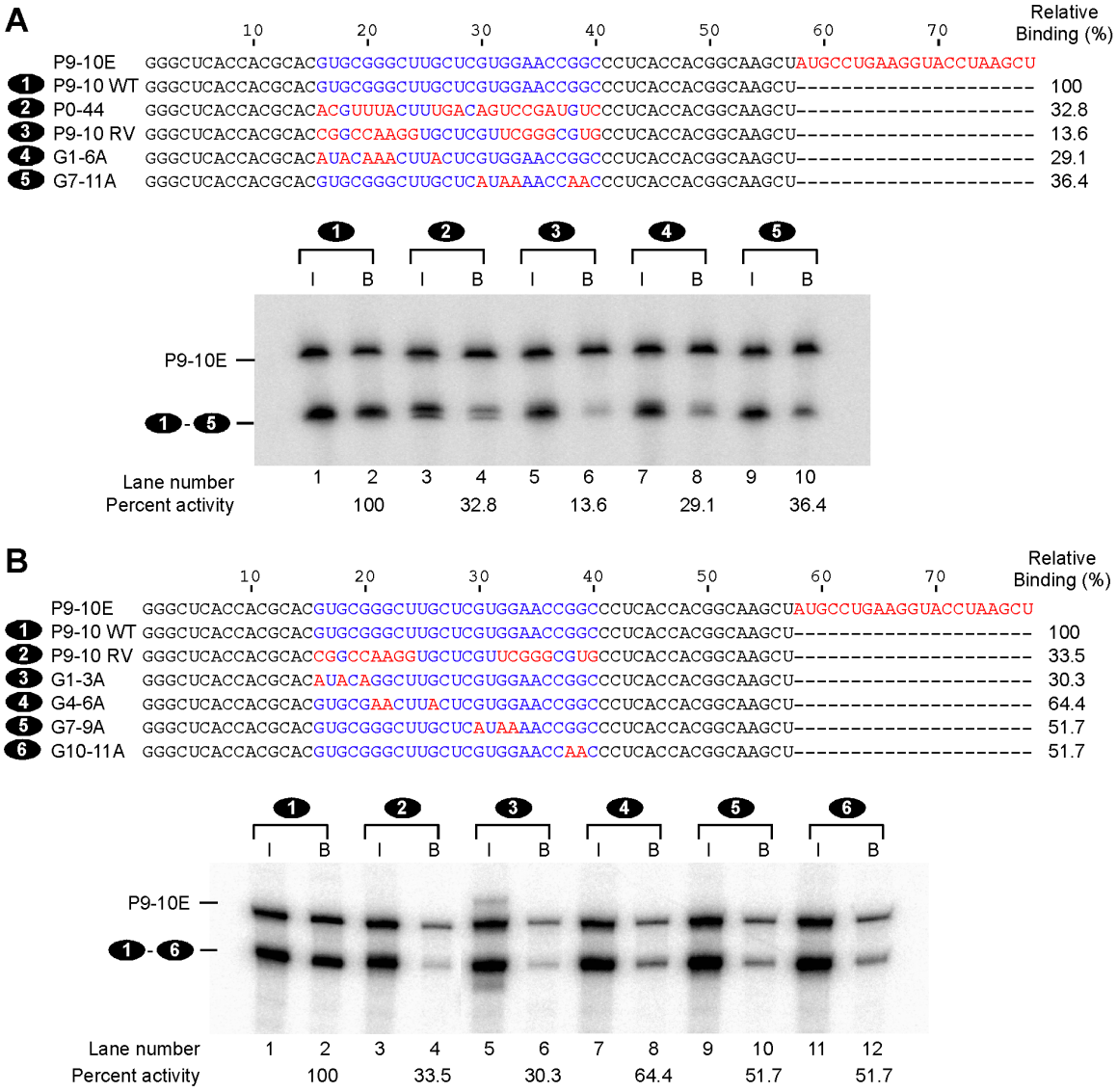


Figure 4. Comparative binding of sequence mutants of clone P9-10

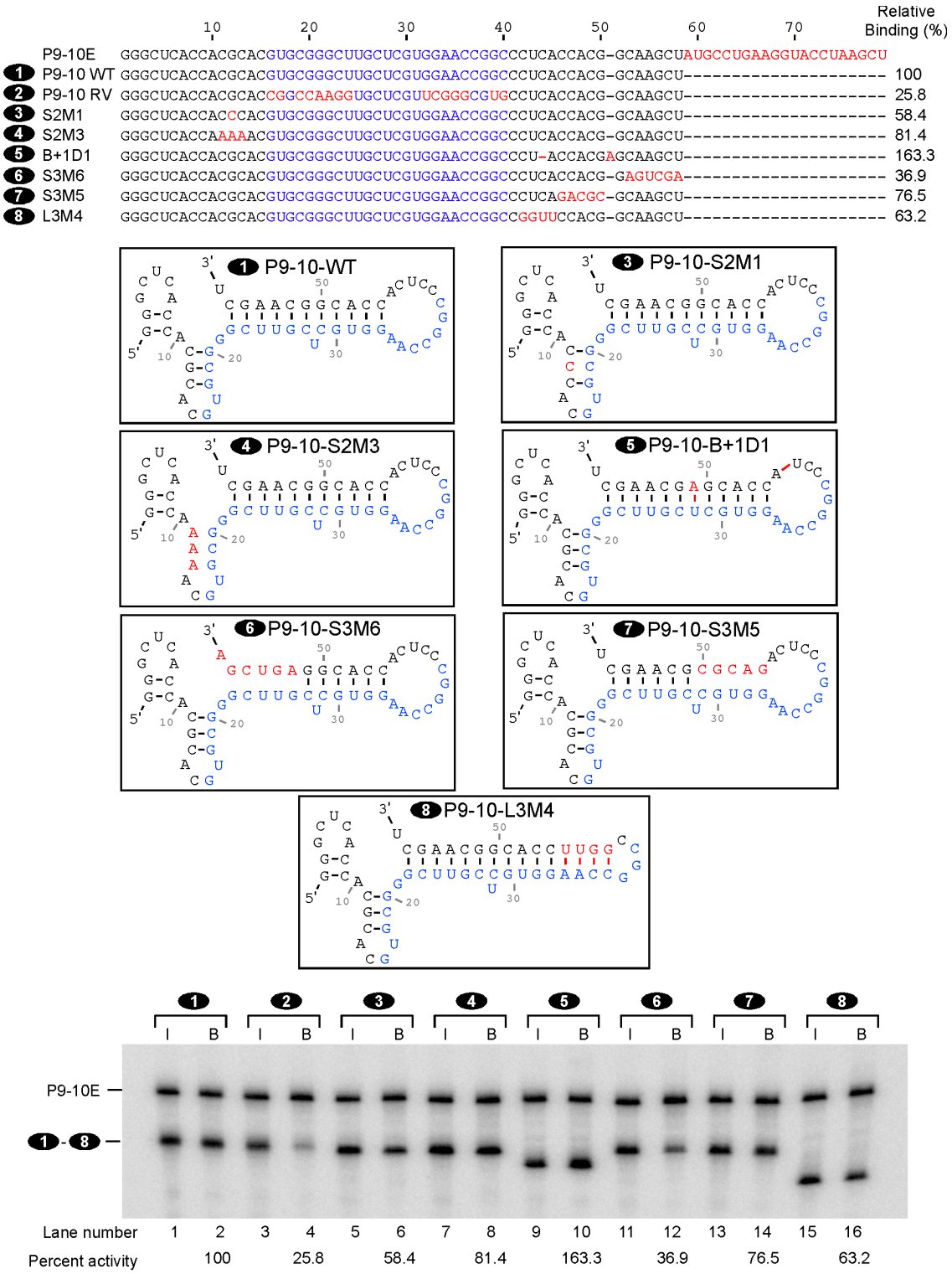


Figure 5. Effect of mutations within constant regions of P9-10

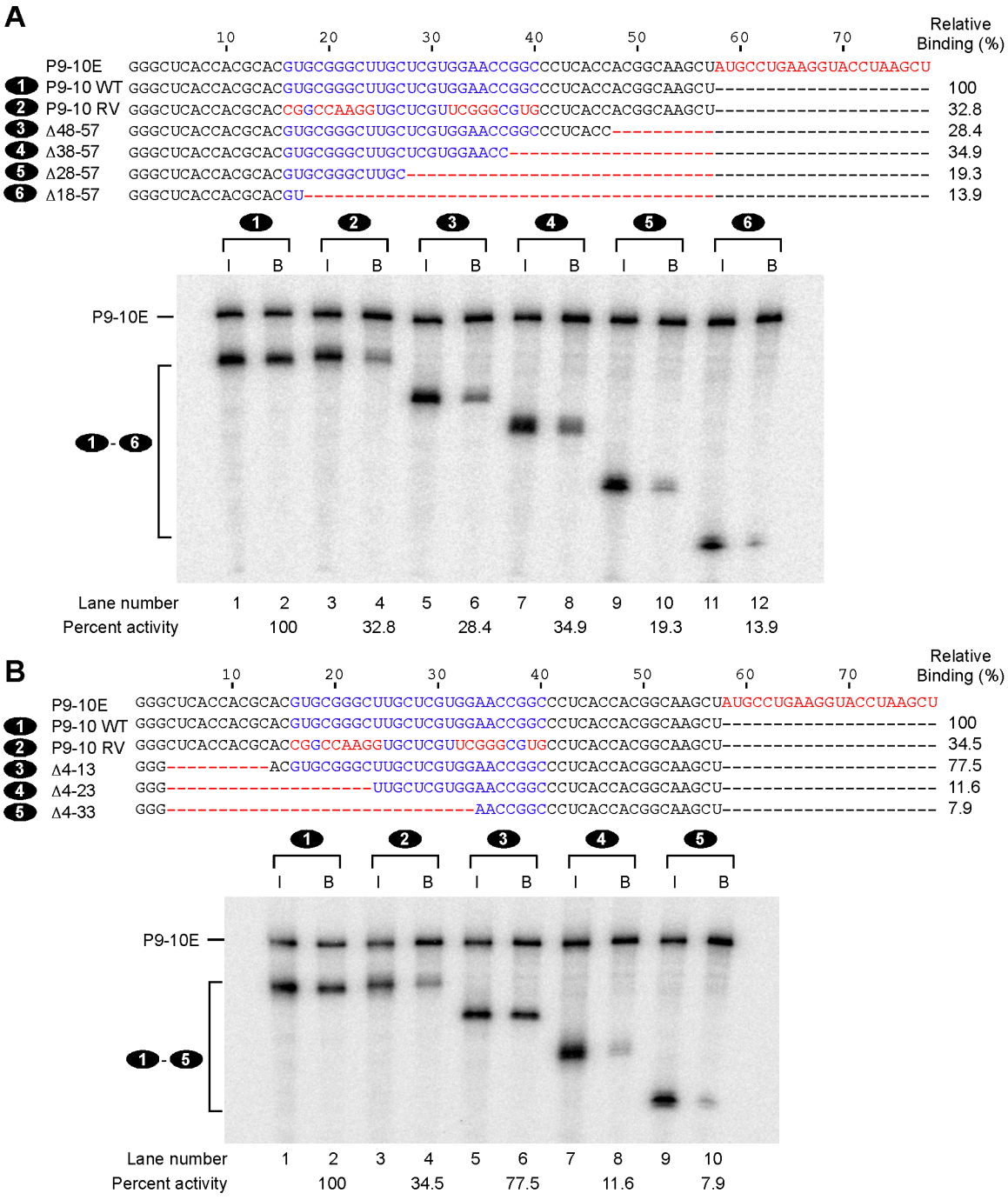


Figure 6. Comparative binding of 5′ and 3′ truncations of clone P9-10

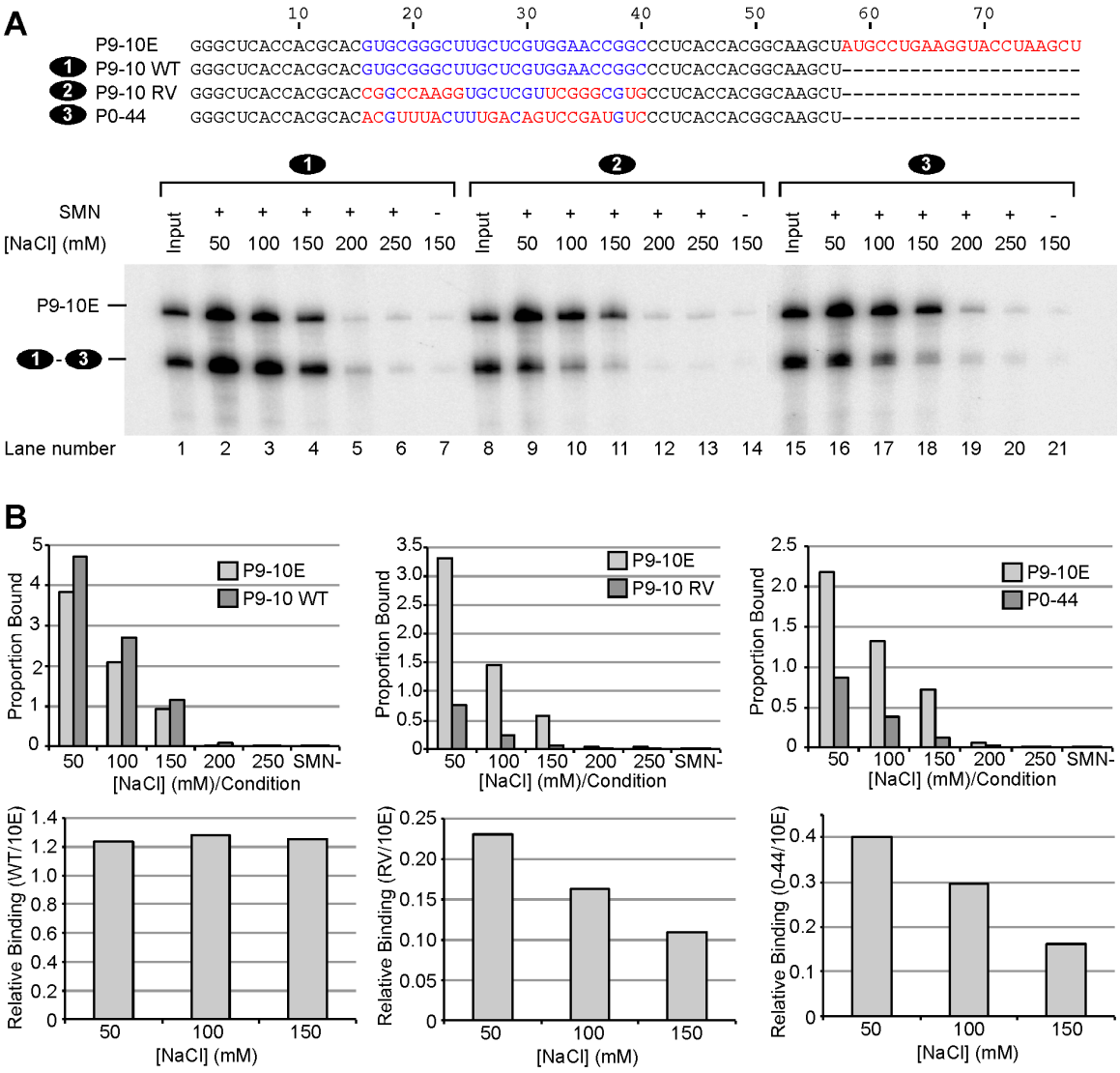


Figure 7. Effect of salt conditions on binding of SMN to RNA

Table 1: Sequences and base composition of initial unselected pool

Clone Number	Randomized Region	Base Occurrence			
		A	U	C	G
P0-1	A C G G G A G U G U U A G G A G G C A G C C A C G	6	3	5	11
P0-2	U G G C G G A G G C G G C G A U A C U C C G U A U	4	5	6	10
P0-3	G U G U A G A U G C C G A U U C G U C A U C G U A	5	8	5	7
P0-4	A U U G G A G G G U C C A G A G G A G G U C A U C	6	5	4	10
P0-5	U G C G C U G U A G U C C U A U U G U C G U G C U	2	10	6	7
P0-6	A U G C C C U U G U C C U A A G G G U U G G C C U	3	8	7	7
P0-7	G G C C C G U A A C U G A A U C C U A G G G G C U	5	5	7	8
P0-8	G G U U A C G A A C U A U U G U G U G A C A U G C	6	8	4	7
P0-9	A G C C U U U U G C U U G U G A A U A U A U A G C	6	10	4	5
P0-10	U A C A G C U C U A U C C C U G U C C A A A U G G	6	7	8	4
P0-11	U U G U U C U A G C C C U A U C U C U A A G G C U G	3	10	7	5
P0-12	A G G A C G A A C C A G G A U C U G C U A G G G U	7	4	5	9
P0-13	G C G A C G C A A C A U G A G A G G C G G C G U A	7	2	6	9
P0-14	A G C C U G U A A A C U G A U G G C C U G G G C G	5	5	6	9
P0-15	C U A C G G G U C U G C G A G U U G C G C G U C A	3	6	7	9
P0-16	G C G U C C A A C G C C C A U A G U G G G C C G U	4	4	9	8
P0-17	C G U C A G G C C G U U A U A G G G G C U A U C	4	6	6	9
P0-18	C A A U U G U A A U A A G U C G U U A G U C A C U	8	9	4	4
P0-19	U A C G A G C G G U U G G C U G G A U G U G G G U	3	7	3	12
P0-20	A G A U U C U A A A G A U G U U A C U C C U C C U	7	9	6	3
P0-21	U C A C A G C A G A A G A C G A A U G C C U G A A	10	3	6	6
P0-22	G U A C U C C G G A G U A C U G U C G U U G G A G	4	7	5	9
P0-23	A G C C U C C U A C U A A C U G G U C U U C U U G	4	9	8	4
P0-24	G U A A U G G G G A C A U G G G U C C U G U C A G	5	6	4	10
P0-25	U U U G A C U A C U A G U C U C G G C G C U U G A	4	9	6	6
P0-26	A A A U G G A C C U A C C G A A U C A C G U C C A	9	4	8	4
P0-27	G G G A C G C C U C G C G G A G C C A U G U G U C	3	4	8	10
P0-28	G G A G G A C G U U G C U C A G A C G C G A U A U	6	5	5	9
P0-29	G G A G U G G G A C C A U U G A C A U U C C C U U	5	7	6	7
P0-30	G A A A G C C G G G U C G U C G G U U U C C C U G	3	6	7	9
P0-31	C U A G A A C U A G G C U G U U U G A U C G G G C	5	7	5	8
P0-32	U C C A C U A U A G G A A U C G U U G C U U G U G	5	9	5	6
P0-33	A U G A A G A U C A G U U G U C C C U G G G C C G U	5	7	5	8
P0-34	U A A G C C G G U C A G A A U G G A G G U U G G G	6	5	3	11
P0-35	G G G A G C G U A C U G U C C G C G C A A C U C G	4	4	8	9
P0-36	G A G C G G C G G A A G U A C U C A A A U C A G C	8	3	6	8
P0-37	C U U C U G U G U U G G A U G A C C C U A C C C C	3	8	9	5
P0-38	C G U C U C U U G G G A U C G U U G U U C C G C G	1	9	7	8
P0-39	G G G C A A G U C C C A U G A G A U C A A C A G U	8	4	6	7
P0-40	A U C G G G C U G A C U G U U C A G G U A A C A G	6	6	5	8
P0-41	A U C C U U C A U A G G C A A G C C U G U C U A A	7	7	7	4
P0-42	A G U C U U U A G A A A U G A C U A C G A A C U G U	8	8	4	5
P0-43	A G C U G C G C A G U G U A A U G C C A G G G C C U	4	5	7	9
P0-44	A C G U U U A C U U U G A C A G U C C G A U G U C	5	9	6	5
P0-45	A G U U G A U G G A A G G U G A A C G C U A G G U	7	6	2	10
P0-46	U G U C G A A G A G A C A A U C A G U C U C G A G	8	5	5	7
P0-47	A U G A C G A A U U A U G U U C U G A G G C U G A	7	8	3	7
P0-48	A C A A G C U C U U U C G A U U C U C G C G C U G	4	8	8	5
P0-49	C C A U G U U G C U G U C A C G C C C G G U A C C	3	6	10	5
P0-50	A U U U C U U G C G C G G U A G U C A C U A G C G	4	8	6	6

Base	Base Occurance by Position																								Overall proportion
A	19	7	14	10	9	7	9	13	13	13	11	10	12	15	12	8	10	5	11	8	8	6	8	8	8	20.8%
U	10	15	10	12	14	8	16	14	9	14	13	11	15	13	20	10	13	13	10	13	13	12	11	12	14	25.8%
C	7	8	10	17	10	15	12	7	9	13	11	13	7	11	7	8	11	19	14	10	13	19	13	14	10	23.6%
G	14	19	15	10	16	19	12	15	17	9	14	15	15	10	10	23	17	12	13	18	14	11	16	14	16	29.8%

Table 2. Conditions used for selection of high-affinity RNA targets of SMN

Pool	[RNA] (nM)	[SMN] (nM)	[NaCl] (mM)	% Binding	Enrichment
P0	100	400	150	0.53%	1.00
P1	100	200	150	1.35%	2.53
P2	100	100	150	2.86%	5.36
P3	100	85	150	5.09%	9.52
P4	100	85	150	8.58%	16.1
P5	100	85	150	14.9%	27.9
P6	100	8.5	150	16.0%	30.0
P7	100	85	150		
P8	100	85	150		
P9				17.7%	35.4

Table 3A: High affinity binders from Pool 9

Clone	Randomized/Selected Region	Base Occurrence A	U	C	G	Binding
P9-55*	G U G C G C G G U A U A G U G G A U U U G G G G G – – – – –	3	7	2	13	117%
P9-10	G U G C G G G C U U G C U C G U G G A A C C G G C – – – – –	2	5	7	11	105%
P9-74	G U G C G C U G G A C U A U U A C G C G U G A G G – – – – –	4	6	5	9	104%
P9-71*	G U G C G C G G A C U U G C A U U G G A C C U G G – – – – –	3	6	6	10	101%
P9-5	G G U G C G U G C U U U C C G U U G U G U U U G G – – – – –	0	11	4	10	87%
P9-81*	C G U G C C G G C U U G U C C G U G U U U G G G G – – – – –	0	8	6	11	86%
P9-29	C G C C G G G U G C U U U C C G U U G G U C U G G G – – – – –	0	8	7	10	85%
P9-67*	G U G C G A A G U G G U A U U G C U G C G G U C G – – – – –	3	7	4	11	81%
P9-78	G U G C G U U U G A U U G C U G U A G G U C U G G G – – – – –	2	10	3	10	76%
P9-75	U G U G C G C C U U C C C G U G U A G U A C U G G G – – – – –	2	8	7	8	74%
P9-27*	U G U G C G A U G C C A U G A G G U U G G G C G G G – – – – –	3	6	4	12	73%
P9-63	G U G C U U U G C U U G A C C G G G A A U C G G G G – – – – –	3	7	5	10	73%
P9-48	G G U G G U G G U U U G A C C G U U G G G G U U U U U G U G G G	1	13	2	14	73%
P9-60	G U G C G U U C A U C U U G U C U C G U A C U G G G – – – – –	2	10	6	7	71%
P9-73	G U G C G C U G G U U U U A A U U U U U G C G C G G – – – – –	2	11	4	8	69%
P9-38*	G U G C G A U A A C G G G G G U U A A G A C G G U G G – – – – –	6	5	3	11	68%
P9-79	A U G G G U G G G G G G A A A G U G U G G G U U C U U – – – – –	4	7	1	13	60%
P9-52	G C G C A A A C G G G G A A A A A A U G U G A G G G G – – – – –	9	2	3	11	59%
P9-49	C G U G C G U U U U U U U G U C G U G C U G C U G G G – – – – –	0	12	5	9	59%
P9-59	G U G G C C U G G G G G U A A U A G C G U G U U A C A A – – – – –	5	6	4	10	58%
P9-32	G U G U G U G G A U U U U U U A G C G U U U A G G G G G G – – – – –	3	10	1	11	57%
P9-28	C G U G G G U C G A U U U U U U G C G G U U C G U G U G G – – – – –	1	11	4	9	53%
P9-35*	G U U U A G G U U G A G G U A U U G C G U G U U U A A – – – – –	4	11	1	9	49%
P9-69	G G G U U U G C G G C U U G C A C G G U U A G G G G G – – – – –	2	6	4	13	47%
P9-26	C U G G U U G G G G G C A A A U U G G A G U G G G C C C G – – – – –	4	5	5	11	47%
P9-41	G U G C G U U G C U G G G U U A U G G U G C G U G – – – – –	1	9	3	12	46%
P9-14	U U G G C G C G G G C U U G U U A U G G A A U G G – – – – –	3	8	3	11	45%
P9-31	C G U G G U U A A A U U U U A A G C G U U G U U G G – – – – –	5	10	2	8	44%
P9-30	A G G U G G G G A U U C G U G G C G U G A C G A U – – – – –	4	6	3	12	44%
P9-22	G U G C G G C C U A G U G A G G A A G U G C G G A – – – – –	5	4	4	12	42%
P9-1	G U G C G G A G U U U U U G A G G U U G U G C G G – – – – –	2	9	2	12	41%
P9-58	G U G C G C U U G C U U G A U G A A U U U C G U A C – – – – –	4	9	5	7	40%
P9-2	G U G C G C G U U U U U A G G U G G G G G A U G G – – – – –	2	8	2	13	40%
P9-36	A G U G C G U G G A G G U A U G C G U G G G U C C – – – – –	3	6	4	12	36%
P9-80	C G U G C U G G G A A A U U U G U U G U G G C C G – – – – –	3	8	4	9	35%
P9-47*	G U G C G C G C U A A G U G A G G A U C G G U G C – – – – –	4	5	5	11	34%
P9-17	U U G C C A G U U U C A U U G C C A U G U A U G G G G – – –	4	10	5	8	34%
P9-53	G U G C G A C C U G U C C G U U U G G U G C G G G – – – – –	1	7	6	11	32%
P9-3	A G U G C G C G U G U U U G U U C G U A G G U C G – – – – –	2	9	4	10	31%
P9-19	A G U A C A A G G A C G A G G A U A U G U G U G G – – – – –	8	5	2	10	27%
P9-37	G U G C G G C C A A A G A G G G U U G C G U U C C – – – – –	4	5	6	10	27%
P9-23	G U G G C A U G U U A C U U U U U A U G G G U G G G – – – – –	3	6	6	10	25%
P9-61*	G U G C G U U G U G U U U A U U G G G A U U A C G – – – – –	3	11	2	9	25%
P9-42	G U G C A A C U U G C A U C U C U C C U U C C C U A – – – – –	4	9	9	3	23%
P9-13	C G U G C G U A G G U U U U G U G C U G G U C G G G – – – – –	1	9	4	11	21%

* - Sequences which were observed twice in selected clones

Base	Base Occurance by Position																											Overall
A	5	0	0	1	3	8	4	4	8	10	6	6	12	8	9	8	7	8	2	7	5	4	3	2	4	0	0	0	0	0	11.8%
U	4	28	14	4	3	12	16	9	18	17	21	22	22	14	17	14	18	14	20	14	14	9	20	7	3	1	0	1	0	0	31.4%
C	8	1	1	23	14	9	9	8	4	6	9	5	3	9	5	6	7	5	3	5	4	15	7	9	5	0	0	0	0	0	15.9%
G	28	16	30	17	25	16	16	24	15	12	9	12	8	14	14	17	13	18	20	19	22	17	15	27	32	2	2	0	1	1	40.8%

Table 3B: Low affinity binders from Pool 9

Clone	Randomized/Selected Region	Base Occurrence				Binding
		A	U	C	G	
P9-57	G A C A C A U U G C C C U G G G G A A U U G G U G – – – – –	5	6	5	9	20%
P9-24	G G G G G C G A U U U G U G A G G U U U G C G U G – – – – –	2	8	2	13	19%
P9-21	G U G U G U G G U C U G U G G G A G G A U G G U G – – – – –	2	8	1	14	17%
P9-18	G U G C A U U G C C G U U U U C U C A U G C C A – – – – –	2	10	7	5	16%
P9-72	G U G G U A A U U G U C U A U G G U G A U U G G G – – – – –	4	10	1	10	15%
P9-46	G U G U G U C G G A G G G U U U G G C U U G C G C – – – – –	1	8	4	12	15%
P9-43	U G U U G C G U A G A A A U U C G U U G G C U U G A – – – –	5	10	3	8	14%
P9-33	U G G G G G C G G U U U G C C G U U U U A U C G C – – – – –	1	9	5	10	13%
P9-50	G U G A A A U G G G G U U G A G G C A C C G C C U – – – – –	5	5	5	10	12%
P9-65	G G G G U G C G U U U U G U C A U G G U U C G G C – – – – –	1	9	4	11	11%
P9-62	C G A U U U U C G C A A A A C A U U U U C C C C A – – – – –	6	8	8	2	10%
P9-56	G G U G G G G A C G A U U G U G C C G G A U U G U – – – – –	3	7	3	12	10%
P9-34*	U U G C C C U U C A U G C C U G G U A C G G G G G – – – – –	2	7	7	9	9.0%
P9-11	U U G G G A U G G G G A U A C U U A G G A G G C G G U A – –	6	7	2	13	8.9%
P9-25	G U G G G C C A A G G U G G G G U A U G C G U G U – – – – –	3	6	3	13	8.8%
P9-51	G U G G G C G U G G U U U U G G U G G G G G C G G – – – – –	0	7	2	16	7.8%
P9-16	G G U G G G G G G G G U U G G C G U G A U U C C U – – – – –	1	7	3	15	7.7%
P9-4	G U G G G C G U G G C U U U U A U U U C G G U G A – – – – –	2	10	3	10	7.5%
P9-66	U A U U G G G G G A U G A A U U G U G G U U G G – – – – –	4	10	0	11	6.4%
P9-64	U U G C U U U U U G G U C C U U G C U U C G C U C C C – – –	0	13	9	5	6.2%
P9-44	G U G C A G U G G U U U A U U U U G C G U U U U G – – – – –	2	13	2	8	4.9%
P9-12	U U G C C A A G A A C U U G C C U U G G C C C G C – – – – –	4	6	9	6	4.8%
P9-68	U U G C C U U G C U U G C C A A A C U U G G G G G – – – – –	3	8	6	8	3.9%
P9-9	A U G G C G G A A A U G C G U C U U U A U A C U C G – – – – –	6	8	5	7	3.4%
P9-15	G U G G G G A G C G G A U G G U U G U G C G A C G U – – – – –	3	6	3	13	3.4%
P9-77	G C G G U G A U C G G A G G A U A G U G C G U C U – – – – –	4	6	4	11	2.8%
P9-76	U G U G C G U U G C G U G G U U U U U U G C U G U – – – – –	0	13	3	9	2.6%
P9-20	G G G G G G C A G G U G G G G U U U U G C G U G A – – – – –	2	6	2	15	2.5%
P9-39	G A C U U C U C U C U U C C A C U C C C U C C C – – – – –	2	8	14	1	2.2%
P9-54	G U G G G C C A A G G U G G G G G U U U G G U G U G – – – – –	2	7	2	14	1.6%
P9-6	G U G A G G A C G G U G A G G C A A U G U G G U A – – – – –	6	5	2	12	1.1%
P9-8	C U C C C C U C C C G C U C C C A C A U C C C U C – – – – –	2	5	17	1	-0.4%
P9-7	A C G G A G U A G G C G C A A A G A G G G G U C G – – – – –	7	2	4	12	-3.3%

Base	Base Occurance by Position																									Overall
A	2	3	1	3	4	6	4	7	5	5	4	4	5	5	5	6	5	5	4	4	4	2	0	0	5	1	0	0	0	0	11.9%
U	9	19	5	6	6	6	13	9	6	5	14	15	13	8	13	10	14	15	16	10	9	8	11	10	6	1	0	0	0	0	31.0%
C	2	2	3	7	7	9	6	5	6	7	4	3	6	6	7	5	4	5	4	5	9	7	12	8	9	1	0	0	0	0	18.0%
G	20	9	24	17	16	12	10	12	16	16	11	11	9	14	8	12	10	8	9	14	11	16	10	15	13	2	0	0	0	0	39.2%

Table 3C: Sequences from Pool 0 for which binding was not tested

Clone	Randomized/Selected Region	Base Occurrence			
		A	U	C	G
P9-40	G U G U U U G G U U U C U U A U U G C G U U G U – – – – –	1	14	3	7
P9-82	U U G C C A A G U U U U U U U G C C A U G G U C C G G – – – –	3	10	6	7
P9-83	A G U G C G U C G U G G U U U U U U G G A G G U C G – – – – –	2	9	3	11
P9-84	U U G C C G C U G C C G U A U G A C U U G A G G A – – – – –	4	7	6	8
P9-85	G U G C G U U G C G U G G A A G A C G G U G C G G – – – – –	3	5	5	12
P9-86	G U G G G C G U G U U A A U U U U G G U G A U G G – – – – –	3	10	1	11
P9-87	A G U G C G U G G A U C U A U G U G C G U U U G G – – – – –	3	9	3	10
P9-88	G U G C U U A A G C U G U G G U C U G U U U G C G U – – – – –	2	10	4	9
P9-89	G U G C G U U U A U U U U C C A C U U U A C A U G – – – – –	4	12	5	4
P9-90	A G G G U G G G G A U G U A G A G U G U G G C A U – – – – –	5	6	1	13

Base	Overall Base Occurance by Position																									Overall
A	10	3	1	4	7	15	10	12	14	17	10	11	18	17	16	16	14	14	6	12	10	8	4	3	10	1	0	0	0	0	11.9%
U	15	54	21	11	12	22	33	21	26	27	43	39	43	26	34	28	36	32	39	29	27	20	34	19	11	2	0	1	0	0	31.9%
C	10	3	4	35	25	19	16	14	11	15	14	10	9	16	13	12	14	12	9	10	13	23	23	19	14	1	0	0	0	0	16.5%
G	53	28	62	38	44	32	29	41	37	29	21	28	18	29	25	32	24	30	34	37	38	37	27	47	51	5	2	0	1	1	39.8%

Table 4A: Trinucleotide occurrence in selected sequences

Second Nucleotide	First nucleotide A	First nucleotide C	First nucleotide G	First nucleotide U	Third Nucleotide
A	14	13	16	11	A
	8	7	16	12	C
	15	3	26	17	G
	17	17	30	22	U
C	7	15	15	5	A
	7	15	31	22	C
	14	22	77	25	G
	14	19	34	16	U
G	7	11	43	30	A
	5	23	36	102	C
	36	41	63	111	G
	20	68	83	52	U
U	8	6	25	26	A
	8	13	23	25	C
	34	26	139	98	G
	36	36	69	60	U

Table 4B: Trinucleotide occurrence in unselected sequences

Second Nucleotide	First nucleotide A	First nucleotide C	First nucleotide G	First nucleotide U	Third Nucleotide
A	6	11	20	6	A
	13	7	18	12	C
	13	19	21	17	G
	15	12	17	10	U
C	11	13	7	10	A
	7	11	23	12	C
	18	14	24	12	G
	22	25	20	12	U
G	14	16	25	19	A
	19	15	26	15	C
	25	18	23	46	G
	19	22	19	1	U
U	9	25	17	3	A
	17	16	35	10	C
	24	28	13	17	G
	11	15	24	6	U

APPENDIX B

CHAPTER III SUPPLEMENTARY FIGURES AND TABLES

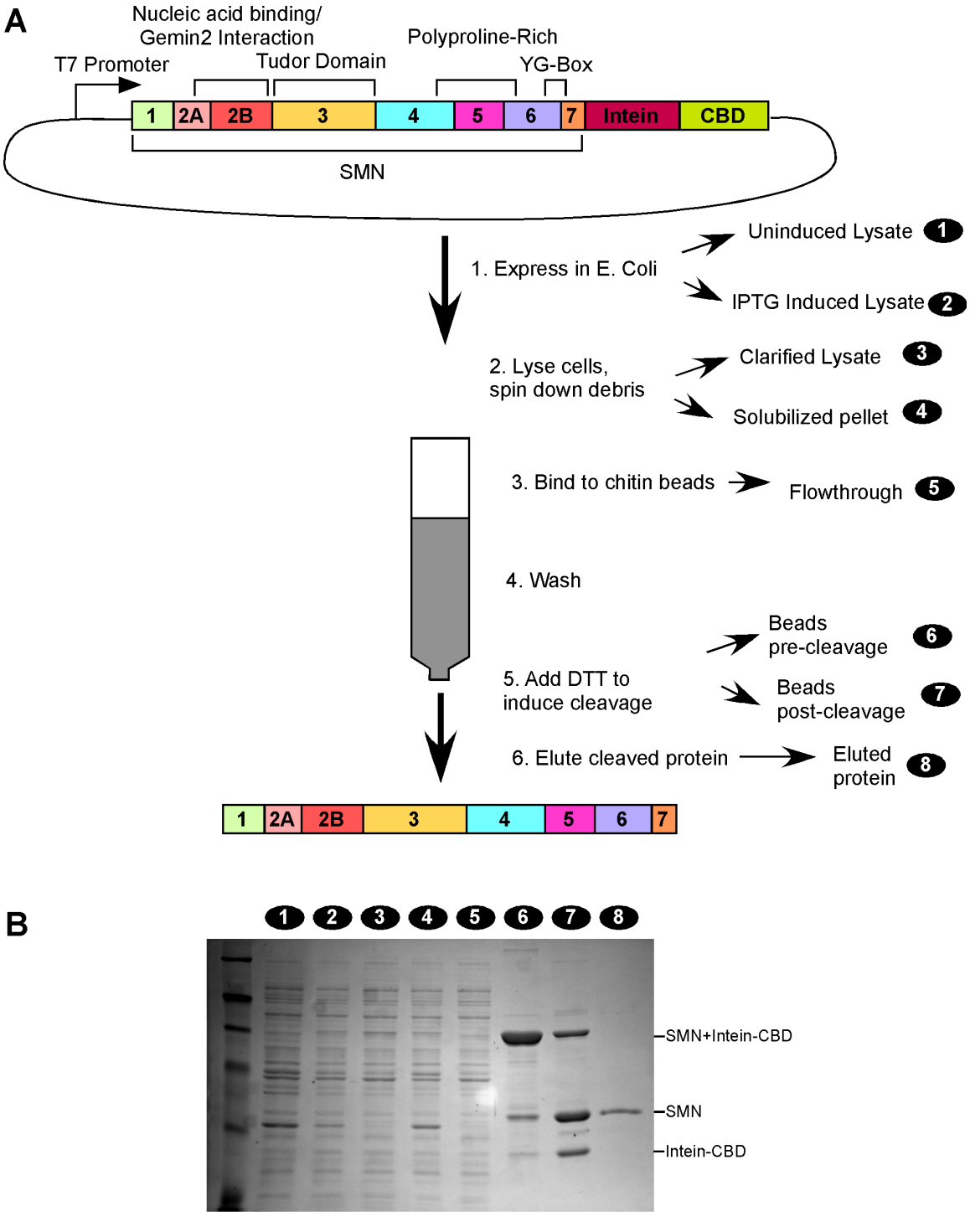
Supplementary Figure Legends

Supplementary Figure 1. Purification of recombinant SMN protein. (A) Purification scheme for IMPACT purification of recombinant SMN. Diagram of expression plasmid is shown. Expression construct contains entire SMN coding sequence followed by a fusion consisting of self-cleaving intein domain and the chitin binding domain. Primary steps of purification are indicated below. **(B)** Diagnostic coomassie-stained SDS-PAGE gel showing aliquots taken at several key steps of the isolation procedure. Final lane shows final, eluted SMN protein.

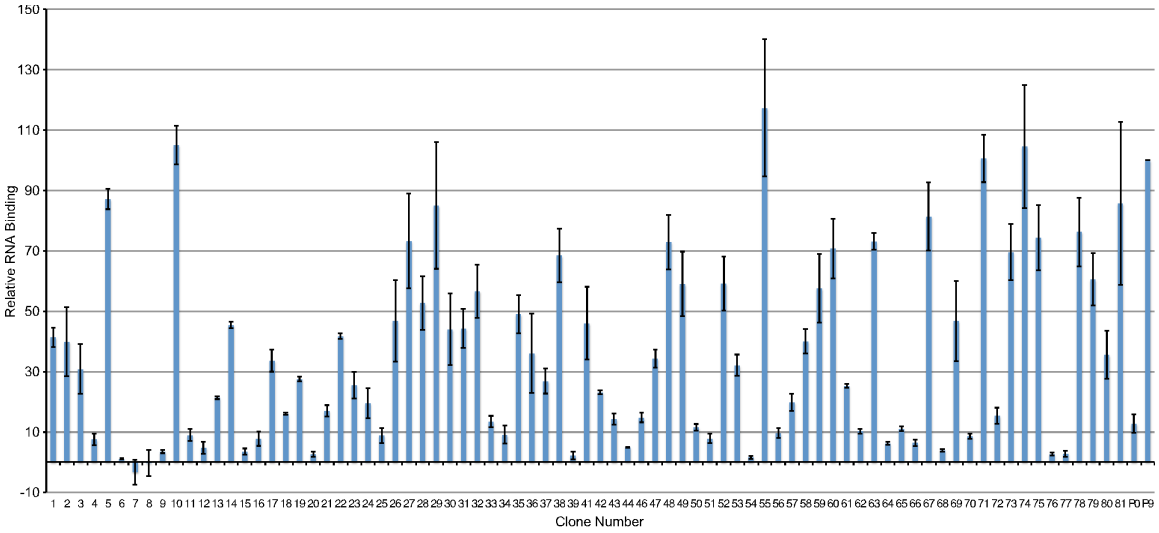
Supplementary Figure 2. Initial screening of final selected pool. 78 clones from pool 9 were screened by the nitrocellulose filter paper binding assay. Clones were screened 8-10 at a time and individual values were normalized by binding value of pooled P9 RNA that was measured in parallel. Y axis gives relative binding to the pooled RNA, X axis indicates the clone number. Error bars represent the standard error of 3 separate binding experiments.

Supplementary Figure 3. Characterization of the top 10 binders from pool 9. (A) Sequences of the 10 highest-performing molecules from a screen of 78 clones. Clone

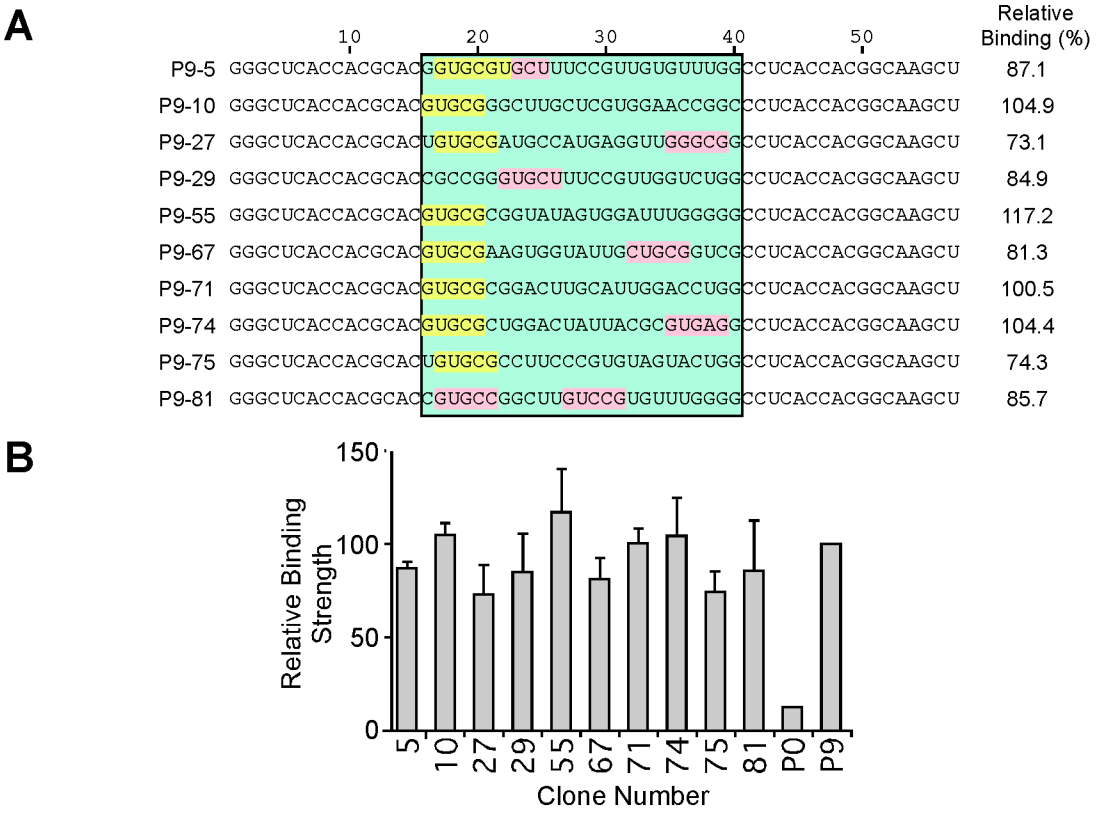
number is given on the left, base position is indicated on top. The selected region is indicated with a green box. Yellow highlighted bases are exact matches for the GUGCG motif identified by MEME, while pink highlighted bases are near-matches, where 4/5 bases of the motif are present. **(B)** Relative binding affinities of each of the top 10 binders, compared to P0 (unselected) and P9 (selected for 9 rounds) pooled RNA. Binding strength (y axis) is expressed as percentage bound corrected to P9 binding. Error bars represent the standard error of 3 separate binding experiments.



Supplementary Figure 1. Purification of recombinant SMN protein



Supplementary Figure 2. Initial screening of final selected poo



Supplementary Figure 3. Characterization of the top 10 binders from pool 9

Supplementary Table 1: Chemicals, reagents, and enzymes used in this study		
Reagent name	Manufacturer	Product Catalog #
ER2566 cells	NEB	Part of Impact kit (E6901)
LB agar	BD Biosciences	244520
LB broth	BD Biosciences	244620
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Fisher	BP1755-10
pTXB3	NEB	Part of Impact kit (E6901)
HEPES	Sigma	H4034-1KG
Sodium Chloride (NaCl)	Sigma	S3014-10KG
Ethylenediaminetetraacetic acid (EDTA)	Sigma	EDS-1KG
IgePal CA-630	Sigma	I8896-10ML
HALT Protease Inhibitor, 100X	Thermo Scientific	78425
PolyPrep Chromatography Columns	Biorad	731-1550
Chitin resin	NEB	S6651L
DL-dithiothreitol (DTT)	Thermo Scientific	R0861
Pierce concentrator columns (20,000 kDa)	Thermo Scientific	87750
Glycerol	Sigma	G5516-1L
Taq DNA polymerase with standard (Mg-Free) buffer	NEB	M0320L
dNTPs (dATP, dCTP, dTTP, dGTP)	Fermentas	R0152
Tris base (Trizma)	Sigma	T1503-5KG
Boric acid	Sigma	B6768-1KG
40% Acrylamide-Bisacrylamide 19:1	Calbiochem	1710
2-log DNA ladder	NEB	N3200S
Megashortscript T7 transcription kit	Ambion	AM1354
Urea	Sigma	S1456-25KG
Micro Bio-spin column P-30, RNase-free	Biorad	7326251
Ammonium Acetate	Sigma	A1542-500G
Sodium Dodecyl Sulfate (SDS)	EMD	7910
Phenol:chloroform 1:1	Calbiochem	6810
Ethanol, absolute	Sigma	E7023-500ML
BA-85 Nitrocellulose filters	Whatman	10 402 578
Magnesium Chloride	Sigma	M8266-100G
Glycogen	Roche	10901393001
Potassium Acetate	Sigma	P1190-100G
Superscript III reverse transcriptase	Invitrogen	18080-044
RNaseOUT RNase inhibitor	Invitrogen	10777-019
EcoRI	NEB	R0101L
HindIII	NEB	R0104L
pUC19	NEB	N3041S
[α-^{32}P]UTP	Perkin-Elmer	NEG507H001MC
rNTPs (ATP, CTP, UTP, GTP)	Amerham	272056, 272066, 272076, 272086
T7 RNA polymerase	NEB	M0251L
Borosilicate liquid scintillation vials	VWR	66022-300
Colloidal Coomassie stain	Biorad	161-0803
Bovine Serum Album (BSA)	Thermo Scientific	23210
Gel loading buffer II	Ambion	AM8546G
Carbenicillin	Sigma	C1389-1G
Acrylamide	Sigma	A9099-1KG
N,N'-Methylene Bisacrylamide	Sigma	146072-100G

Supplementary Table 2: Primers used in this study	
Primer name	Sequence
Temp-0	CGACTCACTATAGGGCTCACCACGCACNNNNNNNNNNNNNNNNNNNNNNNNNCCTCACCACGGCAAG
5'Pool	CAGAGAATTCTAATACGACTCACTATAGGGCTCACCACGCAC
3'Pool	CTTGCCGTGGTGAGG
3'Hind	ATATAAGCTTGCCGTGGTGAGG

CHAPTER IV

HIGH AFFINITY RNA TARGETS OF THE SURVIVAL MOTOR NEURON PROTEIN DEFINED BY UV-CROSSLINKING AND IMMUNOPRECIPITATION

A manuscript under preparation

Eric W Ottesen[1], Ravindra N Singh[1*]

[1]Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA

*Corresponding author: Ravindra N. Singh, Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA; Phone: 1 (515) 294-8505; Email: singhr@iastate.edu

Abstract

Spinal muscular atrophy (SMA) is a genetic disease caused by low levels of the Survival Motor Neuron (SMN) protein. SMN is found in complexes with a number of RNA species within cells and has been shown to interact with RNA in vitro; however, a direct interaction with RNA in cells has not been explored. Here, we performed in vivo crosslinking and immunoprecipitation coupled with high-throughput sequencing (HITS-CLIP) as well as knockdown of SMN levels followed by RNA-Seq to identify RNA targets of SMN within living cells. HITS-CLIP identified a variety of RNA targets of

SMN with an enrichment of mRNAs participating in a number of pathways, including ribosome function and actin cytoskeleton regulation. *SPON2*, *LAMB2*, and *EEF1A2* in particular were all predicted by HITS-CLIP to be bound by SMN and were downregulated upon SMN knockdown, indicating a direct regulatory role for SMN on expression of these genes.

Introduction

Spinal muscular atrophy (SMA) is the leading genetic cause of infant mortality (Wirth et al 2006). SMA is caused by low levels of the Survival Motor Neuron (SMN) protein due to homozygous mutation or deletion of the *SMN1* gene (Lefebvre et al 1995). Humans have an additional copy of the *SMN* gene, *SMN2* (Lorson et al 1999). *SMN2* predominantly produces an exon 7-skipped form of transcript due to a translationally silent point mutation in exon 7 (Lorson et al 1999). In SMA, *SMN2* fails to fully compensate for lack of *SMN1* due to a greatly reduced stability of the exon 7-skipped form of the protein, SMNΔ7 (Cho and Dreyfuss 2010).

RNP crosslinking and immunoprecipitation (CLIP) is a powerful approach to identifying RNA-protein interaction sites that has allowed researchers to identify in vivo interaction sites through irreversible crosslinking a protein to its interacting RNA (Ule et al 2003). CLIP identifies sites of actual contact that may be defined by additional interactions or extended RNA structure. With the advent of next-generation sequencing techniques, the method was modified to include high-throughput sequencing (HITS-CLIP) (Licatalosi et al 2008, Darnell 2010).

SMN is a multifunctional protein which has been implicated in numerous functions such as snRNP assembly (Meister et al 2001, Zhang et al 2008), signal recognition particle (SRP) biogenesis (Piazzon et al 2013) transcription (Zhao et al 2016), translation (Sanchez et al 2013), stress granule formation (Hua et al 2004, Zou et al 2011), endocytosis and synaptic transmission (Dimitriadi et al 2016), and mRNP transport (Rossoll et al 2003, Akten et al 2011, Fallini et al 2011, Fallini et al 2014). Of note, a large number of these functions involve interactions with RNA, either directly or indirectly. In many cases, SMN carries out these functions through its interactions with numerous RNA-binding proteins (RBPs) (Mourelatos et al 2001, Pellizzoni et al 2001, Rossoll et al 2002, Tadesse et al 2008, Glinka et al 2010, Yong et al 2010, Fallini et al 2011, Piazzon et al 2013, Fallini et al 2014, So et al 2016). SMN itself is capable of binding RNA in vitro (Lorson and Androphy 1998, Bertrandy et al 1999), indicating that it may be able to play a more direct role in RNA metabolism.

In vivo RNA targets of SMN have been previously identified by UV crosslinking and pulldown followed by microarray analysis (RIP-Chip) (Rage et al 2013). These experiments were carried out in mouse motor neuron-like NSC-34 cells carrying an inducible FLAG-tagged form of SMN. RIP-Chip of FLAG-tagged SMN revealed a number of transcripts targeted by SMN, and the interaction highly correlated with altered mRNA distribution in cultured neurons upon knockdown of SMN (Rage et al 2013). Of note, due to technical limitations of the microarray platform, sequence information of RNA-SMN interaction sites is not available. In addition, a direct interaction between SMN and RNA was not investigated.

Here we employ HITS-CLIP to uncover in vivo RNA targets of SMN. We chose human neuroblastoma SH-SY5Y cells for this study due to SMA being a neurodegenerative disease and the types of transcripts generated in neuronal cells is distinct from those generated in non-neuronal cells. Our results of HITS-CLIP revealed a wide range of RNA targets including messenger RNAs (mRNAs), small nucleolar RNAs (snoRNAs), long noncoding RNAs (lncRNAs), and snRNAs. We observed a complex sequence preference of RNA-SMN interaction, with enrichment of G- and A-rich sequence motifs near crosslinking sites. In order to determine whether SMN interaction is playing a role in RNA expression or stability, we performed transcriptome profiling of SMN-depleted SH-SY5Y cells. Expressions of several transcripts including *SPON2*, *LAMB2*, and *EEF1A2* that were identified as SMN targets by HITS-CLIP, were impacted by low levels of SMN. Our results support novel functions of SMN through direct sequence-specific interactions with transcripts.

Results

Identifying in vivo interactions through HITS-CLIP

In order to identify RNA sequences in direct contact with the SMN protein in vivo, we performed HITS-CLIP (Figure 1). This method utilizes UV light to form irreversible covalent bonds between nucleic acids and proteins, allowing for highly stringent purification (Ule et al 2003). Of note, it takes into account all interactions, including those that are not possible to recreate in vitro. Since SMA is a neurodegenerative disease (Burghes and Beattie 2009), and we wished to uncover interactions within neurons that

may have an impact on SMA disease progression, we performed these experiments in neuronal SH-SY5Y cells. Briefly, large numbers (>10^7) of neuronal SH-SY5Y cells were grown in culture, then exposed to UV light at 254 nm in order to induce crosslinking between cellular RNAs and proteins. Of note, these crosslinks form only in between molecules that are in close contact (<10 Å), so crosslinking indicates a direct interaction. SMN-containing complexes were purified by immunoprecipitation using anti-SMN antibody. After immunoprecipitation, transcripts were radiolabeled with ^{32}P and the crosslinked complexes were separated on denaturing gels and transferred to nitrocellulose membranes.

The SMN protein runs at a size of 42 kilodaltons (kDa) as shown by Western blotting (Figure 2A, left panel). Due to the low efficiency of the crosslinking reaction, only a very small fraction of total SMN is likely to crosslinked by this method. Imaging by autoradiography reveals a primary band at approximately 48 kilodaltons (kDa), which is consistent with the predicted size of SMN crosslinked to RNA with a modal size of 18-20 nucleotides (Figure 6A, right panel). After excision of the crosslinked band from the nitrocellulose membrane, RNAs were eluted and libraries for Illumina sequencing were prepared, then sequenced at high depth in order to identify all of the cellular interaction partners of SMN. The details of sequencing libraries are outlined in Supplementary Table 2.

After removing duplicate reads to eliminate PCR bias, we mapped CLIP sequence tags to the human genome. As a control for total RNA abundance, we also sequenced and mapped total ribosome-depleted RNA derived from SH-SY5Y cells. The majority of unique mapped reads, which shows enrichment compared to total SH-SY5Y RNA,

overlapped known protein-coding genes (Fig. 2B). The next most abundant type of reads mapped to ribosomal RNA (Fig. 2B). We assume that these reads represent nonspecific pulldown of ribosomes. A non-negligible proportion of reads (11.2%) mapped to known long noncoding RNAs (lncRNAs), although they represented a smaller proportion of total HITS-CLIP reads (Fig. 2B).

Identifying Individual Sites of Interaction

In order to identify individual candidate binding regions, we analyzed the mapped reads to identify peaks with significantly increased read density compared to a random distribution. Prior to peak calling and all future analyses, we removed all reads overlapping the Repeatmasker annotation track. Similar to the reads derived from CLIP, the majority of enriched peaks mapped to protein-coding genes (Figure 2B). A smaller proportion of peaks overlapped snoRNAs and lncRNAs, although both were enriched compared to the proportion of raw reads mapping to these RNA types (Figure 2B). Within protein-coding genes, the majority of reads mapped to introns (Figure 2C). However, a reduced proportion was concentrated into significantly enriched peaks, indicating that few of the intron-derived reads represented high-confidence binding sites (Figure 2C). Within exons, CLIP reads and peaks followed a similar distribution to the SH-SY5Y transcriptome, with a slight enrichment of reads corresponding to the 5′ UTR and coding region (Figure 2C).

In order to determine whether specific families of genes or pathways are disproportionately bound by SMN, we examined the genes overlapping significant CLIP peaks for enriched gene ontology (GO) terms and Kyoto Encyclopedia of Genes and

Genomes (KEGG) pathways. The most enriched KEGG pathway is for ribosomal proteins, followed by a number of pathways related to the cytoskeleton (Fig. 2D). Many of the other pathways involve RNA metabolism and interaction with the extracellular environment (Figure 2D).

Next, we wished to compare our results of HITS-CLIP with previously identified RNA targets of SMN. We compiled a list of human orthologs to the significantly enriched RNAs identified in a previous study by RIP-Chip (Rage et al 2013) and found 236 comparable human genes. We then matched those 236 genes to our list of 488 genes overlapping CLIP peaks and found an overlap of 33 genes (Supplementary Table 2). Of note, this overlap contains *ANXA2*, one of the highest-confidence targets identified in RIP-Chip, as well as U12 minor spliceosomal snRNA, genes encoding 5 separate ribosomal proteins and 7 snoRNAs. In addition, although not found to be significantly enriched by microarray, Rage et al also identified *Vim* as a target by RT-PCR and colocalization assays, which was also identified in our study.

In order to identify significantly enriched sequence motifs within regions potentially bound by SMN, we identified each peak that overlapped in 2 or more of the three replicates of CLIP and extracted the genomic sequence of the surrounding region. We then searched for significantly enriched motifs using MEME (Bailey et al 2009). In each condition that we tested, the highest enriched motif was purine-rich, generally with a repetitive pattern [WGA] x n (Fig. 2E).

SMN binding to specific RNA targets

We closely examined the pattern of mapped CLIP reads to several candidate RNAs (Figure 3). Eukaryotic elongation factor 1A2 (*EEF1A2*) and Laminin β2 (*LAMB2*) are novel high-confidence targets identified in our CLIP assay. The majority of CLIP tags mapping to *EEF1A2* were concentrated in a single peak in the fifth exon (Figure 3A, top panel). In contrast, reads derived from RNA-Seq of SH-SY5Y cells were mapped throughout the length of the gene (Figure 3A, bottom panel). Similarly, CLIP tags mapping to *LAMB2* primarily mapped to the 16th exon (Figure 3B, top panel), compared with an even distribution of RNA-Seq reads (Figure 3B, bottom panel). Vimentin (*VIM*), β-actin (*ACTB*), 7S RNA and U11/U12 RNAs have been previously identified as targets of SMN in previous studies (Lotti et al 2012, Piazzon et al 2013, Rage et al 2013, Rossoll et al 2003). *VIM* CLIP tags primarily mapped to the 7th exon; however, there were also smaller peaks throughout the length of the gene (Figure 3C). *ACTB*, in contrast with the other genes, which had clearly defined peaks, had CLIP tags mapped throughout the length of the mRNA, although there are some regions of increased read density in exons 2, 3, 4, and 6 (Figure 3D). All 3 of the small RNAs examined here had a single primary peak, but reads mapped throughout the length of the RNA (Figure 3E-3G). Interestingly, both U11 and U12 had reads derived from RNA-Seq that mapped beyond the annotated end of the snRNA (Figure 3F-3G, lower panels). This is consistent with a previous report that snRNAs initially exist as precursors that are cleaved during biogenesis (Yong et al 2010). In contrast with RNA-Seq, CLIP tags only mapped to the annotated region of the snRNA genes, suggesting that SMN interacts only with the mature snRNA (Figure 3F-3G, top panels).

Effect of SMN Depletion on SH-SY5Y Transcriptome

Aside from its known roles in snRNP assembly and mRNP transport (Meister et al 2001, Zhang et al 2008, Rossoll et al 2003, Akten et al 2011, Fallini et al 2011), very little is known about the potential biological significance of RNA-SMN interactions. In order to determine whether SMN interactions may be linked to altered RNA expression, we performed depletion experiments using siRNA targeted against the SMN mRNA followed by RNA-Seq. We isolated RNA and protein from SMN-depleted SH-SY5Y cells at 24 and 48 hours after a single transfection with SMN-targeting siRNAs, as well as from cells transfected twice at 48-hour intervals for a total of 96 hours of treatment. As controls, we used a non-targeting pool of siRNAs as well as siRNA against TIA1, an RNA-binding protein known to regulate SMN exon 7 splicing (Singh et al 2011). To confirm efficient knockdown of SMN levels, we performed Western blotting using antibody against SMN. SMN levels were strongly reduced at all time points, with almost no visible expression at 96 hours. (Fig. 4A) Interestingly, there is an interaction between SMN and TIA1 protein levels, as SMN expression is strongly reduced in TIA1 knockdown. Although TIA1 is predicted to increase SMN levels by stimulating inclusion of *SMN2* exon 7 (Singh et al 2011), the effect on the protein levels of SMN appears to be even greater than expected from altered splicing alone. There is also an apparent effect of SMN depletion on TIA1 protein levels, as TIA1 expression levels are noticeably reduced upon SMN depletion. (Figure 4A)

In order to test whether SMN interaction with a given RNA is linked to altered expression or stability, we performed deep sequencing on RNA isolated from the SMN-

depleted SH-SY5Y cells. Because SMN has been implicated in the biogenesis of multiple species of noncoding RNA (ncRNA), and the results of CLIP support an interaction with several ncRNAs, we performed ribosomal RNA depletion before sequencing rather than the standard poly(A) selection generally used for mRNA-Seq. Three replicates of each treatment were sequenced on an Illumina HiSeq 2500, resulting in 22.9-33.4 million reads per sample (Supplementary Table 3).

After mapping the RNA-Seq reads to the human genome (UCSC version hg38), we performed differential expression (DE) analysis. At 24, 48, and 96 hours post transfection, we identified 382, 297, and 3537 genes, respectively, with altered RNA levels (adjusted P value < 0.05) upon SMN depletion (Fig. 7B-D). TIA1 depletion resulted in a much smaller number of altered genes, likely due to the presence of a functionally similar homolog, TIAR (Wang et al 2010). A small number of genes (7, 2, and 12 for 24, 48, and 96 hours) were significantly altered in both SMN and TIA1 knockdown. These genes may represent nonspecific responses to siRNA-mediated knockdown, or may represent genuine overlap between SMN and TIA1 function. Since they represent such a small minority of genes affected by SMN knockdown, they were not eliminated from further analyses. We compared the list of genes altered upon knockdown of SMN with those containing CLIP tags. Of 488 genes with CLIP tags, 113 had altered expression levels upon knockdown of SMN in at least one of the time points tested, which is a highly significant level of overlap ($p = 5.42 \times 10^{-29}$). This indicates that SMN may be involved in regulating RNA expression. The strongest enrichment that we observed was in genes which are upregulated at 96 hours post-transfection ($p = 7.50 \times 10^{-}$

[28]), although it is not known whether this is due to a stabilization of RNA in the absence of SMN or compensatory expression due to loss of SMN interaction.

We performed a KEGG pathway analysis with genes with altered expression levels upon SMN knockdown. The most strongly affected pathway at 24 hours is the MAP kinase (MAPK) signaling pathway (Fig. 7E). At 48 hours post transfection, the enriched pathways bear a strong resemblance to the pathways enriched in the identified binding partners of SMN (Figs. 6D, 7F). In particular, focal adhesion, RNA transport, and ECM-intracellular interaction pathways are all significantly enriched. At 96 hours post-transfection, the enriched KEGG pathways are dominated by protein translation and energy metabolism genes (Fig.7G).

Next, we identified a limited number of strong candidate genes to use for validation. It has previously been reported that CLIP experiments are prone to nonspecific pulldown of highly abundant RNA species (Friedersdorf et al 2014), so we chose to eliminate the top 10% of highest-expressing genes from our RNA-Seq experiment from further analysis. Likewise, genes that were expressed at very low levels in RNA-Seq may represent artifacts or non-specific mapping, so we eliminated the lowest-expressing genes as well. After visual inspection of mapped reads, we chose 15 high-confidence targets (Table 3). Confirming that these have a high probability of being genuine targets of SMN, 6 of the 15 (*SNX12, CDC42EP1, TXNIP, LAMB2, SPON2,* and EEF1A2) high-confidence targets were predicted by RNA-Seq to be either upregulated or downregulated at one or more time points. We next tested the altered expression of these genes by QPCR. All three of the genes predicted to be downregulated after knockdown of SMN (*LAMB2*, *SPON2*, and *EEF1A2*) were genuinely reduced (Figure 8A). *EEF1A2*

showed the strongest response, with greater than 5 fold reduction at 96 hours post-knockdown of SMN. The genomic organization of *EEF1A2* and the location of CLIP reads is shown in Figure 8B. We obtained the greatest density of reads mapping to the first half of exon 5. We observed two single-base deletions among the reads mapping to this region, which is indicative of error-prone readthrough of reverse transcriptase at the site of crosslinking (Zhang and Darnell 2011).

Discussion

We performed HITS-CLIP to identify in vivo RNA targets of the SMN protein in human neuronal SH-SY5Y cells. Similar to a previous study utilizing RIP-Chip performed in the mouse motor neuron-like NSC34 cells (Rage et al 2013), we observed a preponderance of reads mapping to mRNAs (Figure 6B). Surprisingly, despite the critical role for SMN in snRNP biogenesis, reads mapping to snRNAs were underrepresented compared to RNA-Seq of total RNA from SH-SY5Y cells (Figure 6B). This could indicate that interactions with snRNAs are transient and only occur during snRNP assembly, while interactions with other types of RNAs such as mRNAs and lncRNAs could occur throughout the life of the RNA. Although slightly under-represented compared to total RNA-Seq, we also observed a significant portion of predicted binding sites mapped to lncRNAs, which we have recently shown to be affected by low levels of SMN in a mouse model of SMA (Ottesen et al 2016). snoRNAs made up the bulk of the remainder of predicted binding sites. Given that SMN interacts with snoRNP proteins Fibrillarin and GAR1 (Pellizzoni

et al 2001), an interaction with snoRNAs is not surprising. However, a direct role in snoRNP biogenesis has never been conclusively proven.

We observed an enrichment of purine-rich motifs within SMN target regions identified by CLIP, with the most enriched motifs following the repetitive pattern [WGA] x n (Fig. 6E). However, this motif was not universally conserved within predicted binding sites, indicating a binding mode that is more complex than binding to a single motif based on RNA sequence. In addition, many high-confidence target RNAs, including *ACTB*, have reads mapped throughout the length of the mRNA rather than one or two strong peaks, indicating that SMN may be binding multiple locations. Of note, SMN self-assembles to form oligomers, which could potentially bind cooperatively (Martin et al 2012, Gupta et al 2015). In addition, SMN can interact with a number of other RBPs (Rossoll et al 2002, Tadesse et al 2008, Yamazaki et al 2012, Fallini et al 2011, Fallini et al 2014). The lack of a single strong, distinct binding motif may be due to recruitment by a variety of these RBPs.

Many of the RNAs identified by CLIP are also functionally related (Figure 6D). The most enriched group of RNAs code for ribosomal proteins, consistent with the results of a previous study (Rage et al 2013). Many of the other pathways indicated by CLIP function in regulation and maintenance of the actin cytoskeleton (Figure 6D). Consistently, β-actin was one of the first mRNAs identified as an in vivo target of SMN (Rossoll et al 2003). Local translation of actin in the growth cone is critical for neurite outgrowth and steering (Willis et al 2005, Leung et al 2006). A role for SMN in the transport and localized translation of actin and other cytoskeletal components is therefore compatible with the increased sensitivity of motor neurons to low levels of SMN.

Several transcriptome-wide expression studies have been carried out in conditions of low SMN expression (Zhang et al 2013, Maeda et al 2014, Ng et al 2015, Ottesen et al 2016), which allows us to look for similar trends between studies and high-confidence candidates for direct targets of SMN. Surprisingly, there is little overlap between genes identified by multiple studies, or even in different tissues isolated in parallel within the same study (Zhang et al 2013) (Table 1). Of note, the most divergent samples were obtained from laser capture microdissection of tissues, which are likely to undergo more complex regulation. There was a stronger overlap between RNA-Seq experiments performed in cultured cells; for example, we observed 32 up-regulated and 291 downregulated genes which were similarly affected in our study and in another which examined motor neurons derived from induced pluripotent stem cells (iPSCs) derived from SMA mice (Maeda et al 2014) (Table 1). Of note, these include *EEF1A2*, *SPON2*, and *LAMB2*, which we identified as the highest-confidence targets of SMN by CLIP. *EEF1A2* is of particular interest as a potential target of SMN, as this gene is specifically expressed in muscles and motor neurons (Abbott et al 2009) and is mutated in the wasted (wst) mouse model (Chambers et al 1998), which is one of the first mouse models of motor neuron degeneration, with a very similar progression to SMA (Newbery et al 2005).

In addition to identifying in vivo RNA targets of SMN, we also performed Systematic Evolution of Ligands by Exponential Enrichment (SELEX) to identify in vitro RNA targets of the SMN protein (see Chapter 3 of this dissertation). Notably, the sequence preference that we identified by CLIP is distinct from the GUG, UGC, UGG, and GCG trimers observed in SELEX-identified sequences. However, many of the

trinucleotide motifs identified by SELEX were present in CLIP tags, suggesting that specificity of RNA-SMN interactions were driven by additional motifs and/or structural contexts that were not fully captured in CLIP experiment. While SELEX and CLIP performed on the same RNA-binding proteins tend to identify similar motifs (Wang et al 2010, Hafner et al 2010), there have been cases where differences arose even between SELEX experiments performed under differing conditions. For instance, results of two SELEX experiments on ASF/SF2 produced different motifs (Tacke and Manley 1995, Smith et al 2006). A CLIP experiment on ASF/SF2 produced yet another consensus motif (Sanford et al 2008). Further, in vivo RNA-protein interactions are expected to be much more diverse due to the fact that several types of protein-protein complexes could interact with RNA, while in case of in vitro selection only high affinity interactions against a single protein are captured (Jankowsky and Harris 2015). In vivo interactions are also subject to regulation by post-translational modification, of which SMN is a target (Grimmler et al 2005, Husedzinovic et al 2014) or recruitment by one of the many RBP binding partners of SMN (Rossoll et al 2002, Tadesse et al 2008, Yamazaki et al 2012, Fallini et al 2011, Fallini et al 2014).

In summary, we have shown that SMN has a direct role in RNA regulation through physical interactions with specific RNAs in cells. CLIP tags had an enrichment of purine-rich sequences, generally consisting of repeated WGA trinucleotides. By combining HITS-CLIP with RNA-Seq after knockdown of SMN, we show that several candidate target genes have altered expression upon loss of SMN expression. This indicates a possible role for SMN in RNA expression or stability of those genes. Several targets of SMN, such as *ACTB*, *VIM*, and *EEF1A2* have known roles in axon outgrowth

and motor neuron survival, establishing a possible direct link between SMN's RNA binding ability and motor neuron health.

Materials and Methods

Reagents

All reagents and chemicals used in this study and their catalog numbers are listed in Supplementary Table 4

Cell culture and transfection

Unless otherwise stated, all tissue culture media and reagents were purchased from Life Technologies. SH-SY5Y cells were obtained from the American Type Culture Collection (ATCC) and cultured in a 50:50 mix of minimal essential media (MEM) and F12 nutrient mixture supplemented with 10% fetal bovine serum (FBS). To transfect cells with siRNA, cells were trypsinized and 1.2-1.5 x 10^6 cells were reverse transfected with 20 nM smartPOOL siRNA (Dharmacon) using Lipofectamine 2000 (Life Technologies) following the manufacturer's instructions and plated on 100 mm dishes. For the 96-hour time point, cells were trypsinized 48 hours after the initial transfection and 1.5-2 x 10^6 cells were re-transfected in a similar manner to the initial transfection. Trypsinized cells not used for the second transfection were collected and saved as the 48-hour time point. Cells were collected by scraping (24 and 96 hours) or trypsinization (48 hours). Cell suspensions were divided into two portions, one for future RNA isolation and one for

preparation of protein lysates, then centrifuged at 1000 x g for 4 minutes to collect cells; supernatant was removed and pellets flash frozen on dry ice and stored at -80 ˚C.

Crosslinking and immunoprecipitation

SH-SY5Y cells were grown (~1.2-2 x 10^7) in 12-20 150 mm dishes per replicate to approximately 70-80% confluency. Media was removed and plates were washed 2X with ice-cold phosphate-buffered saline (PBS). After removing the second PBS wash, plates were placed on ice and exposed to 150 mJ/cm^2 of 254 nm UV light in a UV Stratalinker (Stratagene). Cells were collected by scraping in 4 mL of PBS per plate and centrifuged at 1000xg for 4 minutes. After removal of supernatant, pellets were flash frozen in liquid nitrogen and stored at -80˚C. For negative (uncrosslinked) control samples, all steps were identical except the UV crosslinking step was omitted. To lyse cells, pellets were resuspended in 250 µl of radioimmunoprecipitation assay (RIPA) buffer supplemented with 1X HALT protease inhibitor (Thermo) per original plate of cells and incubated on ice for 10 minutes. All further treatments and steps are described per 1 mL of lysate. To remove genomic DNA and partially fragment RNPs, lysates were treated with 50 µl of RQ1 RNase-free DNase (Promega) and 10 µl of a 1:50 dilution of RNase A/T1 mix (Ambion) in a 37˚ C water bath for 10 minutes, then placed on ice for 5 minutes. Lysates were then centrifuged at 12,000xg for 15 minutes and supernatant was transferred to a new tube. SMN-containing complexes were immunoprecipitated using 100 µl of protein G Dynabeads (Invitrogen) coupled with 5 µg monoclonal anti-SMN (BD Biosciences) using the method described by Hafner et al (2010). After incubating with lysate for 30 minutes at 4˚C, beads were washed 3 times in 1 ml RIPA buffer. RNA dephosphorylation

and labeling was carried out following the procedure outlined in Ule et al (2005). After labeling and washing, RNP complexes were eluted by boiling in 40 µl SDS-PAGE buffer for 5 minutes, then loaded onto 10% NuPAGE gels (Invitrogen). After SDS-PAGE separation, complexes were transferred onto Protran BA-85 nitrocellulose membranes (Whatman) and exposed to phosphorimager screens, which were scanned using a Fujifilm FPL-5000 imager. Printouts were lined up with reference points on the original membrane and bands corresponding to crosslinked RNP complexes were excised with a razor blade. RNA was eluted by incubating membrane slices in 200 µl of 4 µg/µl proteinase K in 1X PK buffer (100 mM Tris–Cl, pH 7.5, 50 mM NaCl, 10 mM EDTA) for 20 minutes at 37˚C, then adding an equal volume of 1X PK buffer supplemented with 7M Urea and incubating for another 20 minutes. RNA was then purified by phenol:chloroform extraction followed by ethanol precipitation in the presence of 0.5 µl glycogen (Roche). Samples were resuspended in 8 µl RNase-free water, then Illumina sequencing libraries were then prepared using the TruSeq small RNA library preparation kit according to the manufacturer's instructions, with the exception that an additional gel purification step was performed between 3′ and 5′ adapter ligation to prevent adapter dimers. For each replicate, PCR amplification was optimized separately using an aliquot of cDNA. For final library preparation, 20-24 cycles of PCR amplification was carried out, depending on the replicate.

RNA isolation and RNA-seq library preparation

RNA was isolated from pelleted cells using the miRVana miRNA isolation kit (Ambion) following the total RNA procedure. RNA concentration was measured using a UV

spectrophotometer and 6 µg was treated with RQ1 RNase-free DNase following the manufacturer's instructions. After DNase treatment, RNA was repurified on Qiagen RNeasy columns according to the manufacturer's instructions, except that prior to binding to the columns, 700 µl of ethanol was added to the samples instead of 250 µl in order to preserve binding of small RNAs. RNA quantity and quality was then assessed with an Agilent Bioanalyzer using the RNA Nano chip. Ribosome depletion was carried out on 1.5 µg total RNA using a RiboZero gold kit (Epicentre), and the entire rRNA-depleted sample was then used as input for library generation using the TruSeq mRNA-seq library preparation kit (Illumina).

Illumina sequencing, mapping, and peak calling

For sequencing of CLIP libraries, the samples for all three replicates were pooled and sequenced using an Illumina MiSeq following a 50-base single-end sequencing protocol. For sequencing of total SH-SY5Y RNA, 6-7 samples were pooled per lane of an Illumina HiSeq 2000 following a 100-base single-end sequencing protocol. Following sequencing, adapter and quality trimming was carried out using Cutadapt (Martin 2011). Identical reads which may indicate PCR duplication were removed from CLIP libraries using the Fastx toolkit. Mapping to the human genome (version hg38) was carried out using Tophat (Trapnell et al 2009) using the Gencode v20 transcriptome annotation (Harrow et al 2012). Differential gene expression analysis was carried out using DESeq (Anders and Huber 2010). Regions of interest were determined from the CLIP sequencing data by Piranha (Uren et al 2012). Significantly enriched GO terms and KEGG pathways were

identified using the WebGESTALT web server (http://bioinfo.vanderbilt.edu/webgestalt/option.php).

Quantitative PCR

2.5 µg of total RNA was converted to cDNA using SuperScript III reverse transcriptase (Life Technologies) using random primers (Promega). For quantitative PCR, 1.5 µl of a 1:20 dilution of cDNA (equivalent to cDNA produced from 9.375 ng of RNA) was used as template in a 20 µl reaction containing 300 nM forward and reverse primers and 1X FastStart Universal SYBR Green Master containing ROX reference dye (Roche). QPCR was performed using a Stratagene MX3005P thermocycler (Agilent) and relative expression levels were determined using the ΔΔCt method using β-Actin as a reference gene. Primers used for QPCR are given in Supplementary Table 5.

References

Abbott, C.M., Newbery, H.J., Squires, C.E., Brownstein, D., Griffiths, L.A., and Soares, D.C. (2009). eEF1A2 and neuronal degeneration. Biochemical Society Transactions *37*, 1293-1297.

Akten, B., Kye, M.J., Hao, L.T., Wertz, M.H., Singh, S., Nie, D.Y., Huang, J., Merianda, T.T., Twiss, J.L., Beattie, C.E., *et al.* (2011). Interaction of survival of motor neuron (SMN) and HuD proteins with mRNA cpg15 rescues motor neuron axonal deficits. Proceedings of the National Academy of Sciences of the United States of America *108*, 10337-10342.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology *11*, 12.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J.Y., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research *37*, W202-W208.

Bertrandy, S., Burlet, P., Clermont, O., Huber, C., Fondrat, C., Thierry-Mieg, D., Munnich, A., and Lefebvre, S. (1999). The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. Human Molecular Genetics *8*, 775-782.

Burghes, A.H.M., and Beattie, C.E. (2009). Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? Nature Reviews Neuroscience *10*, 597-609.

Chambers, D.M., Peters, J., and Abbott, C.M. (1998). The lethal mutation of the mouse wasted (wst) is a deletion that abolishes expression of a tissue-specific isoform of translation elongation factor 1 alpha, encoded by the Eefla2 gene. Proceedings of the National Academy of Sciences of the United States of America *95*, 4463-4468.

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

Darnell, R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. Wiley Interdisciplinary Reviews-Rna *1*, 266-286.

Fallini, C., Rouanet, J.P., Donlin-Asp, P.G., Guo, P., Zhang, H.L., Singer, R.H., Rossoll, W., and Bassell, G.J. (2014). Dynamics of Survival of Motor Neuron (SMN) Protein Interaction with the mRNA-Binding Protein IMP1 Facilitates Its Trafficking into Motor Neuron Axons. Developmental Neurobiology *74*, 319-332.

Fallini, C., Zhang, H.L., Su, Y.H., Silani, V., Singer, R.H., Rossoll, W., and Bassell, G.J. (2011). The Survival of Motor Neuron (SMN) Protein Interacts with the mRNA-Binding Protein HuD and Regulates Localization of Poly(A) mRNA in Primary Motor Neuron Axons. Journal of Neuroscience *31*, 3914-3925.

Friedersdorf, M.B., and Keene, J.D. (2014). Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. Genome Biology *15*, 16.

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

Grimmler, M., Bauer, L., Nousiainen, M., Korner, R., Meister, G., and Fischer, U. (2005). Phosphorylation regulates the activity of the SMN complex during assembly of spliceosomal U snRNPs. Embo Reports *6*, 70-76.

Gupta, K., Martin, R., Sharp, R., Sarachan, K.L., Ninan, N.S., and Van Duyne, G.D. (2015). Oligomeric Properties of Survival Motor Neuron.Gemin2 Complexes. Journal of Biological Chemistry *290*, 20185-20199.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.C., Munschauer, M.*, et al.* (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell *141*, 129-141.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S.*, et al.* (2012). GENCODE: The reference human genome annotation for The ENCODE Project. Genome Research *22*, 1760-1774.

Hua, Y.M., and Zhou, J.H. (2004). Survival motor neuron protein facilitates assembly of stress granules. Febs Letters *572*, 69-74.

Husedzinovic, A., Oppermann, F., Draeger-Meurer, S., Chari, A., Fischer, U., Daub, H., and Gruss, O.J. (2014). Phosphoregulation of the human SMN complex. European Journal of Cell Biology *93*, 106-117.

Jankowsky, E., and Harris, M.E. (2015). Specificity and nonspecificity in RNA-protein interactions. Nature Reviews Molecular Cell Biology *16*, 533-544.

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., and Zeviani, M. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80*, 155-165.

Leung, K.M., van Horck, F.P.G., Lin, A.C., Allison, R., Standart, N., and Holt, C.E. (2006). Asymmetrical beta-actin mRNA translation in growth cones mediates attractive turning to netrin-1. Nature Neuroscience *9*, 1247-1256.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X.N.*, et al.* (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464-U422.

Lorson, C.L., and Androphy, E.J. (1998). The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. Human Molecular Genetics *7*, 1269-1275.

Lorson, C.L., Hahnen, E., Androphy, E.J., and 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*, 6307-6311.

Lotti, F., Imlach, W.L., Saieva, L., Beck, E.S., Hao, L.T., Li, D.K., Jiao, W., Mentis, G.Z., Beattie, C.E., McCabe, B.D., *et al.* (2012). An SMN-Dependent U12 Splicing Event Essential for Motor Circuit Function. Cell *151*, 440-454.

Maeda, M., Harris, A.W., Kingham, B.F., Lumpkin, C.J., Opdenaker, L.M., McCahan, S.M., Wang, W., and Butchbach, M.E.R. (2014). Transcriptome Profiling of Spinal Muscular Atrophy Motor Neurons Derived from Mouse Embryonic Stem Cells. Plos One *9*, 18.

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

Martin, R., Gupta, K., Ninan, N.S., Perry, K., and Van Duyne, G.D. (2012). The Survival Motor Neuron Protein Forms Soluble Glycine Zipper Oligomers. Structure *20*, 1929-1939.

Meister, G., Buhler, D., Pillai, R., Lottspeich, F., and Fischer, U. (2001). A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nature Cell Biology *3*, 945-949.

Mourelatos, Z., Abel, L., Yong, J.S., Kataoka, N., and Dreyfuss, G. (2001). SMN interacts with a novel family of hnRNP and spliceosomal proteins. Embo Journal *20*, 5443-5452.

Newbery, H.J., Gillingwater, T.H., Dharmasaroja, P., Peters, J., Wharton, S.B., Thomson, D., Ribchester, R.R., and Abbott, C.M. (2005). Progressive loss of motor neuron function in wasted mice: Effects of a spontaneous null mutation in the gene for the eEF1A2 translation factor. Journal of Neuropathology and Experimental Neurology *64*, 295-303.

Ottesen, E.W., Howell, M.D., Singh, N.N., Seo, J., Whitley, E.M., and Singh, R.N. (2016). Severe impairment of male reproductive organ development in a low SMN expressing mouse model of spinal muscular atrophy. Scientific Reports *6*, 17.

Pellizzoni, L., Baccon, J., Charroux, B., and Dreyfuss, G. (2001). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. Current Biology *11*, 1079-1088.

Piazzon, N., Schlotter, F., Lefebvre, S., Dodre, M., Mereau, A., Soret, J., Besse, A., Barkats, M., Bordonne, R., Branlant, C., *et al.* (2013). Implication of the SMN complex in the biogenesis and steady state level of the Signal Recognition Particle. Nucleic Acids Research *41*, 1255-1272.

Rage, F., Boulisfane, N., Rihan, K., Neel, H., Gostan, T., Bertrand, E., Bordonne, R., and Soret, J. (2013). Genome-wide identification of mRNAs associated with the protein SMN whose depletion decreases their axonal localization. Rna-a Publication of the Rna Society *19*, 1755-1766.

Rossoll, W., Jablonka, S., Andreassi, C., Kroning, A.K., Karle, K., Monani, U.R., and 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*, 801-812.

Rossoll, W., Kroning, A.K., Ohndorf, U.M., Steegborn, C., Jablonka, S., and Sendtner, M. (2002). Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons? Human Molecular Genetics *11*, 93-105.

Sanchez, G., Dury, A.Y., Murray, L.M., Biondi, O., Tadesse, H., El Fatimy, R., Kothary, R., Charbonnier, F., Khandjian, E.W., and Cote, J. (2013). A novel function for the survival motoneuron protein as a translational regulator. Human Molecular Genetics *22*, 668-684.

Sanford, J.R., Coutinho, P., Hackett, J.A., Wang, X., Ranahan, W., and Caceres, J.F. (2008). Identification of Nuclear and Cytoplasmic mRNA Targets for the Shuttling Protein SF2/ASF. Plos One *3*, 10.

Singh, N.N., Seo, J.B., Ottesen, E.W., Shishimorova, M., Bhattacharya, D., and Singh, R.N. (2011). TIA1 Prevents Skipping of a Critical Exon Associated with Spinal Muscular Atrophy. Molecular and Cellular Biology *31*, 935-954.

Smith, P.J., Zhang, C.L., Wang, J.H., Chew, S.L., Zhang, M.Q., and Krainer, A.R. (2006). An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Human Molecular Genetics *15*, 2490-2508.

So, B.R., Wan, L.L., Zhang, Z.X., Li, P.L., Babiash, E., Duan, J.Q., Younis, I., and Dreyfuss, G. (2016). A U1 snRNP-specific assembly pathway reveals the SMN complex as a versatile hub for RNP exchange. Nature Structural & Molecular Biology *23*, 225-230.

Tacke, R., and Manley, J.L. (1995). THE HUMAN SPLICING FACTORS ASF/SF2 AND SC35 POSSESS DISTINCT, FUNCTIONALLY SIGNIFICANT RNA-BINDING SPECIFICITIES. Embo Journal *14*, 3540-3551.

Tadesse, H., Deschenes-Furry, J., Boisvenue, S., and Cote, J. (2008). KH-type splicing regulatory protein interacts with survival motor neuron protein and is misregulated in spinal muscular atrophy. Human Molecular Genetics *17*, 506-524.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: A method for identifying protein-RNA interaction sites in living cells. Methods *37*, 376-386.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212-1215.

Uren, P.J., Bahrami-Samani, E., Burns, S.C., Qiao, M., Karginov, F.V., Hodges, E., Hannon, G.J., Sanford, J.R., Penalva, L.O.F., and Smith, A.D. (2012). Site identification in high-throughput RNA-protein interaction data. Bioinformatics *28*, 3013-3020.

Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N.M., Rot, G., Zupan, B., Curk, T., and Ule, J. (2010). iCLIP Predicts the Dual Splicing Effects of TIA-RNA Interactions. Plos

Biology *8*, 16.

Willis, D., Li, K.W., Zheng, J.Q., Chang, J.H., Smit, A., Kelly, T., Merianda, T.T., Sylvester, J., van Minnen, J., and Twiss, J.L. (2005). Differential transport and local translation of cytoskeletal, injury-response, and neurodegeneration protein mRNAs in axons. Journal of Neuroscience *25*, 778-791.

Wirth, B., Brichta, L., and Hahnen, E. (2006). Spinal muscular atrophy: from gene to therapy. Seminars in pediatric neurology *13*, 121-131.

Yamazaki, T., Chen, S., Yu, Y., Yan, B.A., Haertlein, T.C., Carrasco, M.A., Tapia, J.C., Zhai, B., Das, R., Lalancette-Hebert, M.*, et al.* (2012). FUS-SMN Protein Interactions Link the Motor Neuron Diseases ALS and SMA. Cell Reports *2*, 799-806.

Yong, J., Kasim, M., Bachorik, J.L., Wan, L.L., and Dreyfuss, G. (2010). Gemin5 Delivers snRNA Precursors to the SMN Complex for snRNP Biogenesis. Molecular Cell *38*, 551-562.

Zhang, C.L., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nature Biotechnology *29*, 607-U686.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. Cell *133*, 585-600.

Zhang, Z.X., Pinto, A.M., Wan, L.L., Wang, W., Berg, M.G., Oliva, I., Singh, L.N., Dengler, C., Wei, Z., and Dreyfuss, G. (2013). Dysregulation of synaptogenesis genes antecedes motor neuron pathology in spinal muscular atrophy. Proceedings of the National Academy of Sciences of the United States of America *110*, 19348-19353.

Zhao, D.Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z.Y., Schmitges, F.W., Zhong, G.Q., Liu, K., Li, W.G., Moffat, J.*, et al.* (2016). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature *529*, 48-+.

Zou, T., Yang, X.M., Pan, D.M., Huang, J., Sahin, M., and Zhou, J.H. (2011). SMN Deficiency Reduces Cellular Ability to Form Stress Granules, Sensitizing Cells to Stress. Cellular and Molecular Neurobiology *31*, 541-550.

Figure 1. Overview of CLIP procedure. Flowchart of experimental procedure is given. Briefly, SH-SY5Y cells were exposed to UV light at 254 nm to induce crosslinking. Afterwards, cells were lysed and subjected to RNase treatment. RNA-SMN complexes were isolated by immunoprecipitation and RNA was labeled with γ-^{32}P ATP. Complexes were separated by SDS-PAGE and transferred to nitrocellulose membranes. After autoradiography, region of the membrane corresponding to RNA-SMN complexes were removed and subjected to Proteinase K digestion to elute RNA. Finally, RNA was prepared for Illumina sequencing by ligation of adapter sequences and RT-PCR amplification.

Figure 2. RNA-SMN Crosslinking and Immunoprecipitation. (A) Verifying pulldown of SMN protein and RNA-SMN complexes. Protein size in kilodaltons (kDa) is given. Left panel: Western blot of input (I, 1% relative to eluted protein loaded) and eluted protein (E). Right panel: Autoradiogram of eluted protein without (-) and with (+) crosslinking of SH-SY5Y cells prior to immunoprecipitation. *: reactivity to anti-SMN antibody **(B)** Read mapping distribution across the transcriptome for RNA-Seq of SH-SY5Y cells (RNA-Seq) and reads obtained from CLIP libraries (CLIP), and distribution of statistically significant enriched regions (CLIP peaks). Color code for RNA types is given on the right of the graph. **(C)** Mapping distribution of reads and peaks across different regions of protein-coding mRNAs. Color code for mRNA regions is given below the graph. **(D)** Top 20 enriched KEGG pathways with genes overlapping CLIP peaks. Y axis represents the Benjamini and Hochberg (B+H) corrected *P* value of

enrichment, X axis indicates the KEGG pathway name. **(E)** Top enriched motifs among CLIP peaks; from left to right, most enriched 6-mer, 7-mer, 8-mer, and 9-mer.

Figure 3. Mapping of CLIP reads to candidate genes. Genomic view of several candidate genes are shown. **(A-D)** Genomic view of *EEF1A2* **(A)**, *LAMB2* **(B)**, *VIM* **(C)**, and *ACTB* **(D)**. Top panels represent mapping of CLIP tags, bottom panels represent RNA-Seq of SH-SY5Y cells. Height of blue indicate mapping depth of at each position. Blue lines below represent introns, thick boxes indicate coding region, and thin boxes represent untranslated regions. Exon numbers are given below gene diagrams. Direction of transcription is indicated with an arrow. **(E-G)** Overview of 3 small RNAs: 7S RNA **(E)**, U11 **(F)**, and U12 **(G)** snRNAs.

Figure 4. RNA-Seq after siRNA-mediated knockdown of SMN. (A) Western blot verifying SMN and TIA protein knockdowns. A representative sample from 3 replicate experiments is shown. **(B-D)** MA plot depicting average expression level (X axis) against $\log_2$ fold change (Y axis) between WT and SMN knockdown samples at 24 hours **(B)**, 28 hours **(C)**, and 96 hours **(D)** post transfection. Each dot represents one gene; red dots represent genes with significantly altered expression (B+H adjusted *P* val < 0.05). **(E-G)** MA plot depicting average expression level (X axis) against $\log_2$ fold change (Y axis) between WT and TIA1 knockdown samples at 24 hours **(B)**, 28 hours **(C)**, and 96 hours **(D)** post transfection. Labeling and coloring are the same as in **(B-D)**

Figure 5. Pathways enriched in genes affected by SMN knockdown (A-C) Top 20 enriched KEGG pathways among significantly altered genes at 24 hours **(A)**, 28 hours **(B)**, and 96 hours **(C)** post transfection. Y axis represents the Benjamini and Hochberg (B+H) corrected *P* value of enrichment, X axis indicates the KEGG pathway name.

Figure 6. Candidate genes regulated by SMN. Relative expression of 6 candidate CLIP targets. Gene expression was measured by QPCR 96 hours after transfection with siRNA. Y axis indicates relative expression as compared to samples transfected with a nontargeting pool of siRNA. X axis indicates genes tested. Error bars represent standard error of 3 replicates.

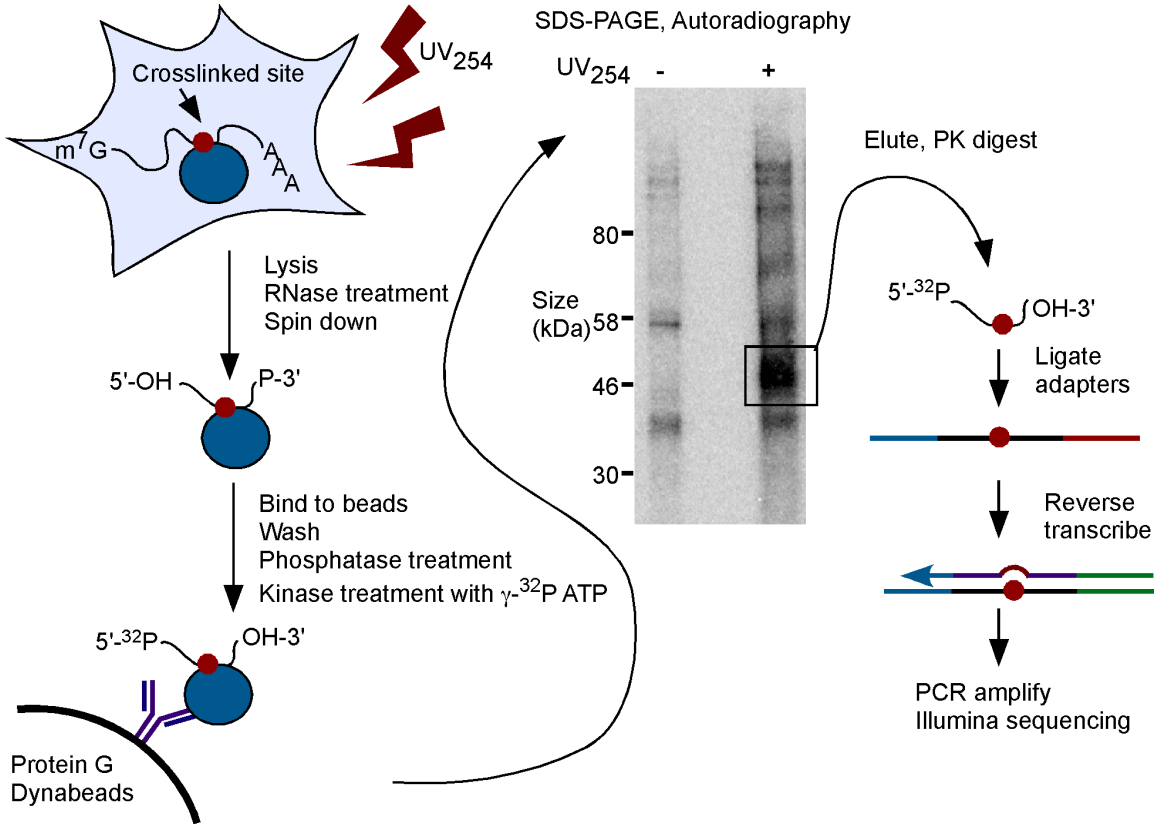


Figure 1. Overview of CLIP procedure

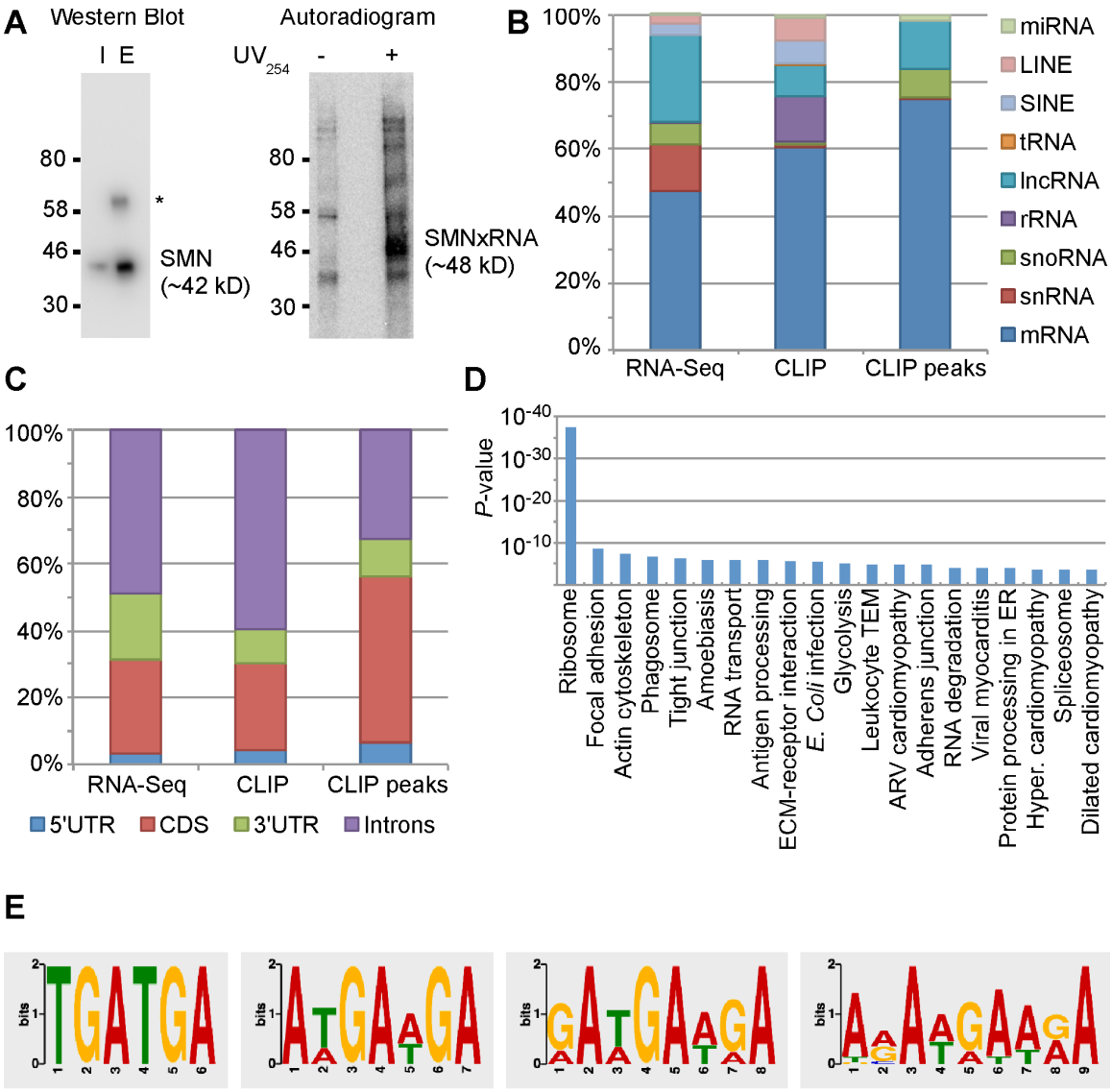


Figure 2. RNA-SMN Crosslinking and Immunoprecipitation

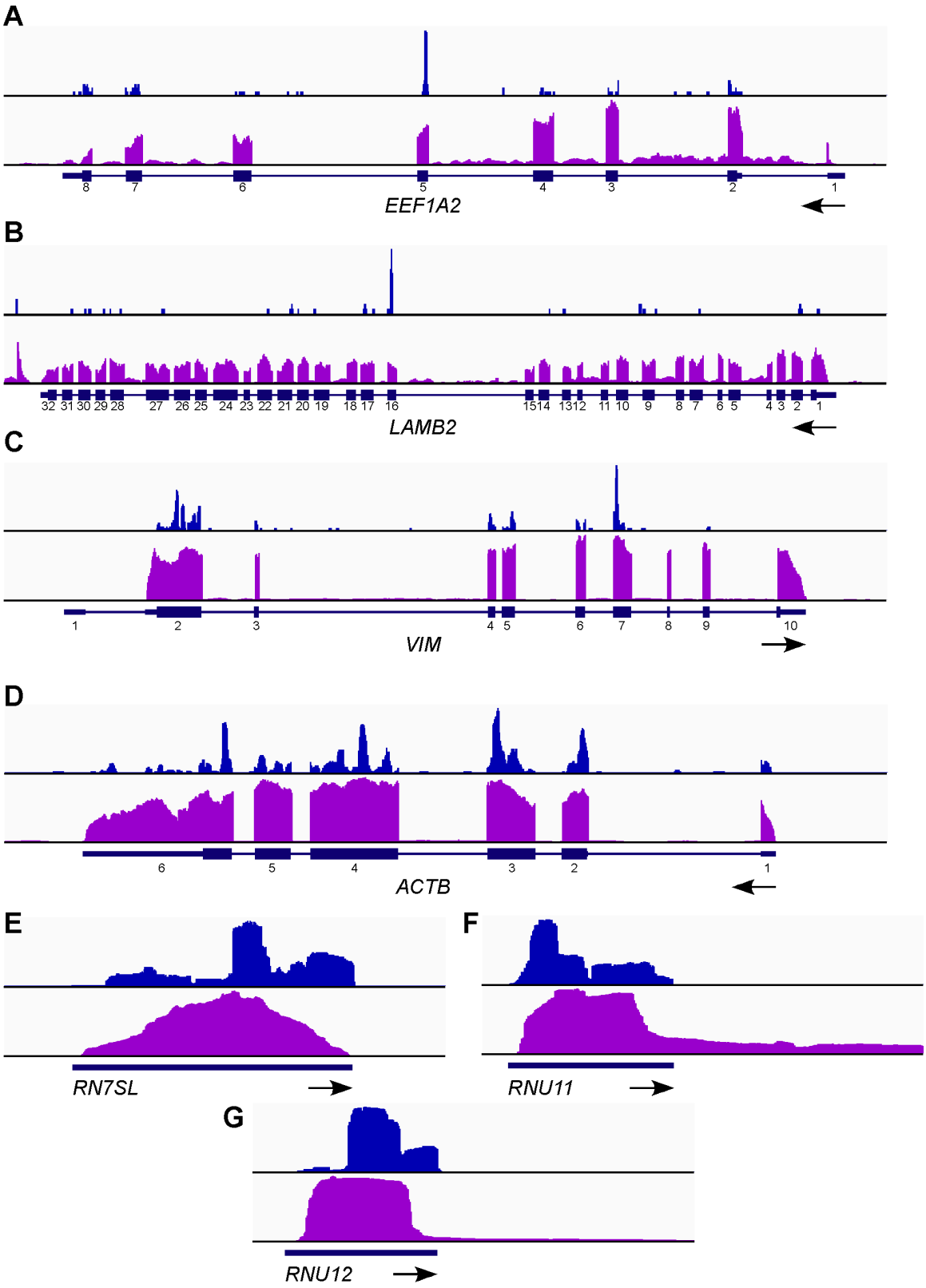


Figure 3. Mapping of CLIP reads to individual genes

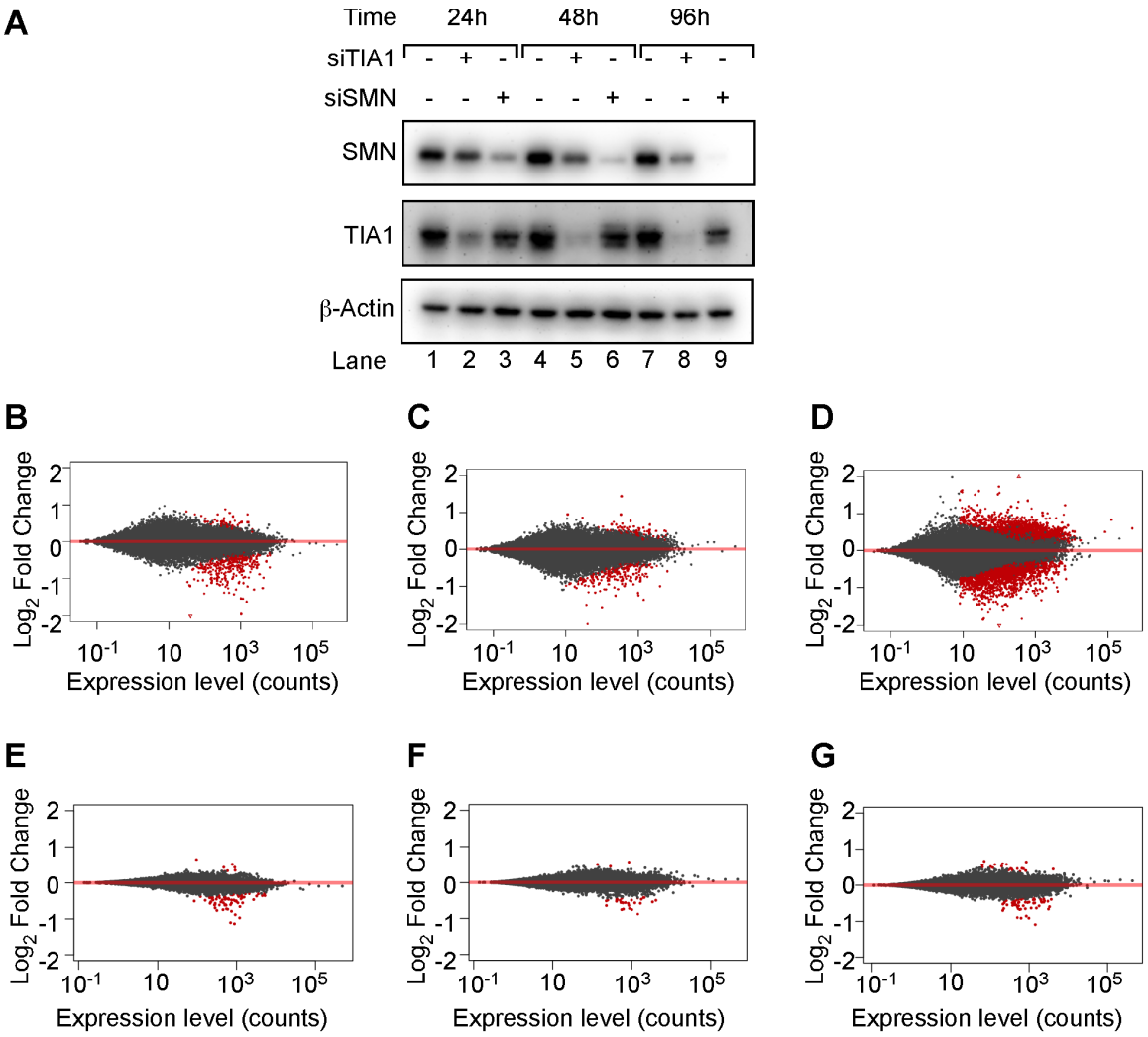


Figure 4. RNA-Seq after siRNA-mediated knockdown of SMN

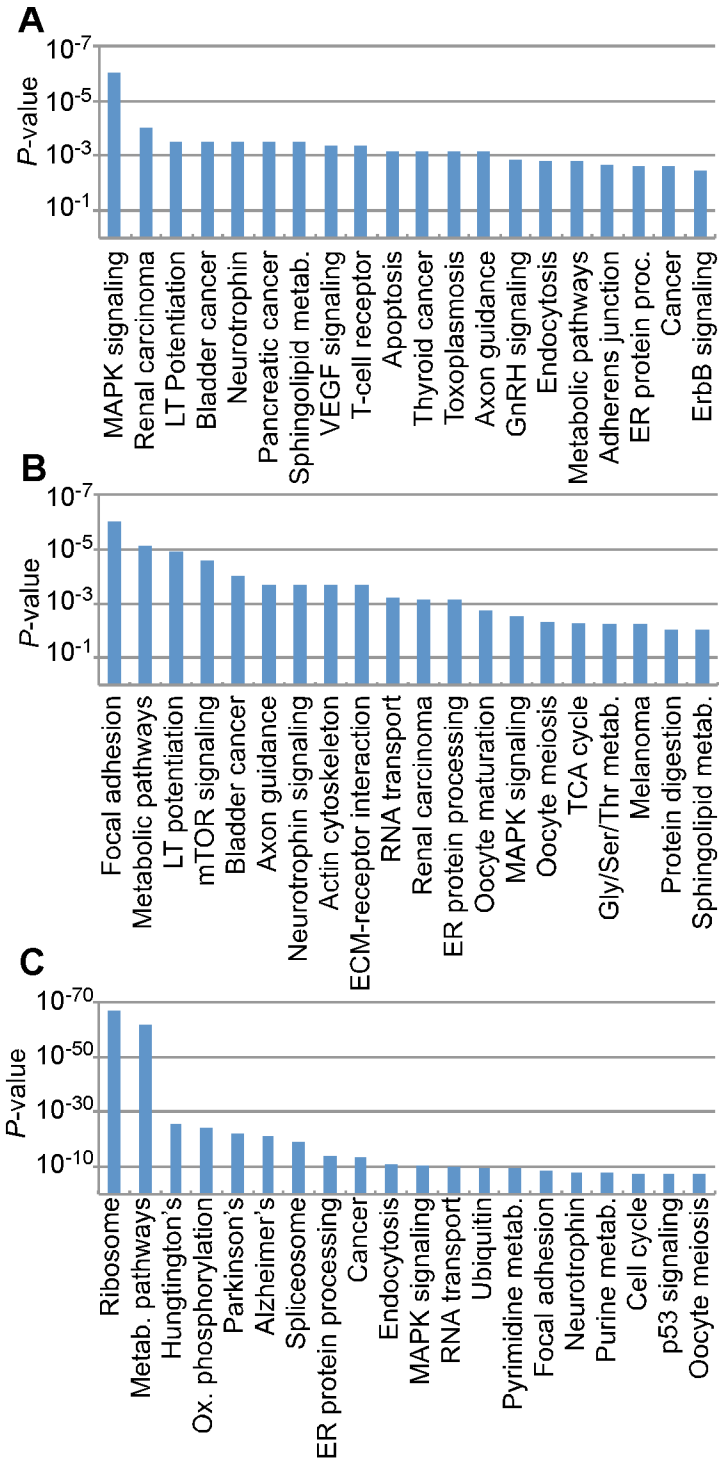


Figure 5. Pathways enriched in genes affected by SMN knockdown

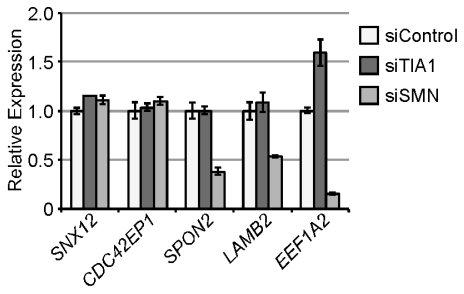


Figure 6. Candidate genes regulated by SMN

Table 1A: Upregulated genes in multiple datasets

	SHSY RNAi	TIA1 RNAi	Zhang et al, motor neurons	Zhang et al, white matter	Maeda et al
SHSY RNAi	1837	3	11	13	32
TIA1 RNAi		15	0	1	0
Zhang et al, motor neurons			118	6	2
Zhang et al, white matter				103	5
Maeda et al					473

No overlaps of 3 or more groups were observed

Table 1B: Downregulated genes in multiple datasets

	SHSY RNAi	TIA1 RNAi	Zhang et al, motor neurons	Zhang et al, white matter	Maeda et al
SHSY RNAi	1880	14	12	9	291
TIA1 RNAi		81	0	0	8
Zhang et al, motor neurons			99	2	10
Zhang et al, white matter				61	7
Maeda et al					2235

SMN1 was observed in SMN RNAi and both Zhang et al datasets
FAHD2A was observed in SMN RNAi, Ng et al, and Zhang et al motor neurons

Table 2. 15 candidate RNAs identified by CLIP and the impact of SMN knockdown on their expression

Ensembl ID	Gene	CLIP score	24h LFC	48h LFC	96h LFC
ENSG00000147164	SNX12	152.82	0.52	0.42	0.61
ENSG00000128283	CDC42EP1	259.36	0.12	0.08	0.55
ENSG00000265972	TXNIP	443.74	0.08	0.36	0.40
ENSG00000101335	MYL9	69.98	0.11	0.00	0.40
ENSG00000099624	ATP5D	115.59	0.22	0.03	0.33
ENSG00000142552	RCN3	103.09	0.06	0.17	0.30
ENSG00000253626	EIF5AL1	15925.16	0.20	0.08	0.22
ENSG00000198003	CCDC151	82168.48	0.05	-0.45	0.12
ENSG00000206503	HLA-A	40.99	-0.02	0.23	0.11
ENSG00000244486	SCARF2	2849.59	-0.17	-0.34	0.10
ENSG00000120029	C10orf76	624.27	0.06	0.04	-0.16
ENSG00000159335	PTMS	129.96	-0.13	0.01	-0.16
ENSG00000172037	LAMB2	211.43	0.06	-0.05	-0.40
ENSG00000159674	SPON2	10473.52	0.14	-0.39	-0.72
ENSG00000101210	EEF1A2	49.74	-0.38	-0.99	-1.82

Yellow highlight-Statistically significant expression changes

APPENDIX C

CHAPTER IV SUPPLEMENTARY TABLES

Supplementary Table 1: CLIP sequencing libraries				
Library	Total reads	Reads > 12bp	Unique reads	Mapped reads
CLIP 1	2914929	2760342	183538	72562
CLIP 2	3146154	3141746	349262	189913
CLIP 3	2810360	2772958	408371	272892

Supplementary Table 2: Genes present in CLIP and Rage et al RIP-Chip

Ensembl Gene ID	Associated Gene Name	Description
ENSG00000034510	TMSB10	thymosin beta 10 [Source:HGNC Symbol;Acc:HGNC:11879]
ENSG00000092841	MYL6	myosin light chain 6 [Source:HGNC Symbol;Acc:HGNC:7587]
ENSG00000100316	RPL3	ribosomal protein L3 [Source:HGNC Symbol;Acc:HGNC:10332]
ENSG00000101335	MYL9	myosin light chain 9 [Source:HGNC Symbol;Acc:HGNC:15754]
ENSG00000112306	RPS12	ribosomal protein S12 [Source:HGNC Symbol;Acc:HGNC:10385]
ENSG00000113140	SPARC	secreted protein, acidic, cysteine-rich (osteonectin) [Source:HGNC Symbol;Acc:HGNC:11219]
ENSG00000118181	RPS25	ribosomal protein S25 [Source:HGNC Symbol;Acc:HGNC:10413]
ENSG00000120725	SIL1	SIL1 nucleotide exchange factor [Source:HGNC Symbol;Acc:HGNC:24624]
ENSG00000122026	RPL21	ribosomal protein L21 [Source:HGNC Symbol;Acc:HGNC:10313]
ENSG00000125977	EIF2S2	eukaryotic translation initiation factor 2 subunit beta [Source:HGNC Symbol;Acc:HGNC:3266]
ENSG00000130255	RPL36	ribosomal protein L36 [Source:HGNC Symbol;Acc:HGNC:13631]
ENSG00000140988	RPS2	ribosomal protein S2 [Source:HGNC Symbol;Acc:HGNC:10404]
ENSG00000142937	RPS8	ribosomal protein S8 [Source:HGNC Symbol;Acc:HGNC:10441]
ENSG00000148303	RPL7A	ribosomal protein L7a [Source:HGNC Symbol;Acc:HGNC:10364]
ENSG00000163430	FSTL1	follistatin like 1 [Source:HGNC Symbol;Acc:HGNC:3972]
ENSG00000166012	TAF1D	TATA-box binding protein associated factor, RNA polymerase I, D [Source:HGNC Symbol;Acc:HGNC:28759]
ENSG00000167526	RPL13	ribosomal protein L13 [Source:HGNC Symbol;Acc:HGNC:10303]
ENSG00000168028	RPSA	ribosomal protein SA [Source:HGNC Symbol;Acc:HGNC:6502]
ENSG00000171858	RPS21	ribosomal protein S21 [Source:HGNC Symbol;Acc:HGNC:10409]
ENSG00000182718	ANXA2	annexin A2 [Source:HGNC Symbol;Acc:HGNC:537]
ENSG00000186468	RPS23	ribosomal protein S23 [Source:HGNC Symbol;Acc:HGNC:10410]
ENSG00000199631	SNORD33	small nucleolar RNA, C/D box 33 [Source:HGNC Symbol;Acc:HGNC:10160]
ENSG00000200084	SNORD68	small nucleolar RNA, C/D box 68 [Source:HGNC Symbol;Acc:HGNC:32729]
ENSG00000206775	SNORD37	small nucleolar RNA, C/D box 37 [Source:HGNC Symbol;Acc:HGNC:10166]
ENSG00000234741	GAS5	growth arrest specific 5 (non-protein coding) [Source:HGNC Symbol;Acc:HGNC:16355]
ENSG00000252481	SCARNA13	small Cajal body-specific RNA 13 [Source:HGNC Symbol;Acc:HGNC:32570]
ENSG00000255717	SNHG1	small nucleolar RNA host gene 1 [Source:HGNC Symbol;Acc:HGNC:32688]
ENSG00000269900	RMRP	RNA component of mitochondrial RNA processing endoribonuclease [Source:EntrezGene;Acc:6023]
ENSG00000270022	RNU12	
ENSG00000276788	SNORD26	small nucleolar RNA, C/D box 26 [Source:HGNC Symbol;Acc:HGNC:10148]
ENSG00000277108	SNORD49B	small nucleolar RNA, C/D box 49B [Source:HGNC Symbol;Acc:HGNC:32721]
ENSG00000277194	SNORD22	small nucleolar RNA, C/D box 22 [Source:HGNC Symbol;Acc:HGNC:10145]
ENSG00000277370	SNORD49A	small nucleolar RNA, C/D box 49A [Source:HGNC Symbol;Acc:HGNC:10189]

Supplementary Table 3: Summary of RNA-Seq libraries				
siRNA	Time	Replicate	Total reads	Mapped reads
siControl	24h	1	27523152	25581717
		2	27011860	25734476
		3	26648822	25097965
	48h	1	24740995	23936976
		2	25413631	24575908
		3	29493699	28455013
	96h	1	22790618	21843487
		2	25354493	24123169
		3	31885211	30392446
siTIA1	24h	1	33339926	30987619
		2	27664336	26210554
		3	22935960	21470938
	48h	1	24984828	24194767
		2	25217183	24415289
		3	30772576	29785296
	96h	1	22920758	21937515
		2	26158856	24889520
		3	29034564	27732772
siSMN	24h	1	24983284	23236101
		2	24975108	23753407
		3	26924239	25092504
	48h	1	23379766	22590497
		2	24879232	24029327
		3	31080328	30027407
	96h	1	23144840	22070042
		2	26679555	25253512
		3	27198256	25790180

Supplementary Table 4: Chemicals, reagents, and enzymes used in this study

Reagent name	Manufacturer	Product Catalog #
Minimum Essential Medium	Life Technologies	11095-080
F12 nutrient mixture	Life Technologies	11765-054
Fetal Bovine Serum	Life Technologies	26140-079
Trypsin 0.25%	Life Technologies	25200-056
PBS	Life Technologies	14190-094
Lipofectamine 2000	Life Technologies	11668-019
TriZOL reagent	Life Technologies	15596-026
UV Stratalinker	Agilent/Stratagene	2400
RIPA buffer	Boston Bioproducts	BP-115
HALT Protease Inhibitor, 100X	Thermo Scientific	78425
RQ1 RNase-free DNase	Promega	M6101
RNase A/T1 mix	Ambion	AM2286
Protein G Dynabeads	Life Technologies	10004D
anti-SMN antibody	BD Biosciences	610646
anti-Gemin2 antibody	Sigma	G6669
anti-Actin antibody	Sigma	A2066
[γ-^{32}P]UTP	Perkin-Elmer	BLU502Z001MC
2X Laemmli buffer	Biorad	161-0737
10% NuPAGE precast gels	Invitrogen	NP0301BOX
Protran BA-85 nitrocellulose	Whatman	10401197
Proteinase K	Sigma	P6556
Tris base (Trizma)	Sigma	T1503-5KG
Sodium Chloride (NaCl)	Sigma	S3014-10KG
Ethylenediaminetetraacetic acid (EDTA)	Sigma	EDS-1KG
Urea	Sigma	S1456-25KG
Phenol:chloroform 1:1	Calbiochem	6810
Glycogen	Roche	10901393001
TruSeq small RNA library preparation kit	Illumina	RS-200-0012
miRVana miRNA isolation kit	Ambion	AM1560
Qiagen RNeasy columns	Qiagen	74104
RNA Nano chip	Agilent	5067-1529
RiboZero Gold rRNA removal kit	Epicentre	MRZG12324
TruSeq mRNA-Seq library preparation kit	Illumina	RS-122-2001
Superscript III reverse transcriptase	Invitrogen	18080-044
FastStart SYBR Green master mix with ROX	Roche	4913914001

Supplementary Table 5: Primers used in this study	
Primer name	Sequence
SNX12-F	CCGCCAAGTAACTTCCTGGAGA
SNX12-R	CACTCAAAGTCACTGTAGCGCC
CDC42EP1-F	GTTGGCAAGCTCAGCTTCGAC
CDC42EP1 R	CTCAGAGGGGAAGGAACCATCC
TXNIP-F	ACTCGTGTCAAAGCCGTTAGGA
TXNIP-R	CAACAGGTGAGAATGAGATGGTGATC
LAMB2-F	TGTCTGCCTTGAGCCTGGTATC
LAMB2-R	ACATCTCTAGCACCAGGACACG
SPON2-F	GACGTAGCGAGTCCTGCTCTTG
SPON2-R	AGATTGTAGACAGCGCCTCAGT
EEF1A2-F	GCGCCTACATCAAGAAGATCGG
EEF1A2-R	TCCTTACGCTCCACCTTCCAG

CHAPTER V

CONCLUSIONS

The Survival Motor Neuron (SMN) protein is a multifunctional protein that is required in all cells. Low levels of SMN result in spinal muscular atrophy (SMA). The most well-characterized function of SMN is in the biogenesis of the spliceosomal small nuclear ribonucleoproteins (snRNPs); however, a number of roles have also been proposed for SMN in diverse processes such as transcription, translation, stress granule formation, mRNA transport, and actin cytoskeleton regulation. Likewise, although SMA is broadly considered to be a disease of motor neurons, secondary symptoms have also been identified in muscle, heart, lungs, and intestine. In this study, we aim to expand the scope of SMN functions, both system-wide and at the molecular level.

In Chapter 2, we identified a previously uncharacterized phenotype of a mild mouse model of SMA (C/C model) which presented as greatly reduced testis size and severe defects in spermatogenesis. As a result, male C/C mice were almost completely infertile and had a ~10-fold reduction in sperm count. We determined that SMN protein expression was extremely high compared to all other tissues tested, indicating an increased requirement for SMN in testes. We characterized the testis transcriptome of C/C mice by RNA-Seq, which revealed drastic alteration of gene expression across a number of critical pathways. We found that SMN levels in C/C testes were developmentally regulated, due to a splicing switch that coincided with the onset of spermatogenesis. Strangely enough, although SMN levels were lowest at early stages of

testis development, we only observed altered gene expression and splicing in sexually mature mice, indicating that either there is a delay between low SMN expression and presentation of the phenotype, or adult testes are affected by loss of SMN expression in other tissues or a small subset of cells within the testis. Our findings underscore the status of SMA as a multi-system disease rather than a strictly neuromuscular disorder and provide new targets for observation in any potential treatment for SMA.

In Chapter 3, we characterized the RNA binding properties of the SMN protein. Although it has long been known that SMN has an RNA binding domain, only a limited amount of research has been done characterizing the specificity of RNA- SMN interactions. We performed iterative in vitro selection (Systematic Evolution of Ligands through Exponential Enrichment or SELEX) to identify high-affinity RNA targets of the SMN protein. Analysis of selected sequences revealed high incidence of GUG, UGC, UGG, and UUG trinucleotides, as well as a GUGCG motif that was strongly enriched. However, mutation of sequences outside of the GUGCG also had a strong effect on binding affinity, indicating that multiple factors, including RNA structure, influenced binding. We demonstrate that both affinity and specificity of RNA-SMN interaction are influenced by salt concentrations. Our results bring new perspective to the mode of action of SMN and may suggest new functions of SMN in RNA metabolism.

In Chapter 4, We used in vivo crosslinking and immunoprecipitation (CLIP) to identify direct interactions between SMN and RNA in living cells. We observed an enrichment of mRNAs in the CLIP-purified RNAs and identify several candidate genes. Although CLIP did not identify the same target motif as SELEX, such a result is not surprising given the increased complexity of in vivo interactions. We identified a variety

of RNA targets of SMN with an enrichment of mRNAs participating in a number of pathways, including ribosome function and actin cytoskeleton regulation. Overall, we confirm that SMN is a bona fide RNA binding protein that exhibits sequence specificity, shedding new light on a number of critical functions of this essential protein.