2016

An engineered U1 snRNP redefines SMN1 exon 7 carrying a pathogenic mutation at the splice donor site

José Bruno Del Rio-Malewski
Iowa State University

An engineered U1 snRNP redefines *SMN1* exon 7 carrying a pathogenic mutation at the splice donor site

by

José Bruno Del Rio-Malewski

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER of SCIENCE

Major: Genetics

Program of Study Committee:
Ravindra N. Singh, Major Professor
Natalia Singh
Jason Ross

Iowa State University

Ames, Iowa

2016

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank Dr. Ravindra Singh, Dr. Natalia Singh, and Dr. Jason Ross for serving as my POS Committee members. Also, I offer my appreciation to the members of the Singh lab for their guidance and technical support.

ABSTRACT

The human genome has two copies of the *Survival Motor Neuron* (*SMN)* gene, *SMN1* and *SMN2*. The absence of *SMN1* results in spinal muscular atrophy (SMA), a leading genetic disease among infants and children. *SMN2* cannot substitute for *SMN1* because of exon 7 skipping. While restoring exon 7 inclusion in *SMN2* is a promising approach for SMA therapy, the method has limitations for specific SMA patients carrying a single *SMN2* allele. Recently, a severe SMA patient carrying a single *SMN2* allele as well as a *SMN1* allele with a deleterious G-to-C mutation (G1C) at the splice donor site of intron 7 has been reported. In this study, we show that an engineered U1 small nuclear ribonucleoprotein (eU1 snRNP) with extended base pairing at the 5′ splice site of intron 7 prevents skipping of exon 7 of *SMN1* carrying the G1C mutation. We also show that eU1 snRNA promotes expression of the full length SMN protein from the *SMN1* allele carrying the G1C mutation. We further demonstrate that eU1 snRNAs annealing to 5′ splice site-like sequences downstream of the canonical intron 7 splice donor site can activate these sites and promote the inclusion of an extended exon 7 from *SMN1* carrying the G1C mutation. Such findings provide a novel method for correcting aberrant splicing in SMA.

CHAPTER 1. GENERAL INTRODUCTION

RNA splicing is a complex process that involves macromolecular machinery and cis-regulatory elements. A key component to RNA splicing is the U1 small nuclear ribonucleoprotein (U1 snRNP), which is composed of a 165 bp snRNA strand containing four stem-loop structures that interact with at least protein factors U1-A, U1-70K, U1-C, and Smith (Sm) proteins (Rogalska et al., 2016). The Sm proteins, which are common to all snRNP complexes, include SmB, SmD1, SmD2, SmD3, SmE, SmF, and SmG (Figure 1; Buratti & Baralle, 2010). During splicing, the U1-C component of the U1 snRNP recognizes the 5′ splice site (5′ss) of an exon, which typically contains a GU residue, before pre-mRNA/U1 snRNA base pairing takes place (Du and Rosbash, 2002). In vertebrates and *S. cerevisiae*, U1 snRNAs contain two pseudouridines in a conserved AUACΨΨACCU sequence at the 5′ end (Wu et al., 2011). Downstream of the 5′ss, protein factor SF1/mBBP recognizes a branchpoint adenine and recruits the U2AF protein. The U2 snRNP is recruited to the branchpoint site and displaces SF1/mBBP and U2AF (Buratti & Baralle, 2010). In the process, the U1-70K protein component of the U1 snRNP, which binds to a consensus GAUCANGAAG motif in stem-loop I, interacts with exonic splicing enhancers to promote exon definition (Surowy et al., 1989; Rogalska et al., 2016). The stem-loop IV structure of the U1 snRNP interacts with the SF3A1 protein component of U2 to promote intron definition (Rogalska et al., 2016). The resulting cross-exon complex recruits a tri-snRNP containing U4, U5, and U6 to form the spliceosome complex that catalyzes intron excision and exon ligation (Wahl et al., 2016). Protein factors further regulate splicing by binding to exonic or intronic splicing enhancers and silencers that define splice sites and enable alternative splicing that can allow a gene to express more than one type of transcript (Wang & Burge, 2008).

Diseases that involve the U1 snRNP emphasize the impact the ribonucleoprotein has on an organism. Amyotrophic lateral sclerosis (ALS), a fatal adult motor neuron disease, is partly caused by U1 snRNP loss of function when mutations in the RNA-binding protein FUS inhibit association between FUS and U1 snRNP. FUS carrying mutations such as R495X and P525L mislocalize in the cytoplasm and have reduced interactions with U1 snRNP, while FUS carrying mutations such as R521G have reduced binding to U1 snRNP (Calvo et al., 2014; Yu et al., 2015; Sun et al., 2015). Cerebro-costo-mandibular syndrome, a malformation disorder, is caused by deleterious mutations in *SNRPB*, the gene that encodes snRNP components SmB and SmB′. Mutations in the second intron of *SNRPB* such as c.164G>C, c.164G>T, c.165G>C, c.166G>C, and c.213+57C>A favor the inclusion of an alternative exon that carries a premature stop codon and reduces the translation of functional SmB and SmB′ for assembling snRNPs (Lynch et al., 2014; Bacrot et al., 2015). Alzheimer's disease (AD) is another condition that may partly involve disrupting U1 snRNP splicing activities. U1 snRNP components U1-70K and U1A are among the proteins that aggregate in the brain tissue of AD patients, while splicing deficiency, premature cleavage, and polyadenylation on cryptic poly(A) sites increase in frequency (Bai et al., 2013). Additionally, HMGA1a, a protein that has increased expression in the brain tissue of sporadic AD patients, forms a complex with the U1 snRNP that inhibits dissociation of U1 snRNP from positions such as the 5′ splice site of exon 5 in *PS2* and impedes RNA splicing (Ohe & Mayeda, 2010). In autoimmune disorders such as systemic lupus erythematosus (SLE), the U1 snRNP itself can trigger inflammation after T and B cells are exposed to the U1-70K component from apoptotic cells (Greidinger et al., 2002; Shin et al., 2012).

In genetic diseases, 50 to 60% of deleterious mutations disrupt RNA splicing (Wang & Cooper, 2007). Familial dysautonomia (FD), an autosomal recessive disorder in which low levels of IKAP protein cause progressive degeneration of the sensory and autonomic nervous systems, often appears to be the result of a T-to-C mutation at base pair 6 in intron 20 of *IKBKAP* (Slaugenhaupt et al., 2001). The T-to-C mutation apparently weakens the already weak splicing signals surrounding exon 20 in *IKBKAP* and promotes exon skipping (Ibrahim et al., 2007). Individuals afflicted with Bardet-Biedl syndrome (BBS), a malformation syndrome, often carry mutations at *BBS1,* a gene necessary for proper assembly of the BBSome, which is involved in transporting proteins to cilia. In one family affected by BBS, a G-to-A mutation in the last position of exon 5 in *BBS1*, which is part of the splice donor site, was found to be triggering exon 5 skipping (Schmid et al., 2011). Netherton syndrome, a severe autosomal recessive skin disorder, can be caused by a variety of mutations at *SPINK5*. Mutations affecting splice donor sites in *SPINK5* have included a G-to-A mutation in the last position of exon 1, a T-to-A mutation in the second position of intron 2, a G-to-A mutation at the fifth position of intron 2, and a G-to-A mutation in the first position of intron 23 (Bitoun et al., 2002). Splice acceptor sites in *SPINK5* have been disrupted by mutations such as a G-to-A mutation in the last position of intron 20 and a G-to-A mutation at the last position of intron 17 (Bitoun et al., 2002). Additionally, a C-to-T mutation in the ninth position of exon 11 in *SPINK5* was shown promoting exon skipping by strengthening hnRNPA1 binding sites, while weakening Tra2β binding sites (Dal Mas, Fortugno, et al., 2015). Cystic fibrosis is triggered by deleterious *CFTR* alleles that alter chloride secretion across the apical membrane of epithelial cells. In *CFTR*, splice donor sites may be disrupted by mutations such as a G-to-T mutation at the first position of intron 5, a G-to-A mutation at the fifth position of intron 14b, and a G-to-A mutation at the

first position of intron 23 (Fanen et al., 1992). Also, a G-to-A mutation at the last position of intron 10 in *CFTR* disrupts a splice acceptor site, while an A-to-G mutation within intron 17a at position -26 generates an alternative splice acceptor site (Fanen et al., 1992). In intron 11 of *CFTR*, an A-to-G mutation at position 1.6 kb produces a new exon (Chillon et al., 1995). In cases of cancer, mutations affecting splicing have been detected in tumor suppressor genes. Notably, cancer patients have carried mutations such as a G-to-A mutation in the first position of intron 7 in *hSNF5*, a G-to-T mutation at the last position of intron 3 in *APC*, and a G-to-A mutation in the last position of exon 4 in *CCND1* (Venables, 2004; Friedl & Aretz, 2005; David & Manley, 2010; Eaton et al., 2011).

Mutations that induce aberrant splicing are especially prevalent in cases of spinal muscular atrophy (SMA). SMA is both the second most common autosomal recessive disorder and a leading hereditary cause of infant death (Lefebvre et al., 1995; Prior, 2007). The disorder is caused by deletions or mutations within the *SMN1* gene, which typically encodes the Survival Motor Neuron protein (SMN) that has multiple, vital functions in cellular metabolism (Singh & Singh, 2011). SMA patients suffer from not having enough SMN protein, which causes anterior horn cells of the spinal cord to degenerate and results in progressive limb and trunk paralysis (Lefebvre et al., 1995). *SMN2*, an almost identical duplicate of *SMN1* that is unique to humans, cannot compensate for the loss of *SMN1* due to a C-to-T mutation at the 6th position of exon 7 that promotes exon 7 skipping and consequently produces low levels of full-length SMN protein (Lefebvre et al., 1995; Lorson et al., 1999; Rochette et al., 2001). In conjunction with a UUA codon that is located between positions 43 and 45 in human exon 7, the C-to-T mutation may have evolved to cap SMN levels after the *SMN* gene duplicated in the human lineage (Singh et

al., 2004). Exon 7 inclusion in *SMN* genes is also mediated by cis-enhancers and silencers such as SF2/ASF and ISS-N1, respectively (Figure 2; Singh et al., 2007).

While *SMN2* can still produce enough full-length transcript for most types of cells, *SMN2* cannot support motor neurons and muscle cells (Boyer et al., 2013). Since almost all SMA patients at least still carry *SMN2*, research for developing SMA therapy has often involved trying to correct *SMN2* splicing to include exon 7 (Singh and Singh, 2011; Howell et al., 2014). SMA patients who retain a copy of *SMN1* carry a variety of mutations that cause a loss of SMN function. While null alleles of *SMN1* are often caused by missense mutations or deletions, relatively rare *SMN1* mutations that cause splice-site changes have also been reported. One recently identified *SMN1* mutation is a G-to-C substitution at the exon 7 splice donor site that induces exon 7 skipping during RNA splicing (Ronchi et al., 2015).

One potential method for correcting splicing in deleterious *SMN1* alleles such as the mutant carrying the G-to-C substitution at the exon 7 splice donor site is to increase base pairing between U1 snRNA and the 5′ss of exon 7. In *SMN1* mutants carrying mutations at exon 7, mutated U1 snRNA can promote splicing in the presence of RNA structures that typically inhibit U1 snRNP binding at the 5′ss of exon 7 (Figure 3; Singh et al., 2007). Similarly, in *BBS1* mini-genes carrying mutations at the splice donor site of exon 5 that induce exon skipping, U1 snRNAs designed to base pair with mutated splice donor sites have been shown to restore exon 5 (Schmid et al., 2013). Alternatively, U1 snRNAs designed to bind to intronic sequences downstream of a canonical splice site can also correct splice defects. In a *SPINK5* mutant that exhibits exon 11 skipping, *F9* mutants that exhibit exon 5 skipping, *CFTR* mutants that exhibit

exon 12 skipping, and in *SMN2,* U1 snRNAs that target sequences downstream of splice donor sites can inhibit exon skipping by apparently recruiting splicing factors onto an upstream exon (Dal Mas, Fortugno, et al., 2015; Alanis et al., 2012; Dal Mas, Rogalska, et al., 2015).

In Chapter 2, we demonstrate that, in the context of the *SMN1* mutant minigene carrying a G-to-C substitution at the exon 7 splice donor site, U1 snRNAs modified to bind to the mutation site can prevent exon 7 skipping. Modified U1 snRNAs targeting the mutated 5′ splice site of exon 7 both strongly promote intron 7 retention and weakly activate cryptic splice sites that allow elongated versions of exon 7 to be included during splicing. Additionally, we show that U1 snRNAs designed to bind to potential splice sites downstream of the mutated exon 7 splice donor site can promote the inclusion of elongated versions of exon 7 greatly enough to enable full-length SMN protein expression. Chapter 3 describes the general conclusions.

REFERENCES

Alanis EF, Pinotti M, Dal Mas A, Balestra D, Cavallari N, Rogalska ME, Bernardi F, Pagani F (2012). An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. Hum Mol Genet. *21(11)*, 2389-2398. doi: 10.1093/hmg/dds045.

Bai B, Hales CM, Chen PC, Gozal Y, Dammer EB, Fritz JJ, et al (2013). U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. Proc Natl Acad Sci USA *110(41)*,16562–16567. doi: 10.1073/pnas.1310249110.

Bacrot S, Doyard M, Huber C, Alibeu O, Feldhahn N, Lehalle D, Lacombe D, Marlin S, Nitschke P, Petit F, Vazquez MP (2015). Mutations in SNRPB, Encoding Components of the Core Splicing Machinery, Cause Cerebro-Costo-Mandibular Syndrome. Human Mutation *36(2)*, 187-190. doi: 10.1002/humu.22729.

Bitoun E, Chavanas S, Irvine AD, Lonie L, Bodemer C, Paradisi M, Hamel-Teillac D, Ansai SI, Mitsuhashi Y, Taïeb A, de Prost Y (2002). Netherton syndrome: disease expression and spectrum of SPINK5 mutations in 21 families. Journal of investigative dermatology *118(2)*, 352-361. doi: 10.1046/j.1523-1747.2002.01603.x.

Boyer JG, Ferrier A, Kothary R (2013). More than a bystander: the contributions of intrinsic skeletal muscle defects in motor neuron diseases. Front Physiol. *4*, 356. doi: 10.3389/fphys.2013.00356.

Buratti E, Baralle D (2010). Novel roles of U1 snRNP in alternative splicing regulation. RNA Biology 7(4), 412-419. PMID: 20523112.

Calvo A, Moglia C, Canosa A, Brunetti M, Barberis M, Traynor BJ, Carrara G, Valentini C, Restagno G, Chiò A (2014). A de novo nonsense mutation of the FUS gene in an apparently familial amyotrophic lateral sclerosis case. Neurobiology of Aging *35(6)*, 1513-e7. doi:10.1016/j.neurobiolaging.2013.12.028.

Chillon M, Dörk T, Casals T, Gimenez J, Fonknechten N, Will K, Ramos D, Nunes V, Estivill X (1995). A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+ 1.6 kbA--> G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. American journal of human genetics *56(3)*, 623. PMCID: PMC1801150.

Dal Mas A, Fortugno P, Donadon I, Levati L, Castiglia D, Pagani F (2015). Exon-Specific U1s Correct SPINK5 Exon 11 Skipping Caused by a Synonymous Substitution that Affects a Bifunctional Splicing Regulatory Element. Human Mutation *36(5)*, 504-12. doi: 10.1002/humu.22762.

Dal Mas A, Rogalska ME, Bussani E, Pagani F (2015). Improvement of SMN2 pre-mRNA processing mediated by exon-specific U1 small nuclear RNA. American Journal of Human Genetics *96(1)*, 93-103. doi: 10.1016/j.ajhg.2014.12.009.

David CJ, Manley JL (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev. *24*, 2343-2364. doi: 10.1101/gad.1973010.

Du H, Rosbash M (2002). The U1 snRNP protein U1C recognizes the 5′ splice site in the absence of base pairing. Nature *419(6902)*, 86-90. doi: 10.1038/nature00947.

Eaton KW, Tooke LS, Wainwright LM, Judkins AR, Biegel JA (2011). Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors. Pediatric blood & cancer *56(1)*, 7-15. doi: 10.1002/pbc.22831.

Fanen P, Ghanem N, Viduad M, Besmond C, Martin J, Costes B, Plassa F, Goossens M (1992). Molecular Characterization of Cystic Fibrosis: 16 Novel Mutations Identified by Analysis of the Whole Cystic Fibrosis Conductance Transmembrane Regulator (CFTR) Coding Regions and Splice Site Junctions. Genomics *13*, 770-776. doi: 10.1016/0888-7543(92)90152-I.

Friedl W, Aretz S (2005). Familial adenomatous polyposis: experience from a study of 1164 unrelated german polyposis patients. Hereditary Cancer in Clinical Practice *3(3)*, 95-114. doi: 10.1186/1897-4287-3-3-95.

Greidinger EL, Foecking MF, Schäfermeyer KR, Bailey CW, Primm SL, Lee DR, Hoffman RW (2002). T cell immunity in connective tissue disease patients targets the RNA binding domain of the U1-70kDa small nuclear ribonucleoprotein. The Journal of Immunology *169(6)*, 3429-3437. doi: 10.4049/ jimmunol.169.6.3429.

Howell MD, Singh NN, Singh RN (2014). Advances in therapeutic development for spinal muscular atrophy. Future Med Chem. *6(9)*, 1081-99. doi: 10.4155/fmc.14.63.

Ibrahim EC, Hims MM, Shomron N, Burge CB, Slaugenhaupt SA, Reed R (2007). Weak definition of IKBKAP exon 20 leads to aberrant splicing in familial dysautonomia. Human Mutation *28(1)*, 41-53. doi: 10.1002/humu.20401.

Lefebvre S, Bürglen 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. *180(1)*, 155-65. doi:10.1016/0092-8674(95)90460-3.

Lorson CL, Hahnen E, Androphy EJ, Wirth B (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad Sci USA. *96(11)*, 6307-11. doi: 10.1073/pnas.96.11.6307.

Lund M, Kjems J (2002). Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end. RNA *8(2)*, 166-79. doi: 10.1017/S1355838202010786.

Lynch DC, Revil T, Schwartzentruber J, et al (2014). Disrupted auto-regulation of the spliceosomal gene SNRPB causes cerebro–costo–mandibular syndrome. Nature Communications *5*, 4483. doi: 10.1038/ncomms5483.

Ohe K, Mayeda A (2010). HMGA1a trapping of U1 snRNP at an authentic 5′ splice site induces aberrant exon skipping in sporadic Alzheimer's disease. Molecular and cellular biology *30(9)*, 2220-2228. doi: 10.1128/MCB.00114-10.

Prior TW (2007). Spinal muscular atrophy diagnostics. J Child Neurol. 22(8), 952-6. doi: 10.1177/0883073807305668.

Rochette CF, Gilbert N, Simard LR (2001). *SMN* gene duplication and the emergence of the *SMN2* gene occurred in distinct hominids: *SMN2* is unique to *Homo sapiens*. Hum Genet 108, 255-266. doi: 10.1007/s004390100473.

Rogalska ME, Tajnik M, Licastro D, Bussani E, Camparini L, Mattioli C, Pagani F (2016). Therapeutic activity of modified U1 core spliceosomal particles. Nature Communications *7*, 11168. doi: 10.1038/ncomms11168.

Ronchi D, Previtali SC, Sora MG, Barera G, Del Menico B, Corti S, Bresolin N, Comi GP (2015). Novel splice-site mutation in SMN1 associated with a very severe SMA-I phenotype. J Mol Neurosci. *56(1)*, 212-215. doi: 10.1007/s12031-014-0483-4.

Schmid F, Glaus E, Barthelmes D, Fliegauf M, Gaspar H, Nürnberg G, Nürnberg P, Omran H, Berger W, Neidhardt J (2011). U1 snRNA-mediated gene therapeutic correction of splice defects caused by an exceptionally mild BBS mutation. Human mutation *32(7)*, 815-824. doi: 10.1002/humu.21509.

Schmid F, Hiller T, Korner G, Glaus E, Berger W, Neidhardt J (2013). A gene therapeutic approach to correct splice defects with modified U1 and U6 snRNPs. Human Gene Therapy *24(1)*, 97-104. doi: 10.1089/hum.2012.110.

Shin MS, Kang Y, Lee N, et al (2012). U1-snRNP activates the NLRP3 inflammasome in human monocytes. Journal of Immunology *188(10)*, 4769-4775. doi: 10.4049/jimmunol.1103355.

Singh NN, Androphy EJ and Singh RN (2004). In vivo selection reveals features of combinatorial control that defines a critical exon in the spinal muscular atrophy genes. RNA *10*, 1291-1305. doi: 10.1261/rna.7580704.

Singh NN, Singh RN (2011). Alternative splicing in spinal muscular atrophy underscores the role of an intron definition model. RNA Biology *8(4)*, 600-606. doi: 10.4161/rna.8.4.16224.

Slaugenhaupt SA, Blumenfeld A, Gill SP, Leyne M, Mull J, Cuajungco MP, Liebert CB, Chadwick B, Idelson M, Reznik L, Robbins CM (2001). Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. The American Journal of Human Genetics *68(3)*, 598-605. doi: 10.1086/318810.

Sun S, Ling SC, Qiu J, Albuquerque CP, Zhou Y, Tokunaga S, Li H, Qiu H, Bui A, Yeo GW, Huang EJ (2015). ALS-causative mutations in FUS/TLS confer gain and loss of function by

altered association with SMN and U1-snRNP. Nature communications *6*, 6171. doi: 10.1038/ncomms7171.

Surowy CS, Van Santen VL, Scheib-Wixted SM, Spritz RA (1989). Direct, Sequence-Specific Binding of the Human U1-70K Ribonucleoprotein Antigen Protein to Loop I of U1 Small Nuclear RNA. Molecular and cellular biology *9(10)*, 4179-4186. doi: 10.1128/MCB.9.10.4179.

Venables JP (2004). Aberrant and Alternative Splicing in Cancer. Cancer Research *64*, 7647-7654. doi: 10.1158/0008-5472.CAN-04-1910.

Wahl MC, Will CL, Lührmann R (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell. *136(4)*, 701-718. doi: 10.1016/j.cell.2009.02.009.

Wang Z, Burge CB (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA *14*, 802-813. doi: 10.1261/rna.876308.

Wang GS, Cooper TA (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet. *8(10)*, 749-61. doi: 10.1038/nrg2164.

Wu G, Yu AT, Kantartzis A, Yu YT (2011). Functions and mechanisms of spliceosomal small nuclear RNA pseudouridylation. Wiley Interdisciplinary Reviews: RNA *2(4)*, 571-81. doi: 10.1002/wrna.77.

Yu Y, Chi B, Xia W, Gangopadhyay J, Yamazaki T, Winkelbauer-Hurt ME, Yin S, Eliasse Y, Adams E, Shaw CE, Reed R (2015). U1 snRNP is mislocalized in ALS patient fibroblasts bearing NLS mutations in FUS and is required for motor neuron outgrowth in zebrafish. Nucleic Acids Research *43(6)*, 3208-3218. doi: 10.1093/nar/gkv157.

FIGURE LEGENDS

Figure 1. Diagrammatic represenation of U1 snRNP annealed to the 5′ splice site (5′ss) of exon 7 in *SMN1*. The U1 snRNP contains four stem-loop structures that interact with U1-A, U1-70K, U1-C, and the Smith proteins SmB, SmD1, SmD2, SmD3, SmE, SmF, and SmG (Rogalska et al. 2016; Buratti & Baralle, 2010). The boxes represent exon 7 (E7) and exon 8 (E8) of *SMN1*, while the line connecting the two boxes represents intron 7. Circles and ovals represent U1 snRNP protein factors. The upper case letters in the *SMN1* sequence indicate exonic nucleotides while lower case letters indicate intronic nucleotides. Numbering starts from the first position of intron 7.

Figure 2. Diagrammatic representation of known splicing regulatory elements in exon 7 and intron 7 of *SMN1*. Binding sites for SF2/ASF and Tra2-β1 are positive (+) cis-elements that promote exon 7 inclusion. Negative (-) cis-elements Extinct, the hnRNP A1 binding site, 3′-Cluster, and ISS-N1 promote exon 7 skipping. TSL2 is an RNA stem-loop structure that promotes exon 7 skipping (Singh et al. 2007). Upper case letters indicate exonic sequences while lower case letters indicate intronic sequences. Numbering starts from the first position of intron 7.

Figure 3. Diagrammatic representation of engineered U1 snRNA that promoted exon 7 inclusion in *SMN1* minigenes that carried mutations strengthening TSL2. The engineered U1 snRNA also promoted exon 7 inclusion in *SMN2* minigenes (Singh et al. 2007). Mutated nucleotides as compared to the wild-type snRNA are shown in red. Black circles indicate Watson-Crick and

wobble base pairs formed between U1 snRNAs and the 5′ss of exon 7. Upper case letters indicate exonic sequences while lower case letters indicate intronic sequences.

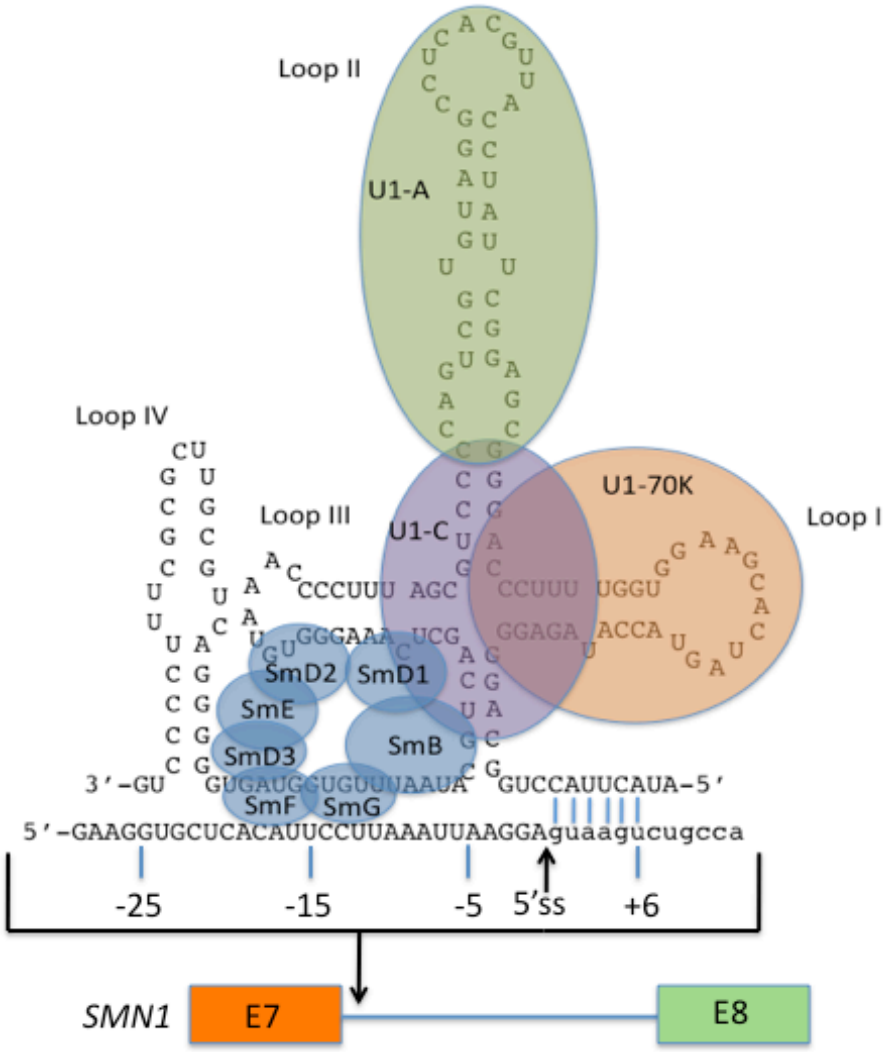


Figure 1. U1 snRNP and RNA Splicing.

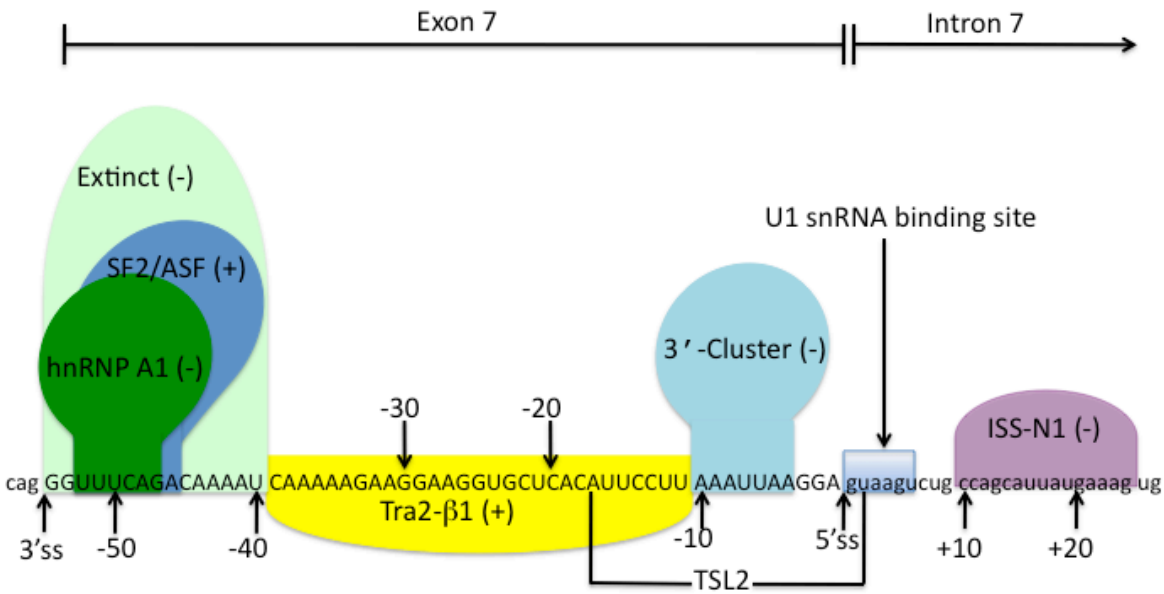


Figure 2. Diagrammatic representation of known splicing regulatory elements in exon 7 and intron 7 of *SMN1*.

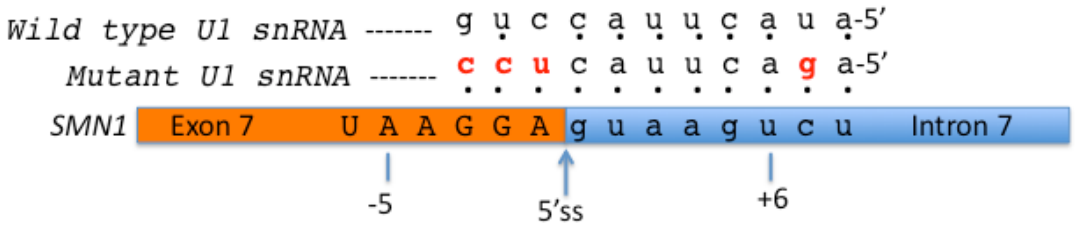


Figure 3. Engineered U1 snRNA that promotes exon 7 inclusion in mutated *SMN1*.

CHAPTER 2. AN ENGINEERED U1 SNRNP REDEFINES *SMN1* EXON 7 CARRYING A PATHOGENIC MUTATION AT THE SPLICE DONOR SITE

José Bruno Del Rio-Malewski[1,2], Natalia N. Singh[2], Matthew D. Howell[2] and Ravindra N. Singh[1,2]

A manuscript in preparation

---

[1]Interdepartmental Genetics and Genomics, Iowa State University, Ames, Iowa 50011

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

*Corresponding author (singhr@iastate.edu).

ABSTRACT

Humans carry two highly similar copies of the *survival motor neuron* (*SMN)* gene, *SMN1* and *SMN2*. Most spinal muscular atrophy (SMA) patients lack a functional copy of *SMN1*. *SMN2* is prone to exon 7 skipping and cannot compensate for the loss of *SMN1*. Previous research in developing treatments for SMA has focused on preventing exon 7 skipping in *SMN2*, but the approach has limitations for patients carrying a single *SMN2* and pathogenic mutations that disable splice sites of *SMN1* exon 7. Engineered U1 snRNAs (eU1) that have extended base pairing with the 5′ splice site of exon 7 have been shown to promote exon 7 inclusion in *SMN1* containing exonic mutations that exhibit exon 7 skipping. In this study, eU1 snRNAs that have extended base pairing with the 5′ splice site of intron 7 or base pair with sequences downstream of the canonical splice donor site are shown to also inhibit exon 7 skipping in a *SMN1* allele containing a deleterious G-to-C (G1C) mutation at the first position of intron 7. These eU1 snRNAs prevent exon 7 skipping in the mutant *SMN1* allele by promoting intron 7 retention and by activating cryptic splice donor sites downstream of the canonical splice donor site. eU1 snRNAs also promote expression of full length SMN protein in the mutant *SMN1* allele. Our findings open up yet another avenue for the treatment of SMA patients carrying pathogenic mutations at splice sites.

INTRODUCTION

Alternative pre-mRNA splicing optimizes the coding potential of a genome by enabling single genes to produce multiple transcripts/proteins (Xing and Lee, 2007; Lee and Rio, 2016). More than 95% of human genes containing two or more exons undergo alternative splicing (Nilsen and Graveley, 2010). For most human genes, GU and AG residues at the beginning and the end of an intron serve as the 5′ splice site (5′ss) and 3′ss, respectively (Burge et al., 1999). A combinatorial control of cis-elements including exonic or intronic splicing enhancers (ESEs or ISEs) and silencers (ESSs or ISSs) define splice sites from among abundantly available GU and AG residues within pre-mRNA (Wang and Burge, 2008). Generally, serine and arginine-rich proteins (SR proteins) and heteronuclear ribonucleoproteins (hnRNPs) promote and suppress exon inclusion by binding to ESEs/ISEs and ESSs/ISSs, respectively. However, there are exceptions to the rule as SR proteins and hnRNPs can have negative and positive roles, respectively. RNA structures provide another layer of splicing regulation by sequestering or exposing regulatory motifs (Buratti and Baralle, 2004; Warf and Berglund, 2010; McManus and Graveley, 2011). A RNA sequence can possess multiple interchangeable structures and a particular structural confirmation can be stabilized/favored by a specific protein(s). However, very little is known how information stored in RNA structures is utilized to regulate pre-mRNA splicing.

GURAGU residues at the first six positions and the YAG residues at the last three positions of an intron define a typical 5′ss and 3′ss in human pre-mRNA, respectively. In addition, the definition of the 3′ss requires presence of a polypyrimidine tract (PPT) and a branch point adenosine residue upstream of the YAG sequence. Other than the authentic splice sites,

pre-mRNAs harbor cryptic splice sites with potential to be activated upon mutations within these cryptic sites and/or at the nearby authentic splice sites. It has been shown that the non-authentic 5′ss within an exon are suppressed by overlapping ESSs (Wang et al., 2004). Suppression of non-authentic intronic splice sites appears to involve a complex network of interactions with overlapping ISSs (Wang et al., 2013). Genetic diseases caused by point mutations triggering splicing changes without affecting the coding sequence underscore the critical role of splicing regulatory elements (Wang et al., 2012; Xiong et al., 2015).

Splicing is catalyzed by the spliceosome, a macromolecular machine in which five small ribonucleoproteins (U1, U2, U4, U5 and U6 snRNPs) play an indispensable role (Matlin and Moore, 2007; Wahl et al., 2009). The two transesterification steps of pre-mRNA splicing are evolutionarily conserved and mimic the splicing of group II introns that are progenitors of the spliceosomal introns present in lower organisms (Gaur et al., 1997). Spliceosomal assembly begins with the recruitment of U1 snRNP at the 5′ss of an exon. The RNA component (U1 snRNA) of U1 snRNP is 165 nucleotides long and folds into four stem-loop structures. The protein components of U1 snRNP include a 7-membered ring of Sm proteins (Smith antigens), U1-A, U1-70K and U1-C proteins. Both, U1 snRNA and U1-C protein contribute towards the recognition of the 5′ss of an exon (Du and Rosbash, 2002). A driving force behind the recruitment of U1 snRNP at the 5′ss of an exon is the base pairing between the 5′ tail of U1 snRNA and the first six positions (GURAGU) of the intron.

The abundance of U1 snRNP happens to be significantly higher than other snRNPs in human cells (Baserga and Steitz, 1993). Consequently, U1 snRNP is recruited at more sites than

are used as the authentic 5′ss during pre-mRNA splicing. Recruitment of U1 snRNP at multiple sites on pre-mRNA offers several benefits including but not limited to suppression of cryptic exons, maintenance of mRNA length and providing directionality to transcription (Pagani et al., 2002; Almada et al., 2013; Kaida et al., 2010; Berg et al., 2012). An increasing number of reports support coupling of splicing with transcription and chromatin structure (Shukla and Oberdoerffer, 2012; Naftelberg, 2015; Yu and Reed, 2015; Saldi et al., 2016). However, the mechanism of coupling is not yet fully understood.

Humans possess two almost identical copies of *survival motor neuron* gene, *SMN1* and *SMN2* (Lefebvre, 1995). While *SMN1* predominantly produces full-length SMN protein (referred to afterwards as "SMN"), *SMN2* mostly produces SMNΔ7, a truncated protein, due to overwhelming skipping of exon 7 (Lorson et al., 1999; Vitte et al., 2007). In the testis however, both *SMN* genes predominantly produce SMN, possibly due to high demand for SMN during male reproductive organ development (Ottesen et al., 2016). Compared to SMN, SMNΔ7 is less stable due to gaining of a degron signal at the C-terminus (Cho and Dreyfuss, 2010). Consequently, low levels of SMN due to deletions of and/or mutation in *SMN1* lead to spinal muscular atrophy (SMA), one of the leading genetic causes of infant mortality (Monani et al., 2014; Ahmad et al., 2016). The spectrum of SMA is broad and ranges from *in utero* death (type 0), infants who are symptomatic before the age of six months and die within two years of age (type I), individuals who are symptomatic before eighteen months and cannot walk independently but can potentially survive into adulthood (type II), individuals who can walk independently and have a near average to average lifespan (type III), and individuals with adult-onset progressive muscle weakness (type IV) (Nurputra et al., 2013; Howell et al., 2014). High

SMN2 copies that can produce cumulative high levels of SMN and SMNΔ7 ameliorate the symptoms of SMA (Parsons et al., 1998; Wirth et al., 2006; Tiziano et al., 2007). Severity of SMA is modified by several factors including but not limited to NAIP, GTF2H2, H4F5, p44 and Plastin (Tran et al., 2008; Oprea et al., 2008; Amara et al., 2012; He et al., 2013). A recent genome-wide analysis of 12 male SMA patients showed association of epigenetic changes in regulators of Rab and Rho GTPases with SMA severity (Zheleznyakova et al., 2013).

Skipping of *SMN2* exon 7 has been attributed to a critical C-to-T mutation at the 6th position of exon 7 (Lorson et al., 1999). Various cis-elements and transacting factors that regulate *SMN2* exon 7 splicing have been reported (Singh and Singh, 2011; Singh et al., 2015). In vivo selection of the entire exon 7 revealed the weak 5′ss as one of the limiting factors for its inclusion (Singh et al., 2004b, 2004c). Consistent with the results of in vivo selection, abrogation of the terminal stem-loop 2 (TSL2) structure that sequesters the 5′ss was found to promote *SMN2* exon 7 inclusion (Singh et al., 2007). Further supporting the results of *in vivo* selection, an engineered U1 snRNP (eU1 snRNP) with extended the base pairing between the 5′ss of exon 7 and the 5′ tail of U1 snRNA promoted *SMN2* exon 7 inclusion (Singh et al., 2007). These results also provided the first direct evidence that the efficient recruitment of U1 snRNP at the 5′ss of *SMN2* exon 7 is necessary and sufficient to restore *SMN2* exon 7 inclusion. One of the major impediments to an efficient recruitment of U1 snRNP at the 5′ss of *SMN2* exon 7 appears to be the presence of the inhibitory element ISS-N1 that spans from the 9th to 24th positions of intron 7 (Singh et al., 2006). Consistently, deletion or an antisense oligonucleotide (ASO)-mediated sequestration of ISS-N1 fully restores *SMN2* exon 7 inclusion (Singh et al., 2006). Independent studies employing ASO-based strategies targeting ISS-N1 have shown unparalleled efficacies *in*

vivo (Sivanesan et al., 2013). Sequences downstream of ISS-N1 constitute the binding site for TIA1, a splicing factor that promotes recruitment of U1 snRNP at the 5′ss of an exon (Singh et al., 2011). Interestingly, recruitment of a eU1 snRNP in the vicinity of ISS-N1 and/or TIA1 binding site has been found to restore *SMN2* exon 7 inclusion (Dal Mas et al., 2015; Rogalska et al., 2016). In addition to ASOs and eU1 snRNP particles, several small compounds that correct *SMN2* exon 7 splicing have potential for SMA therapy (Seo et al., 2013; Howell et al., 2014).

Research in the last decade has made tremendous progress in determining *SMN* exon 7 splicing regulation and novel approaches to correct *SMN2* exon 7 splicing. However, SMA still has no cure and strategies to treat severe cases remain very daunting. Feasibility of a therapeutic approach becomes further limited if the patient carries a single *SMN2* allele in combination with the deletion and/or lethal mutations of *SMN1* alleles. One such classic example of a severe SMA patient that survived for only four months has recently been reported (Ronchi et al., 2015). This patient carried a single *SMN2* allele along with a single *SMN1* allele harboring a lethal G-to-C mutation at the first position (G1C) of intron 7 (Ronchi et al., 2015). For the ease of understanding we designate $SMN1^{G1C}$ as the *SMN1* mutant carrying the G1C mutation. Here we examine the regulation of $SMN1^{G1C}$ exon 7 splicing employing a translation competent minigene system. Recapitulating the severe impact on SMA patient carrying the $SMN1^{G1C}$ gene, the $SMN1^{G1C}$ minigene showed complete skipping of exon 7. As a therapeutic strategy, we screened a library of eU1 snRNPs to identify a lead eU1 snRNP capable of preventing $SMN1^{G1C}$ exon 7 skipping as well as activating a potential 5′ss downstream of the abrogated 5′ss in $SMN1^{G1C}$. Our approach uncovered two 5′ss, one at the 23rd position and the other at the 51st position within intron 7. Partially overlapping with ISS-N1, the 23rd position emerged as the most favorable 5′ss

among all eU1 snRNPs employed within the first 60 nucleotides of intron 7. Activation of the 5′ss at the 23rd position of intron 7 extended the length of exon 7 by 23 nucleotides. However, since translation stop codon of SMN is located within exon 7, activation of the 5′ss at the 23rd position of intron 7 had no consequence to the SMN coding sequence. We demonstrate that the eU1 snRNP-based approach to activate the 5′ss at the 23rd position of intron 7 has therapeutic implications for a wide range of pathogenic mutations in SMA.

RESULTS

Transcripts derived from *SMN1*^{G1C} minigene undergo complete exon 7 skipping

To recapitulate the splicing of *SMN1* carrying pathogenic G1C mutation (Ronchi et al., 2015), we generated a *SMN1*^{G1C} minigene in which a single G-to-C mutation was introduced at the first position of intron 7 (Figure 4A). As a template, we used *SMN1* minigene as previously described (Singh et al., 2004a). We determined the splicing pattern of exon 7 of *SMN1*^{G1C} minigene by transfecting *SMN1*^{G1C} minigene (plasmid) into HeLa cells similarly as recently described (Singh et al., 2013). In parallel, we also performed control experiments with *SMN1* and *SMN2* minigenes. Transcripts were isolated ~24 h post transfection followed by DNase treatment and RT-PCR. Samples were analyzed on a 6% native polyacrylamide gel (OmniPur, 1690-OP). As expected, transcripts generated from *SMN1* minigene showed predominant inclusion of exon 7 (Figure 4B, lane 4). Also as expected, transcripts generated from *SMN2* minigene showed both inclusion and skipping of exon 7 (Figure 4B, lane 3). Recapitulating the splicing of exon 7 from endogenous *SMN1*^{G1C} gene of an SMA patient (Ronchi et al., 2015), *SMN1*^{G1C} minigene showed complete skipping of exon 7 (Figure 4B, lane 5). In the cases of the *SMN2* and *SMN1*^{G1C} minigenes, we also detected a faint band corresponding to the unspliced precursor RNA (Figure 4B, 1063 bp band).

Effect of engineered U1 snRNA on splicing of exon 7 of *SMN1*^{G1C} minigene

The presence of a G residue at the first intronic position is critical for both the base pairing with the U1 snRNA and the catalytic process during splicing (Lund & Kjems, 2002). In the case of *SMN1* and *SMN2*, the wild-type U1 snRNA (wU1) base pairs with six continuous intronic residues (GUAAGU) at the 5′ss of exon 7. The G1C mutation reduces the base pairing between

wU1 snRNA and the 5′ss of exon 7 to only five continuous intronic residues (Figure 5A). To assess the effect of a strengthened base-pairing between U1 snRNA at the 5′ss of exon 7 on splicing *SMN1*^{G1C} exon 7, we generated an eU1 snRNA (eU1^{11}) carrying five substitutions that increased the U1 snRNA base pairing to eleven continuous residues (three exonic and eight intronic residues). We co-transfected HeLa cells with *SMN1*^{G1C} minigene and eU1^{11} and examined the splicing pattern of exon 7 of *SMN1*^{G1C} transcripts. Unlike wU1, eU1^{11} effectively prevented *SMN1*^{G1C} exon 7 skipping and promoted intron 7 retention at all concentrations examined (Figure 5B, lanes 6-8). These results suggested that a strong RNA:RNA duplex formed by eU1^{11} at the 5′ss of exon 7 favors intron 6 removal but is not conducive for the removal of intron 7. Since the translation stop codon of *SMN* is located within exon 7, retention of intron 7 will have no consequence to protein sequence. However, intron 7 retention will increase the size of the 3′ untranslated region (3′UTR) by 444 nucleotides. The presence of intron 7 and an increase in the size of mRNA may adversely affect the nuclear export and/or translation of the transcript.

Effect of the 5′ss:U1 snRNA duplex size on *SMN1*^{G1C} exon 7 splicing

Inspired by the results of eU1^{11} that prevented *SMN1*^{G1C} exon 7 skipping, we next inquired if the reduction in the size of the 5′ss:eU1 duplex will reduce intron 7 retention and favor exon 7 inclusion from *SMN1*^{G1C}. We generated eU1^{10A}, eU1^{10B}, eU1^{1+8C} and eU1^{2+6} that reduced the 5′ss:eU1 duplex size to 10, 9, 8 and 6 continuous Watson-Crick base pairs, respectively. While eU1^{1+8C} and eU1^{2+6} formed their duplexes by solely annealing to the intronic sequence, eU1^{10A} and eU1^{10B} had extended base pairing at 2 positions within the exonic sequence (Figure 6A). eU1^{10A}, eU1^{10B}, eU1^{1+8C} and eU1^{2+6} retained the C:G base pairing at the first position of

SMN1^{G1C} intron 7. We observed that eU1^{10A} was as effective as eU1^{11} in preventing exon 7 skipping and retaining intron 7 of *SMN1*^{G1C} (Figure 6B). However, with further decrease in 5′ss:eU1 duplex size, we observed more skipping of exon 7 and less retention of *SMN1*^{G1C} intron 7 (Figure 6B). There was no effect of eU1^{1+8C} and eU1^{2+6} on splicing of *SMN1*^{G1C} exon 7 (Figure 6B). We made additional eU1 snRNAs by shifting base pairing positions to see if other variants of short (seven or eight base pairs) 5′ss:eU1 duplexes could stimulate inclusion of *SMN1*^{G1C} exon 7. eU1^{1+8A}, which formed an 8 bp duplex involving two last residues of exon 7, prevented skipping of exon 7 and induced retention of *SMN1*^{G1C} intron 7 (Figure 6C). Our combined results also indicate that the positioning of the 5′ss:eU1 duplex is critical for preventing exon 7 skipping (Figures 6B and 6C). None of the eU1 snRNAs promoted intron 7 removal from *SMN1*^{G1C}. Additionally, some eU1s activated a cryptic splice site (Figures 6B, 6C, 6D, 425 bp product). We also tested higher concentrations of eU1^{10A} to see if removal of *SMN1*^{G1C} intron 7 could be induced. While higher concentrations prevented *SMN1*^{G1C} exon 7 skipping, they failed to induce removal of intron 7 (Figure 6D). These results suggested that *SMN1*^{G1C} is inherently incompetent to initiate the first step of the transesterification process at the 5′ss of exon 7.

Activation of an alternative 5′ss downstream of *SMN1*^{G1C} exon 7

Another likely mechanism that could induce inclusion of *SMN1*^{G1C} exon 7 is the activation of an alternative 5′ss downstream of exon 7. Such activation would increase the size of exon 7 without affecting the protein sequence. As potential 5′ splice sites, there are five GU dinucleotides within the first 60 residues of *SMN1*^{G1C} intron 7. The 1st, 2nd, 3rd and 4th GU dinucleotides occupy the 5th/6th (V1 site), 24th/25th (V2 site), 39th/40th (V3 site), 52nd/53rd (V4 site) intronic positions, respectively (Figure 7A). We examined the effect of V-series eU1 snRNAs that annealed to V1,

V2, V3 and V4 sites within intron 7. In particular, we wanted to determine whether the V-series eU1 snRNAs could promote the inclusion of elongated forms of exon 7 in *SMN1*^{G1C} more effectively than eU1^{1+8A}. For the sake of clarity we term eU1^{V1}, eU1^{V2}, eU1^{V3} and eU1^{V4} to eU1 snRNAs that target V1, V2, V3 and V4 sites, respectively. All V-series eU1 snRNAs formed an 11 bp long 5′ss:eU1 duplex (Figure 7A). A clue that alternative 5′ss downstream of exon 7 could be activated came from eU1^{V4} snRNA that produced two bands from *SMN1*^{G1C} that ran slower than the band from *SMN1* that corresponded to transcripts that include exon 7 (Figure 7B, lanes 1 and 9). Cloning and sequencing of the bands that eU1^{V4} produced from *SMN1*^{G1C} revealed the inclusion of extended forms of exon 7 due to activation of V2 and/or V4 sites. Activation of V2 and V4 sites added 23 and 51 residues at the 3′ end of exon 7, respectively. We call these versions of exon 7 with 23 and 51 nucleotide extensions exon 7^{L1} and exon 7^{L2}, respectively. Interestingly, both eU1^{V2} and eU1^{V3} produced exon 7^{L1}-included variant as the major spliced product. Apart from eU1^{V2}, the other V-series eU1 snRNAs and eU1^{1+8A} produced both exon 7^{L1} and exon 7^{L2}-included transcripts. In all cases, the proportion of exon 7^{L1} -included transcripts was substantially higher than exon 7^{L2}-included transcripts (Figure 7B; lanes 5-9).

Effect of eU1 snRNAs on splicing of exon 7 from endogenous *SMN1* and *SMN2*

While eU1 snRNAs altered exon 7 splicing in context of *SMN* minigenes, we wanted to know if eU1 snRNAs also affected endogenous *SMN1* and *SMN2*. We transfected HeLa cells with wU1, mutant U1, eU1^{V2} or eU1^{V3} expression vectors (Figure 8A). Compared to wU1, mutant U1 has extended base pairing with the wild-type 5′ ss of *SMN1* and *SMN2* intron 7 (Singh et al., 2007; Figure 8A). As compared to cells transfected with wU1, mutant U1, eU1^{V2} and eU1^{V3} promoted exon 7 inclusion from *SMN2* (Figure 8B, lanes 2-4). Taken together, these results demonstrate

that an extended base pairing at either the wild-type splice donor site or at the downstream V2 or V3 splice donor sites can alter endogenous *SMN2* splicing.

The processed *SMN1*^{G1C} transcripts encompassing exon 7^{L1} generate full-length SMN

After determining that the novel exon 7^{L1} is included in transcripts generated from *SMN1*^{G1C} minigene when co-transfected with various eU1 snRNA constructs (Figure 7B), we next examined whether transcripts containing exon 7^{L1} could generate a stable protein. We transfected HeLa cells with the translation competent minigenes *SMN1* and *SMN1*^{G1C} (Figure 9A). As a control we used cDNA clones *cSMN* and *cSMNΔE7*. All minigenes or cDNA clones carried FLAG tag at the N-terminus. As expected, *SMN1* minigene generated SMN that migrated similar to SMN generated from cDNA clone of *SMN1* (Figure 9A, lanes 1 and 2). The *SMN1*^{G1C} minigene generated multiple bands, two of which corresponded to SMN and SMNΔ7 proteins (Figure 9A, lane 3). We also performed a titration experiment to determine how much *FLAG-SMN1* would be required to observe SMN expression (Figure 9B). We observed that transfecting 0.6 µg of *FLAG-SMN1* was sufficient to obtain notable SMN expression (Figure 9B) and we used this concentration for subsequent experiments.

Since the *FLAG-SMN1*^{G1C} minigene generated protein, we next examined whether altering splicing with the eU1^{V3} would also alter the translated protein. For comparison on western blot, we generated the *FLAG-SMNI6Δ7* minigene to simulate intron 6 retention. The *FLAG-SMNI6Δ7* minigene was predicted to generate protein and contains the first 62 bp of intron 6, which includes a natural stop codon, but lacks exon 7, intron 7, and the 3′ss of exon 8 (Figure 9C). HeLa cells were transfected with various FLAG constructs and allowed to express

protein. The *cSMN*- and *SMN1*-transfected cells generated a full-length SMN protein (Figure 10C, lanes 1 and 4, respectively). The *SMNI6Δ7*-transfected cells generated a SMN protein that had a similar molecular weight as full-length SMN protein (Figure 10C, lane 3). *FLAG-SMNΔ7*-transfected cells generated SMNΔ7 protein (Figure 9C, lane 2). Similarly, *FLAG-SMN1*^{G1C}-transfected cells also generated SMNΔ7 protein. However, when cells were co-transfected with *FLAG-SMN1*^{G1C} and eU1^{V3}, the cells produced full-length SMN (Figure 9C, lanes 6-8). Thus, activation of the V2 site results into generation of full-length SMN. These results indicate that V2 splice site could serve as a therapeutic target to correct defective splicing in SMA patients carrying the pathogenic G1C mutation in *SMN1*.

DISCUSSION

SMA is one of the frequent hereditary causes of infant mortality. A vast majority of SMA patients carry *SMN2* that produces low levels of SMN. Except for gene therapy, most therapeutic approaches under development rely on employment of the existing *SMN2* to enhance levels of SMN via one of the three potential mechanisms: (i) increased *SMN2* transcription, (ii) correction of *SMN2* exon 7 splicing, and (iii) stabilization of the SMN protein. The expected beneficial effects of the *SMN2*-dependent approaches have a ceiling threshold that cannot be easily crossed by low copy numbers of *SMN2*. In this study, we have addressed this issue for a subset of SMA patients with one *SMN2* allele and a defective *SMN1* carrying a lethal G1C mutation at the splice donor site. The approach is inspired by our previous study in which we showed that an extended 5′ss:eU1 RNA duplex is favorable for *SMN2* exon 7 inclusion (Singh et al., 2007). We reasoned that this approach could also be applicable for the prevention of exon 7 skipping in $SMN1^{G1C}$. Indeed, eU1 snRNAs with extended base pairing at the 5′ss carrying the G1C mutation fully prevented exon 7 skipping. However, the major splice product retained intron 7, suggesting that G1C is incompetent to carryout splicing reaction. Such an outcome is not totally surprising, since a G residue at the first position of an intron is involved in catalysis (Lund & Kjems, 2002).

Intron 7 retention has no consequence to protein sequence in the case of *SMN* genes. Therefore, an ASO-based approach to promote retention of *SMN2* intron 7 through blocking of the 3′ss of exon 8 has been considered as one of the possible therapeutic avenues (Lim and Hertel, 2001). However, concerns remain that a transcript retaining an intron may not be efficiently exported out of the nucleus and efficiently translated. To overcome these issues in the case of $SMN1^{G1C}$, we explored activating a potential 5′ss immediately downstream of the

mutated 5′ss in *SMN1*^{G1C} minigene. Among four potential candidates examined, the 5′ss (V2 site) partially overlapping the ISS-N1 emerged as the optimal candidate. The activation of V2 was robust in the case of eU1 snRNAs that annealed to V2 or V4 sites. In addition to eU1 snRNA targeting V2 site, other eU1 snRNAs also activated V4 site. However, the extent of V4 activation did not reach to the level of V2 activation by any of the eU1 snRNAs examined.

Previous studies have shown that eU1 snRNAs targeting ISS-N1 site strengthen the wild-type 5′ss of exon 7 and promote *SMN2* exon 7 inclusion (Dal Mas, Fortugno et al., 2015). Consistent with these findings, eU1 snRNAs used in our study also promoted inclusion of exon 7 from endogenous *SMN2*. However, the rules of splicing appear to change in the absence of an active 5′ss of *SMN* exon 7. In the scenario of the loss of the wild-type 5′ss of *SMN* exon 7, V2 site becomes the next most favorable 5′ss. Consistent with a transcriptome-wide study supporting that silencer elements suppress cryptic 5′ss (Wang et al., 2004), we hypothesize that ISS-N1 may serve such a role by sequestering the V2 site. Several factors may contribute to why V2 site was the most activated 5′ss by a eU1 snRNA in the case of *SMN1* carrying the G1C mutation. The sequence composition of V2 is very close to the consensus GTRAGT motif of the 5′ss. The binding sites of TIA1 and/or TIAR are located just downstream of V2. TIA1 and TIAR are known to promote *SMN2* exon 7 inclusion possibly through enhanced recruitment of U1 snRNP downstream of the ISS-N1 site (Singh et al., 2011). We have recently reported the secondary structure of *SMN* intron 7. Interestingly, V2 site is located in the most accessible region in the structure of intron 7. While all of the above-mentioned factors may have contributed to the selection of V2 site, presence of eU1 snRNA with extensive base pairing at V2 site appears to be

the driving force behind V2 activation. Consistently, we could not detect appreciable level of V2 activation in *SMN1*^{G1C} minigene in the absence of a eU1 snRNA.

Inclusion of exon 7^{L1} has no consequence to SMN protein. Consistently, a eU1 snRNA that targeted V2 site produced SMN from a translation competent *SMN1*^{G1C} minigene. These results confirm that an exon 7^{L1}-containing transcript is nuclear export- and translation-competent. Our findings open up new therapeutic possibilities for a group of SMA patients who would not benefit from most therapeutic approaches currently under development. Since an U1 snRNA targeting V2 site also promotes exon 7 inclusion from wild-type *SMN* genes, the approach is relevant to a substantial majority of SMA patients. While we have discovered a novel therapeutic molecule for treating SMA, determining a method to deliver this molecule *in vivo* is still a challenge. One of the most plausible ways of delivering eU1 snRNAs is through most frequently used AAV9-derived viral vectors used in gene therapy. Recent years have witnessed tremendous improvements towards gene therapy of SMA (Benkhelifa-Ziyyat et al., 2013; Meyer et al., 2015). Compared to gene therapy that generates SMN from an artificial gene, a eU1 snRNA generates SMN using an endogenous gene. Hence, a eU1 snRNA-based approach described here provides a better alternative, since it restores the ability of the defective endogenous gene to fine-tune the tissue-specific expression of SMN.

MATERIALS AND METHODS

Construction of Minigene Containing Pathogenic Mutation

The *SMN1*^{G1C} minigene was constructed by inducing a G-to-C substitution at the first position of intron 7 in pSMN1ΔI6 (Singh et al., 2007). First, two fragments were amplified using primer pair 5SMN1Ex6 and 3SMN1E7I7 and primer pair 5SMN1E7I7 and 3SMN1Ex8 using pSMN1ΔI6 as a template. The two fragments were run on 1% agarose gel, isolated using QIAquick Gel Extraction Kit (Qiagen, 28704 and 28706), and then PCR-ligated using primers 5SMN1Ex6 and 3SMN1Ex8. After gel purification, the PCR product was digested with XhoI and NotI for 1 hour, run on 1% agarose gel, and isolated using QIAquick Gel Extraction Kit. The purified PCR product was cloned into pCI vector that had been digested with XhoI and NotI for 1 hour and gel purified. The identity of the *SMN1*^{G1C} minigene was verified using Sanger DNA sequencing (DNA Facility of the Iowa State University Office of Biotechnology, IA).

The *FLAG-SMN1* and *FLAG-SMN1*^{G1C} minigenes were constructed by adding three copies of FLAG tag (3XFLAG) sequence (5′-ATGGACTACAAAGACCATGACGGTGATTAT AAAGATCATGACATCGACTACAAAGACGACGATGACAAGACGCGTTCTAGA-3′) and *SMN1* exons 1 through 5 to the 5′ ends of pSMN1ΔI6 and *SMN1*^{G1C} respectively. First, NdeI cut fragments of 3XFLAG-SMN (a pCI-neo-based expression vector that has 3XFLAG sequence followed by *SMN1* exons 1 through 7) containing exons 2a through 6 were PCR ligated with pSMN1ΔI6 and *SMN1*^{G1C} using primer pair 5Ex4 and 3SMN1Ex8. The PCR ligation products were then PCR ligated with ApoI cut fragments of 3XFLAG-SMN containing 3XFLAG sequence and exons 1 through 5. The PCR products were digested with XhoI and NotI for 1 hour, run on 1% agarose gel, and isolated using QIAquick Gel Extraction Kit. The purified PCR product was cloned into pCI vector that had been digested with XhoI and NotI overnight and gel-

purified. The identities of the *FLAG-SMN1* and *FLAG-SMN1*^{G1C} minigenes were confirmed using Sanger DNA sequencing.

The *FLAG-SMNI6Δ7* minigene was constructed by deleting the last 155 bp of intron 6, exon 7, intron 7, and the first 24 bp of exon 8 in *FLAG-SMN1*. Two fragments were amplified using primer pair 5Ex8a and PCI-DN and primer pair PCI-UP and 3In7a using *FLAG-SMN1* as a template. The two fragments were PCR-ligated using PCI-UP and PCI-DN. The PCR product was cleaved with XhoI and NotI for 1 hour, run on 1% agarose gel, and isolated using QIAquick Gel Extraction Kit. The purified PCR product was cloned into pCI vector that had been digested with XhoI and NotI overnight and gel purified. The identity of the *FLAG-SMNI6Δ7* minigene was confirmed using Sanger DNA sequencing. Gels were made using Omnipur agarose (OmniPur, 2120-OP).

Construction of U1 snRNA Expression Vectors

The mutated U1 snRNA expression vectors were constructed using PCR to introduce mutations at the 5′ end of U1 RNA. The site-specific mutations were generated using primer 3′pUCBU1 with primer 5′pUCBU-mutG1C for eU1^{11}, 5′pUCBU-mut4 for eU1^{10A}, 5′pUCBU-mut3 for eU1^{10B}, M3V2 for eU1^{1+8A}, M3V3 for eU1^{1+8B}, M3V4 for eU1^{2+7}, 5′pUCBU-mut2 for eU1^{1+8C}, M2V2 for eU1^{1+8D}, 5′pUCBU-mut1 for eU1^{2+6}, MI7V1 for eU1^{V1}, MI7V2 for eU1^{V2}, MI7V3 for eU1^{V3}, MI7V4 for eU1^{V4}, and plasmid pUCBU1 (Singh et al., 2007) as a template. The PCR products were digested with Bg1II and XhoI for 1hour, run on 1% agarose gel, and isolated using QIAquick Gel Extraction Kit. The purified PCR products were cloned into pUCBΔU1 vector (Singh et al., 2007) that had been digested with Bg1II and XhoI for 1 hour and gel purified. The

identities of the eU1 snRNAs were verified using Sanger DNA sequencing. Integrated DNA Technologies (Coralville, IA) and the DNA Facility of the Iowa State University Office of Biotechnology (Ames, IA) supplied the oligonucleotides.

Cell Culture, Transfection, and *in vivo* splicing

HeLa cells were obtained from the American Type Culture Collection and were maintained in Dulbecco's Modified Eagle's Medium (Thermo Fisher Scientific,11965) containing 10% fetal bovine serum (Thermo Fisher Scientific, 26140079). Transfections were conducted using X-tremeGENE (Roche Applied Science, 04476093001 and 04476115001) following the manufacturer's instructions. HeLa cells were plated in either a 24-well or 6-well plate at a density of $0.9x10^5$ cells per well or $0.4x10^6$ cells per well, respectively, a day before transfection. The ratio between nucleic acids and X-tremeGENE HP DNA Transfection Reagent was 1:2. Unless stated otherwise, in co-transfection experiments using either 24-well or 6-well plates, cells were transfected with 0.05 µg of a minigene and 0.05 µg of a U1 snRNA expression vector or 0.1 µg of minigene and 0.2 µg of U1 snRNA expression vector respectively. Whenever necessary, GFP expression vector was added in required amounts to maintain the total amount of DNA (0.5 µg in 24-well plates and 2 µg in 6-well plates).

Total RNA was isolated 20-24 hours after transfection using Trizol reagent (Invitrogen, 15596018 and 15596026) followed by treatment with RQ1 DNase (Promega, M6101) for 30 minutes at 37 ˚C. 2 µl of RQ1 DNase Stop was then added to each reaction and incubation continued for 10 minutes at 65 ˚C. cDNA was generated using SuperscriptIII reaction kit (Invitrogen, 18080-044) and oligo $(dT)_{12-18}$ primer (Invitrogen, 18418012). Unless stated

otherwise, 2 µl of DNase reaction were used per 10 µl reverse transcriptase (RT) reaction. After mixing 2 µl of DNase eaction with 0.5 µl of oligo $(dT)_{12-18}$, 0.5 µl 10 mM dNTP, and 4 µl of water, RT reaction solutions were heated to 65 ˚C for 5 minutes and incubated on ice for 5 minutes to denature the RNA. After adding 2 µl of 5X First Strand Buffer, 0.5 µl of 0.1M DTT, and 0.5 µl of SuperScript RT, RT reactions were incubated at 50 ˚C for 60 minutes and then heated at 70 ˚C for 15 minutes to inhibit the reverse transcriptase.

Minigene-specific spliced products were amplified using Taq polymerase (New England Biolabs, M0273S; Invitrogen, 10342-053) and primer pair Ex8P2-2 and PCI-UP or 5CTRL. PCR amplification was performed for 30 cycles and used an annealing temperature of 55 ˚C. Endogenous *SMN1* and *SMN2* spliced products were amplified using primers P25 and P31 for 28 cycles and used an annealing temperature of 54 ˚C. In order to distinguish between amplified spliced products from endogenous *SMN1* and *SMN2*, 20 µl of PCR product was used per 100 µl DdeI restriction digest reaction that was allowed to incubate overnight (Singh et al., 2011). The Qiaquick PCR purification kit (Qiagen, 28704 and 28706) was used to purify DdeI-digested PCR products. The PCR products were eluted into 30 µl of EB buffer (10 mM Tris-Cl, pH 8.5). PCR products were run at 200 volts on 6% acrylamide gels, which were prepared using 40% acrylamide/bisacrylamide stock solution (Omnipur, 1690-OP). PCR products were then stained with ethidium bromide (Thermo Fisher Scientific, 15585-011). Spliced products were revealed using UVP Biospectrum AC Imaging System (UVP, 81-0346-01). When minigene-specific spliced products needed to be quantified, PCR amplification was performed in the presence of [α-^{32}P] dATP (Perkin-Elmer Life Sciences, BLU503H250UC) using primers P1 and P2 for 19 cycles and used an annealing temperature of 60 ˚C. PCR products were run on 6% acrylamide

gels at 220 volts. Radioactive acrylamide gels were dried using a Model 583 Gel Dryer (Bio-Rad, 1651797) and the spliced products were visualized using a Fujifilm FLA-5100 imaging system (FUJIFILM Life Science). The PCR products were quantified using ImageGauge software (Fuji Photo Film Inc.). The identities of spliced products were verified using Sanger DNA Sequencing.

Western blot analysis

Protein was extracted from HeLa cells 24 hours after transfection. Growth media was removed from each well of transfected HeLa cells and the cells were washed once with 2 ml of cold DPBS (Thermo Fisher Scientific, 14190250). The DPBS was removed and 1 ml of cold DPBS was added to each well of transfected HeLa cells. The HeLa cells were scraped and collected into pre-chilled 1.5 ml tubes. The HeLa cell samples were spun at 3500g for 1 minute at 4 ˚C and the supernatant of each sample was removed. The cell pellet of each sample was re-suspended in 70 µl of radioimmunoprecipitation assay (RIPA) buffer (Boston BioProducts, BP-115), containing Halt™ Protease and Phosphotase Inhibitor Cocktail (Thermo Fisher Scientific, 78440), and was lysed on ice for 30 minutes. The samples were spun at 12000g for 10 minutes at 4 ˚C and the supernatant of each sample was collected in pre-chilled 1.5 ml tubes. Protein concentrations were calculated using Bradford Protein Assay Kit II (Bio-Rad, 500-0002). 2X Laemmli buffer (Bio-Rad, 161-0737) containing 5% beta-mercaptoethanol (Calbiochem, 444203) was added to protein samples at a 1:1 ratio, and boiled for 5 minutes.

Unless stated otherwise, 10 µg of total protein per sample was resolved on an 11% SDS-polyacrylamide gel and transferred onto a polyvinylidene diflouride (PVDF) membrane (Bio-

Rad, 1704156) using the Trans-Blot Turbo Transfer System (Bio-Rad, 1704155). PVDF membranes were blocked in 5% non-fat milk dissolved in 1X TBST buffer for 30 minutes at 37 ˚C. 1X TBST was composed of 0.05M Tris (pH 7.5), 0.45M NaCl, and 0.05% Tween 20. FLAG-tagged proteins were probed by incubating blocked PVDF membrane with mouse monoclonal anti-FLAG M2 antibody (Sigma-Aldrich, F3165) diluted 1:4000 in 5% non-fat milk dissolved in 1X TBST for 30 minutes at 37 ˚C. PVDF membranes were then washed three times in 1X TBST for 10 minutes each. Proteins were visualized using Bio-Rad Clarity western ECL substrate (Bio-Rad, 1705060 and 1705061) and PVDF membranes were scanned using a UVP Biospectrum AC Imaging System (UVP, 81-0207-01). PVDF membrane images were developed using exposure times between 11 to 13 seconds.

REFERENCES

Ahmad S, Bhatia K, Kannan A, Ganwani L (2016). Molecular mechanisms of neurodegeneration in Spinal Muscular Atrophy. J Exp Neurosci *10*, 39-49. doi:10.4137/JEN.S33122.

Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499(7458)*, 360-363. doi: 10.1038/nature12349.

Amara A, Adala L, Charfeddine IB, Mamaï O, Mili A, Lazreg TB, H'mida D, Amri F, Salem N, Boughammura L, Saad A (2012). Correlation of SMN2, NAIP, p44, H4F5 and Occludin genes copy number with spinal muscular atrophy phenotype in Tunisian patients. European Journal of Paediatric Neurology *16(2)*, 167-74. doi:10.1016/j.ejpn.2011.07.007.

Baserga SJ, Steitz JA (1993). 14 The Diverse World of Small Ribonucleoproteins. Cold Spring Harbor Monograph Archive *24*, 359-381. doi: 10.1101/087969380.24.359.

Benkhelifa-Ziyyat S, Besse A, Roda M, Duque S, Astord S, Carcenac R, Marais T, Barkats M (2013). Intramuscular scAAV9-SMN injection mediates widespread gene delivery to the spinal cord and decreases disease severity in SMA mice. Mol Ther. *21(2)*, 282-90. doi: 10.1038/mt.2012.261.

Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L, Dreyfuss G (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell. *150(1)*, 53-64. doi: 10.1016/j.cell.2012.05.029.

Buratti E, Baralle FE (2004). Influence of RNA secondary structure on the pre-mRNA splicing process. Mol. Cell. Biol. *24*, 10505-10514. doi: 10.1128/MCB.24.24.10505-10514.2004.

Burge CB, Tuschl T, Sharp PA (1999). Splicing of precursors to mRNAs by the spliceosomes. The RNA World, R.F. Gesteland and J.F. Atkins, eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), pp. 525–560. doi: 10.1101/087969589.37.525.

Cho S, Dreyfuss G (2010). A degron created by SMN2 exon 7 skipping is a principal contributor to spinal muscular atrophy severity. Genes Dev. *24(5)*, 438-442. doi: 10.1101/gad.1884910.

Dal Mas A, Fortugno P, Donadon I, Levati L, Castiglia D, Pagani F (2015). Exon-Specific U1s Correct SPINK5 Exon 11 Skipping Caused by a Synonymous Substitution that Affects a Bifunctional Splicing Regulatory Element. Human Mutation *36(5)*, 504-12. doi: 10.1002/humu.22762.

Dal Mas A, Rogalska ME, Bussani E, Pagani F (2015). Improvement of SMN2 pre-mRNA processing mediated by exon-specific U1 small nuclear RNA. American Journal of Human Genetics *96(1)*, 93-103. doi: 10.1016/j.ajhg.2014.12.009.

Du H, Rosbash M (2002). The U1 snRNP protein U1C recognizes the 5′ splice site in the absence of base pairing. Nature *419(6902)*, 86-90. doi: 10.1038/nature00947.

Gaur RK, McLaughlin LW, Green MR (1997). Functional group substitutions of the branchpoint adenosine in a nuclear pre-mRNA and a group II intron. RNA *3(8)*, 861-869. PMCID: PMC1369531.

He J, Zhang QJ, Lin QF, Chen YF, Lin XZ, Lin MT, Murong SX, Wang N, Chen WJ (2013). Molecular analysis of SMN1, SMN2, NAIP, GTF2H2, and H4F5 genes in 157 Chinese patients with spinal muscular atrophy. Gene. *518(2)*, 325–329. doi:10.1016/j.gene.2012.12.109.

Howell MD, Singh NN, Singh RN (2014). Advances in therapeutic development for spinal muscular atrophy. Future Med Chem. *6(9)*, 1081-99. doi: 10.4155/fmc.14.63.

Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468(7324)*, 664-668. doi: 10.1038/nature09479.

Lee Y, Rio DC (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem. *84*, 291-323. doi: 10.1146/annurev-biochem-060614-034316.

Lim SR, Hertel KJ (2001). Modulation of survival motor neuron pre-mRNA splicing by inhibition of alternative 3' splice site pairing. The Journal of Biological Chemistry *276(48)*, 45476-45483. doi: 10.1074/jbc.M107632200.

Lorson CL, Hahnen E, Androphy EJ, Wirth B (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad Sci USA. *96(11)*, 6307-11. doi: 10.1073/pnas.96.11.6307.

Lund M, Kjems J (2002). Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end. RNA *8(2)*, 166-79. doi: 10.1017/S1355838202010786.

Matlin AJ, Moore MJ (2007). Spliceosome assembly and composition. Adv Exp Med Biol. *623*, 14-35. PMID: 18380338.

McManus CJ, Graveley BR (2011). RNA structure and the mechanism of alternative splicing. Curr. Opin. Genet. Dev. *21*, 373-379. doi:10.1016/j.gde.2011.04.001.

Meyer K, Ferraiuolo L, Schmelzer L, Braun L, McGovern V, Likhite S, Michels O, Govoni A, Fitzgerald J, Morales P, Foust KD, Mendell JR, Burghes AH, Kaspar BK (2015). Improving single injection CSF delivery of AAV9-mediated gene therapy for SMA: a dose-response study in mice and nonhuman primates. Mol Ther. *23(3)*, 477-487. doi: 10.1038/mt.2014.210.

Monani UR, De Vivo DC (2014). Neurodegeneration in spinal muscular atrophy: from disease phenotype and animal models to therapeutic strategies and beyond. Future Neurol. *9(1)*, 49-65. doi: 10.2217/fnl.13.58.

Naftelberg S, Schor IE, Ast G, Kornblihtt AR (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu Rev Biochem. *84*, 165-98. doi: 10.1146/annurev-biochem-060614-034242.

Nilsen TW, Graveley BR (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463(7280)*, 457-463. doi:10.1038/nature08909.

Nurputra DK, Lai PS, Harahap NI, Morikawa S, Yamamoto T, Nishimura N, et al (2013). Spinal muscular atrophy: from gene discovery to clinical trials. Ann Hum Genet. *77(5)*, 435-63. doi: 10.1111/ahg.12031.

Oprea G, Kröber S, McWhorter M, Rossoll W, Müller S, Krawczak M, Bassell G, Beattie C, Wirth, B. (2008). Plastin 3 is a protective modifier of autosomal recessive Spinal Muscular Atrophy. Science *320*, 524-527. doi: 10.1126/science.1155085.

Ottesen EW, Howell MD, Singh NN, Seo J, Whitley EM, Singh RN (2016). Severe impairment of male reproductive organ development in a low SMN expressing mouse model of spinal muscular atrophy. Sci Rep. *6*, 20193. doi: 10.1038/srep20193.

Pagani F, Buratti E, Stuani C, Bendix R, Dörk T, Baralle FE (2002). A new type of mutation causes a splicing defect in ATM. Nat Genet. *30(4)*, 426-9. doi: 10.1038/ng858.

Parsons DW, McAndrew PE, Iannaccone ST, Mendell JR, Burghes AH, Prior TW (1998). Intragenic telSMN mutations: frequency, distribution, evidence of a founder effect, and modification of the spinal muscular atrophy phenotype by cenSMN copy number. Am. J. Hum. Genet. *63(6)*, 1712–1723 (1998). doi: 10.1086/302160.

Rogalska ME, Tajnik M, Licastro D, Bussani E, Camparini L, Mattioli C, Pagani F (2016). Therapeutic activity of modified U1 core spliceosomal particles. Nature Communications *7*, 11168. doi: 10.1038/ncomms11168.

Ronchi D, Previtali SC, Sora MG, Barera G, Del Menico B, Corti S, Bresolin N, Comi GP (2015). Novel splice-site mutation in SMN1 associated with a very severe SMA-I phenotype. J Mol Neurosci. *56(1)*, 212-215. doi: 10.1007/s12031-014-0483-4.

Saldi T, Cortazar MA, Sheridan RM, Bentley DL (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. J Mol Biol. *428(12)*, 2623-2635. doi: 10.1016/j.jmb.2016.04.017.

Seo J, Howell MD, Singh NN, Singh RN (2013). Spinal muscular atrophy: an update on therapeutic progress. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease *1832(12)*, 2180-2190. doi: 10.1016/j.bbadis.2013.08.005.

Shukla S, Oberdoerffer S (2012). Co-transcriptional regulation of alternative pre-mRNA splicing. Biochim Biophys Acta. *1819(7)*, 673-683. doi: 10.1016/j.bbagrm.2012.01.014.

Singh NN, Androphy EJ and Singh RN (2004a). An extended inhibitory context causes skipping of exon 7 of *SMN2* in spinal muscular atrophy. Biochemical and Biophysical Research Communications *315*, 381-388. doi: 10.1016/j.bbrc.2004.01.067.

Singh NN, Androphy EJ and Singh RN (2004b). In vivo selection reveals features of combinatorial control that defines a critical exon in the spinal muscular atrophy genes. RNA *10*, 1291-1305. doi: 10.1261/rna.7580704.

Singh NN, Androphy EJ and Singh RN (2004c) Regulation and regulatory activities of alternative splicing of the SMN genes. Critical Reviews in Eukaryotic Gene Expression *14*, 271-285. doi: 10.1615/CritRevEukaryotGeneExpr.v14.i4.30.

Singh Nk, Singh NN, Androphy EJ, Singh RN (2006). Splicing of a critical exon of human *Survival Motor Neuron* is regulated by a unique silencer element located in the last intron. Molecular and Cellular Biology *26(4)*, 1333-1346. doi: 10.1128/MCB.26.4.1333-1346.2006.

Singh NN, Singh RN, Androphy EJ (2007). Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. Nucleic Acids Research *35(2)*, 371-89. doi: 10.1093/nar/gkl1050.

Singh NN, Seo J, Ottesen EW, Shishimorova M, Bhattacharya D, Singh RN (2011). TIA1 prevents skipping of a critical exon associated with spinal muscular atrophy. Molecular and Cellular Biology *31(5)*, 935-954. doi: 10.1128/MCB.00945-10.

Singh NN, Singh RN (2011). Alternative splicing in spinal muscular atrophy underscores the role of an intron definition model. RNA Biology *8(4)*, 600-606. doi: 10.4161/rna.8.4.16224.

Singh NN, Lee BM, Singh RN (2015). Splicing regulation in spinal muscular atrophy by an RNA structure formed by long-distance interactions. Ann N Y Acad Sci. *1341*, 176-187. doi: 10.1111/nyas.12727.

Sivanesan S, Howell MD, DiDonato CJ, Singh RN (2013). Antisense oligonucleotide mediated therapy of spinal muscular atrophy. Translational neuroscience *4(1)*, 1-7. doi: 10.2478/s13380-013-0109-2.

Tiziano FD, Bertini E, Messina S, et al (2007). The Hammersmith functional score correlates with the SMN2 copy number: a multicentric study. Neuromuscul. Disord. *17(5)*, 400–403. doi: 10.1016/j.nmd.2007.02.006.

Tran VK, Sasongko TH, Hong DD, Hoan NT, Dung VC, Lee MJ, Gunadi Takeshima Y, Matsuo M, Nishio H (2008). *SMN2* and NAIP gene dosages in Vietnamese patients with spinal muscular atrophy. Ped Int *50*, 346–3451. doi: 10.1111/j.1442-200X.2008.02590.x.

Vitte J, Fassier C, Tiziano FD, Dalard C, Soave S, Roblot N, et al (2007). Refined characterization of the expression and stability of the SMN gene products. Am J Pathol. *171(4)*, 1269-1280. doi: S0002-9440(10)62390.

Wahl MC, Will CL, Lührmann R (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell. *136(4)*, 701-718. doi: 10.1016/j.cell.2009.02.009.

Wang Z, Burge CB (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA *14*, 802-813. doi: .

Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB (2004). Systematic identification and analysis of exonic splicing silencers. Cell. *119(6)*, 831-845. doi: .

Wang J, Zhang J, Li K, Zhao W, Cui Q (2012). SpliceDisease database: linking RNA splicing and disease. Nucleic Acids Research *40(Database issue)*, D1055-D1059. doi: 10.1093/nar/gkr1171.

Wang Y, Xiao X, Zhang J, Choudhury R, Robertson A, Li K, Ma M, Burge CB, Wang Z (2013). A Complex Network of Factors with Overlapping Affinities Repress Splicing through Intronic Elements. Nat Struct Mol Biol. 20(1), 36-45. doi: 10.1038/nsmb.2459.

Warf MB, Berglund JA (2010). Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem. Sci. *35*, 169-178. doi: 10.1016/j.tibs.2009.10.004.

Wirth B, Brichta L, Schrank B, et al (2006). Mildly affected patients with spinal muscular atrophy are partially protected by an increased SMN2 copy number. Hum. Genet. *119(4)*, 422–428. doi: 10.1007/s00439-006-0156-7.

Xing Y, Lee C (2007) Relating alternative splicing to proteome complexity and genome evolution. Adv Exp Med Biol *623*, 36-49. PMID: 18380339.

Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RK, Hua Y, Gueroussov S, Najafabadi HS, Hughes TR, Morris Q, Barash Y, Krainer AR, Jojic N, Scherer SW, Blencowe BJ, Frey BJ (2015). The human splicing code reveals new insights into the genetic determinants of disease. Science. *347(6218)*, 1254806. doi: 10.1126/science.1254806.

Yu Y, Reed R (2015). FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP. Proc Natl Acad Sci U S A. *112(28)*, 8608-8613. doi: 10.1073/pnas.1506282112.

Zheleznyakova GY, Voisin S, Kiselev AV, Almén MS, Xavier MJ, Maretina MA, Tishchenko LI, Fredriksson R, Baranov VS, Schiöth HB (2013). Genome-wide analysis shows association of epigenetic changes in regulators of Rab and Rho GTPases with spinal muscular atrophy severity. European Journal of Human Genetics *21(9)*, 988-93. doi: 10.1038/ejhg.2012.293.

FIGURE LEGENDS

Figure 4. *SMN1*^{G1C} minigene exhibits exon 7 skipping. **(A)** Diagrammatic represenation of the *SMN1*^{G1C} minigene with the intron 7 mutation shown in red. Exon 6, exon 7, and exon 8 are indicated as boxes. Intron 6 and intron 7 are indicated as lines. Sizes of exons and introns are indicated in base pairs (bp). **(B)** Splicing pattern of *SMN2*, *SMN1* and *SMN1*^{G1C} minigenes in HeLa cells. Molecular weights are indicated to the left while band identities are shown to the right. Above the gel, DNase indicates whether the total RNA was treated with DNase prior to reverse transcription and RT indicates whether reverse transcriptase was added to a reaction mixture. HeLa cells were plated in 24-well plates at a density of $0.9x10^5$ cells per well. Cells were transfected with 0.05 µg of mini-gene and 0.45 µg of GFP expression vector. Total RNA was isolated from HeLa cells 24 hours after transfection followed by treating 6 µg of RNA with DNase in 20 µl DNase reaction. cDNA was generated using 4 µl of DNase reaction per 10 µl RT reaction. PCR products were generated using primers PCI-UP and 3Ex8P2-2 and used 1 µl of RT reaction was used per 20 µl Taq PCR reaction. PCR amplification was performed for 30 cycles and the annealing temperature (Tm) was 55 ˚C. The PCR products were run on a 6% acrylamide gel and were stained with ethidium bromide.

Figure 5. eU1 snRNA with extended base pairing at the 5′ splice site (5′ss) of intron 7 in the *SMN1*^{G1C} minigene promotes intron 7 retention. **(A)** The sequence of the eU1 snRNA compared to wild-type (wU1) U1 snRNA and the 5′ss of intron 7 in *SMN1*^{G1C}. Mutated nucleotides are indicated by red letters. Black circles represent both Watson-Crick and wobble base pairing between U1 snRNA and the 5′ss of exon 7. Exon 7 sequence is shown in upper-case letters while intron 7 sequence is shown in lower-case letters. Numbering starts from the first position of

intron 7. **(B)** Splicing pattern of *SMN1*^{G1C} minigene in the presence of wU1 and eU1 snRNA in HeLa cells. Band sizes are indicated to the left, band identifies are indicated to the right, and the contents for each transfection are indicated above the gel. For lanes 3-8, the amount of each U1 snRNA expression vector transfected is indicated. HeLa cells were plated in the same manner described in Figure 4B. Cells were transfected with 0.05 µg of mini-gene and 0.05 µg, 0.2 µg, or 0.5 µg of a given U1 snRNA expression vector. GFP expression vector was added when necessary to maintain a total amount of 0.5 µg of DNA in each transfection. Total RNA was isolated from HeLa cells 22 hours after transfection followed by treating 6 µg of RNA with DNAse in 20 µl DNase reaction. cDNA was generated using 7 µl of DNase reaction per 20 µl RT reaction. PCR products were generated, run on a 6% acrylamide gel, and stained in the same manner described in Figure 4B.

Figure 6. The effect of different eU1 snRNAs with differing degrees of extended base pairing with the 5′ ss of exon 7 on *SMN1*^{G1C} splicing. **(A)** The sequences of the eU1 snRNAs compared to the wU1 snRNA. Names of U1 snRNAs are shown on the left. Red letters indicate mutations compared to wild-type U1 snRNA (wU1). Black circles indicate Watson-Crick and wobble base pairs formed between U1 snRNAs and the 5′ss of exon 7. Exon 7 sequence is shown in upper-case letters while intron 7 sequence is shown in lower-case letters. **(B)** Splicing pattern of minigenes and eU1 snRNA in HeLa cells. Band sizes are indicated to the left and band identities are indicated to the right of the gel. The minigenes and eU1 transfected for each sample are indicated above the gel. HeLa cells were plated in 6-well plates at a density of 4.8×10^5 cells per well. Cells were transfected with 1 µg of mini-gene and 1 µg of either U1 snRNA expression vector or GFP expression vector. Total RNA was isolated from HeLa cells 22 hours after

transfection followed by treating 8 µg of RNA with DNAse in 20 µl DNase reaction. cDNA was generated in the same manner described in Figure 4B. PCR products were generated in the same manner described in Figure 4B, except 1 µl cDNA was used per 25 µl Taq PCR reaction. PCR products were run on a 6% acrylamide gel and stained in the same manner described in Figure 4B. **(C)** Splicing pattern of minigenes with different eU1 in HeLa cells. HeLa cells were plated as described in Figure 6B. Cells were transfected with 0.1 µg of mini-gene and 0.2 µg of a given U1 snRNA expression vector. GFP was added when necessary to maintain a total amount of 2 µg of DNA in each transfection. Total RNA was isolated from HeLa cells 21 hours after transfection followed by treating 1 µg of RNA with DNase in 10 µl DNase reaction. cDNA was generated using 2 µl of DNase reaction per 10 µl RT reaction. PCR products were generated in the same manner described in Figure 4B. PCR products were run on a 6% acrylamide gel and stained in the same manner described in Figure 4B. **(D)** Splicing of *SMN1*^{G1C} minigene with different concentrations of eU1^{10A} in HeLa cells. Molecular size is indicated to the left, band identities are indicated to the right, and the minigenes transfected in HeLa cells are indicated at the top of the gel. HeLa cells were plated as described in Figure 6B. Cells were transfected with 0.1 µg of mini-gene and 0.2 µg, 0.5 µg, or 1 µg of eU1^{10A} snRNA expression vector. GFP expression vector was added when needed to maintain a total amount of 2 µg of DNA in each transfection. Total RNA was isolated from HeLa cells 23 hours after transfection followed by treating 6 µg of RNA with DNAse in 20 µl DNase reaction. cDNA was generated in the same manner described in Figure 5B. PCR products were generated in the same manner described in Figure 4B, except 1 µl cDNA was used per 25 µl Taq PCR reaction. The PCR products were run on a 6% acrylamide gel and stained in the same manner described in Figure 4B.

Figure 7. The effects of eU1 snRNAs targeting different splice donor sites in the *SMN1*^{G1C} minigene. **(A)** eU1 snRNAs and their annealing sites within intron 7 in *SMN1*^{G1C}. Mutated nucleotides in are indicated by red letters. Exon 7 sequence is shown in upper-case letters while intron 7 sequence is shown in lower-case letters. Mutated 5′ss is shown in red. Distances between splice sites are indicated as nucleotides (nt). GU dinucleotides are marked with blue squares. Numbering starts from the first position of intron 7. **(B)** Splicing patterns of *SMN1*^{G1C} minigene in the presence of different eU1 snRNAs in HeLa cells. Product identities and their sizes are indicated on the right side of the gel. Ex6S denotes a shortened form of exon 6 in which the GU dinucleotide at the 61st/62nd positions in exon 6 act as a cryptic splice donor site. The percentages of exon 7 were calculated from the total value of exon 7-included and exon 7-skipped products. HeLa cells were plated as described in Figure 4B. Cells were transfected with 0.05 µg of minigene and 0.05 µg of U1 snRNA expression vector. Empty pCI vector was added when needed to maintain a total amount of 0.5 µg of DNA in each transfection. Total RNA was isolated from HeLa cells 24 hours after transfection followed by DNase treating RNA. DNAse treated RNA was phenol extracted and ethanol precipitate. cDNA was generated using 0.5 µg of total RNA per 5 µl RTase reaction. PCR products were generated using primers P2 and P1 and used 4 µl of RTase reaction per 50 µl PCR reaction. PCR amplification was performed for 19 cycles in the presence of [α-^{32}P] dATP and used an annealing temperature of 60 ˚C. The PCR products were separated on a 6% acrylamide gel.

Figure 8. eU1 snRNA can alter splicing of endogenous *SMN2*. **(A)** U1 snRNA mutants and their annealing sites at the 5′ss of exon 7 and within intron 7. Red nucleotides show mutations that extend the base pairing at the splice donor site. eU1^{V2} snRNA and eU1^{V3} snRNA are shown to

activate the V2 alternative splice site. Exon 7 sequence is shown in upper-case letters while intron 7 sequence is shown in lower-case letters. **(B)** Splicing pattern of endogenous *SMN1* and *SMN2* in HeLa cells in the presence of different U1 snRNAs as indicated in Figure 8A. To distinguish between *SMN1* and *SMN2*, PCR products were digested with DdeI (Singh et al., 2011). The identity of the band (*SMN1*- or *SMN2*-derived) is indicated to the right of the gel and the presence or absence of intron 7 or exon 7 is indicated to the left of the gel. HeLa cells were plated as described in Figure 4B. Cells were transfected 0.05 µg of a given U1 snRNA expression vector and 0.45 µg of pCI vector. Total RNA was isolated from HeLa cells and treated with DNase in the same manner described in Figure 6C. PCR products were generated using primers P25 and P31 and used 1 µl of RT reaction per 25 µl Taq PCR reaction. PCR amplification was performed for 28 cycles and used an annealing temperature of 54˚C. PCR products were digested with DdeI overnight and then run on a 6% acrylamide gel and stained in the same manner described in Figure 4B.

Figure 9. Protein expression in FLAG-tagged *SMN* minigenes. **(A)** Diagrammatic representation and protein expression from *FLAG-SMN1*^{G1C} minigene. HeLa cells were transfected with various minigenes and the translated protein was analyzed by Western blot using an anti-FLAG antibody conjugated to horseradish peroxidase. *FLAG-SMN1* was identical to *FLAG-SMN1*^{G1C} except the first position of intron 7 did not carry the G-to-C mutation (shown in red). *FLAG-cSMN* contained the nine *SMN1* exons with no introns. *FLAG-cSMNΔ7* contained all exons except exon 7 and had no introns. The molecular weight ladder is indicated to the left of the blot and the size for full-length FLAG-SMN (38.4 kDa) and truncated FLAG-SMNΔ7 (36.96 kDa) are indicated to the right. HeLa cells were plated in the same manner described in Figure 6B. Cells were

transfected with 2 µg of a given FLAG-tagged *SMN* mini-gene. Total protein was isolated from HeLa cells 24 hours after transfection. A total of 10 µg of protein was used for samples corresponding to *SMN1*, *cSMN*, *SMN1*^{G1C}, while a total of 20 µg of protein was used for the sample corresponding to *cSMN*Δ*E7*. Lysates were run on 11% SDS-polyacrylamide gel. The image was developed using an exposure time of 13 seconds. **(B)** Titration experiment to determine the required amount of *FLAG-SMN1* minigene to observe FLAG-tagged SMN expression. Minigenes are the same as described in **(A)**. Molecular-weight marker is indicated to the left of the gel. HeLa cells were plated in the same manner described in Figure 6B. Cells were transfected with 0.2 µg, 0.6 µg, or 2 µg of FLAG-tagged mini-gene. pCI was added when necessary to maintain a total of 2 µg of DNA in each transfection. Protein was isolated from HeLa cells and prepared as described in Figure 9A. A total of 10 µg of protein was used per sample. Proteins were resolved as described in Figure 9A. The image was developed using an exposure time of 11 seconds. **(C)** Diagrammatic representation and splicing pattern of *SMNI6*Δ*7* minigene, which is designed to simulate intron 6 retention in *SMN1*. HeLa cells were plated in the same manner described in Figure 6B and were transfected with 2 µg of *SMNI6*Δ*7* minigene. Total RNA was isolated 24 hours after transfection from an aliquot of HeLa cells collected in 200 µl of DPBS Buffer. DNase treatment of RNA and generation of cDNA were done in the same manner described in Figure 6C. PCR products were generated using primers 5CTRL and PS-SMN-exon8 and used 1 µl of RT reaction per 25 µl Taq PCR reaction. PCR amplification conditions were the same as described in Figure 4B. The PCR products were run on a 6% acrylamide gel and stained in the same manner described in Figure 4B. **(D)** FLAG western blot to examine protein expression in HeLa cells. The minigenes transfected in HeLa cells are indicated at the top of the blot. Molecular-weight marker appears to the left of the blot. The full-

length FLAG-SMN (38.4 kDa) and FLAG-SMNΔ7 (36.96 kDa) bands are indicated to the right of the gel. HeLa cells were plated in the same manner described in Figure 6B. Cells were transfected with 0.6 µg of a FLAG-tagged *SMN* mini-gene and 0.6 µg, 1 µg, or 1.4 µg of eU1^{V3} snRNA expression vector. Protein was isolated from HeLa cells and prepared as described in Figure 9A. While a total of 10 µg of protein was used for samples, a total of 15 µg of protein was used for the sample corresponding to *cSMNΔE7*. Proteins were resolved and probed as described in Figure 9A. The image was developed using an exposure time of 12 seconds.

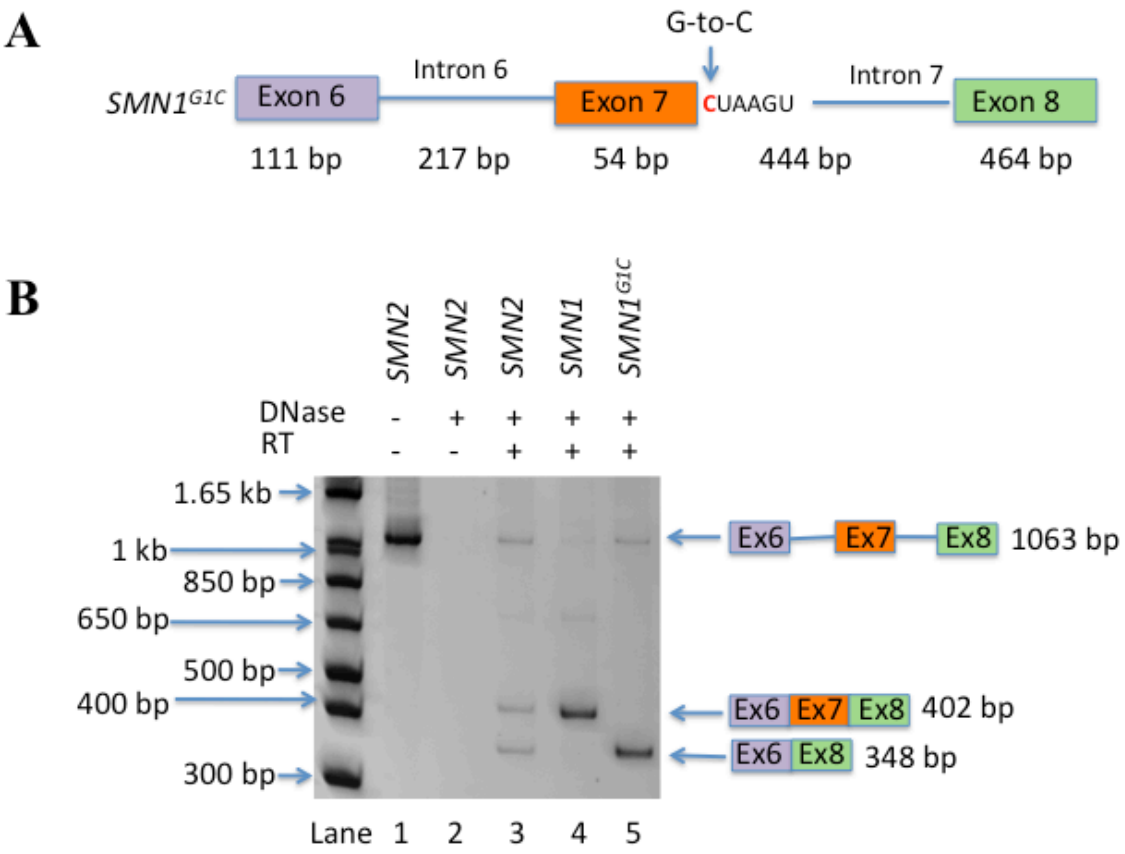


Figure 4. Splicing patterns of various *SMN* minigenes.

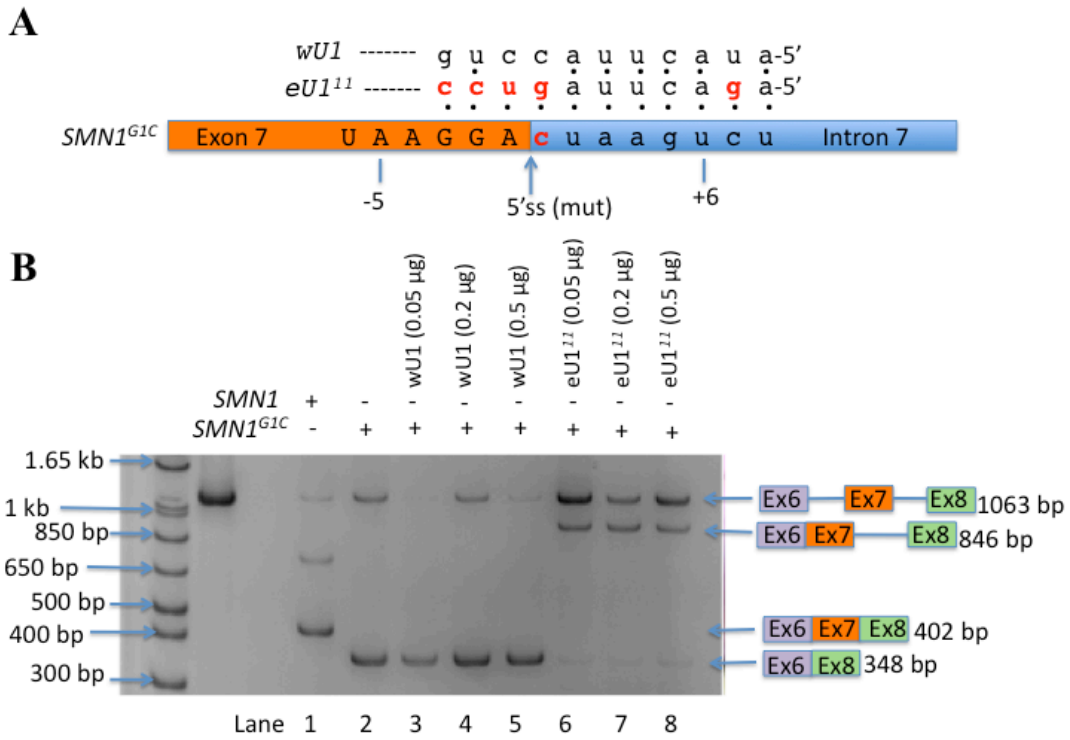
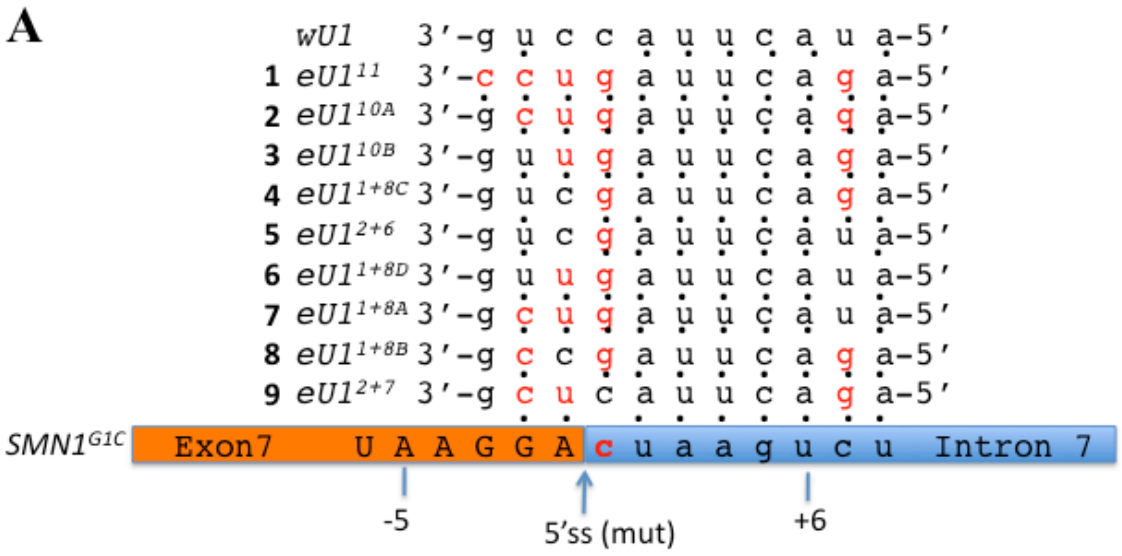
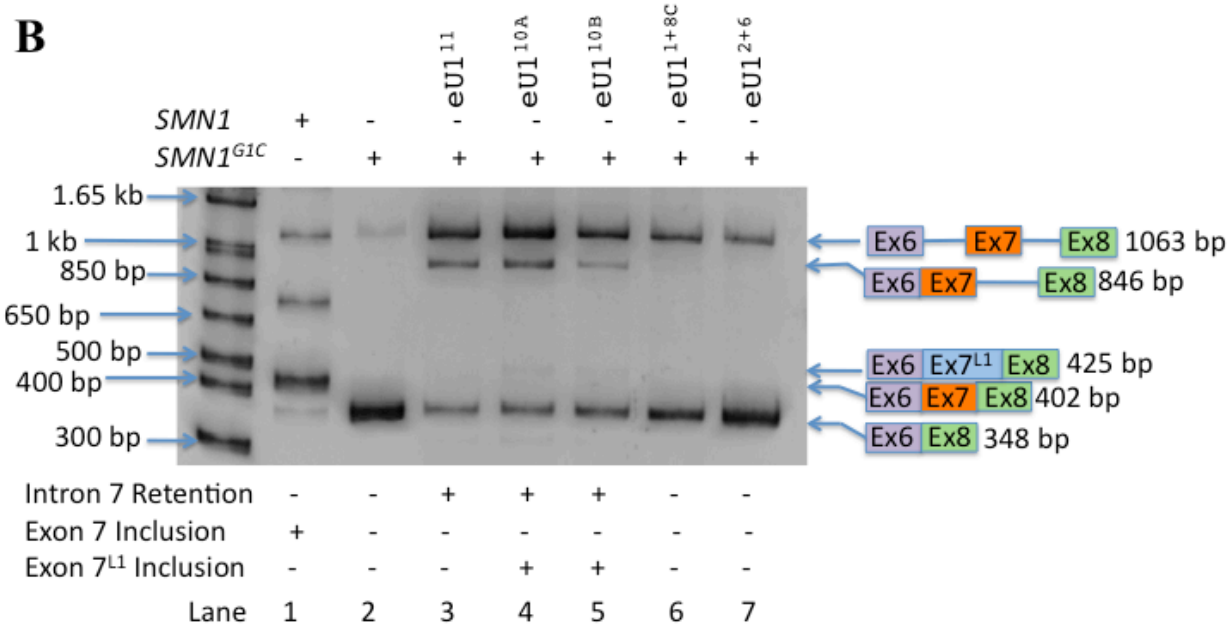


Figure 5. eU1 snRNA with extended base pairing at the 5′ splice site (5′ss) of intron 7 in the *SMN1*^{G1C} minigene promotes intron 7 retention.





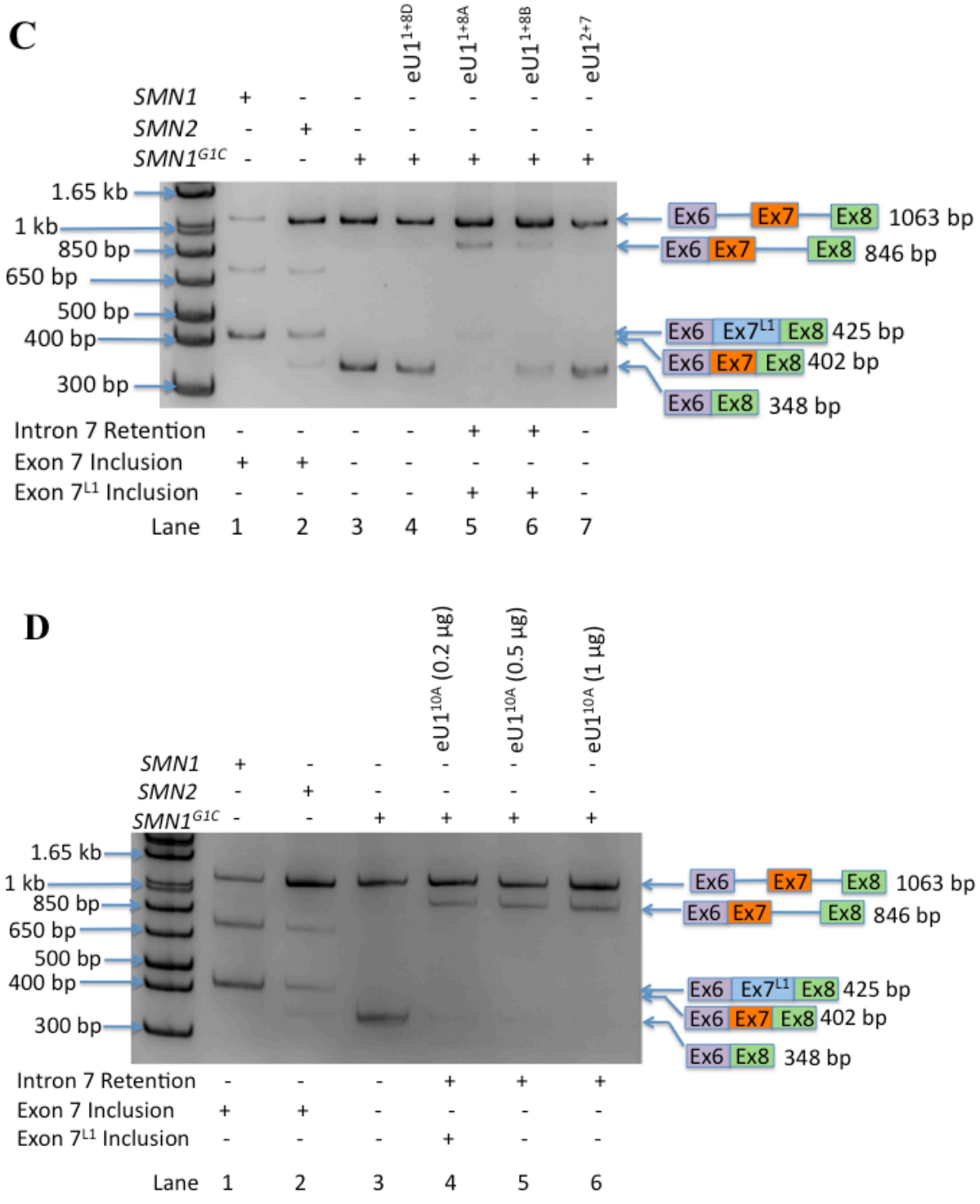


Figure 6. The effect of eU1 snRNAs with differing degrees of extended base pairing with the 5′ ss of intron 7 on *SMN1*^{G1C} splicing.

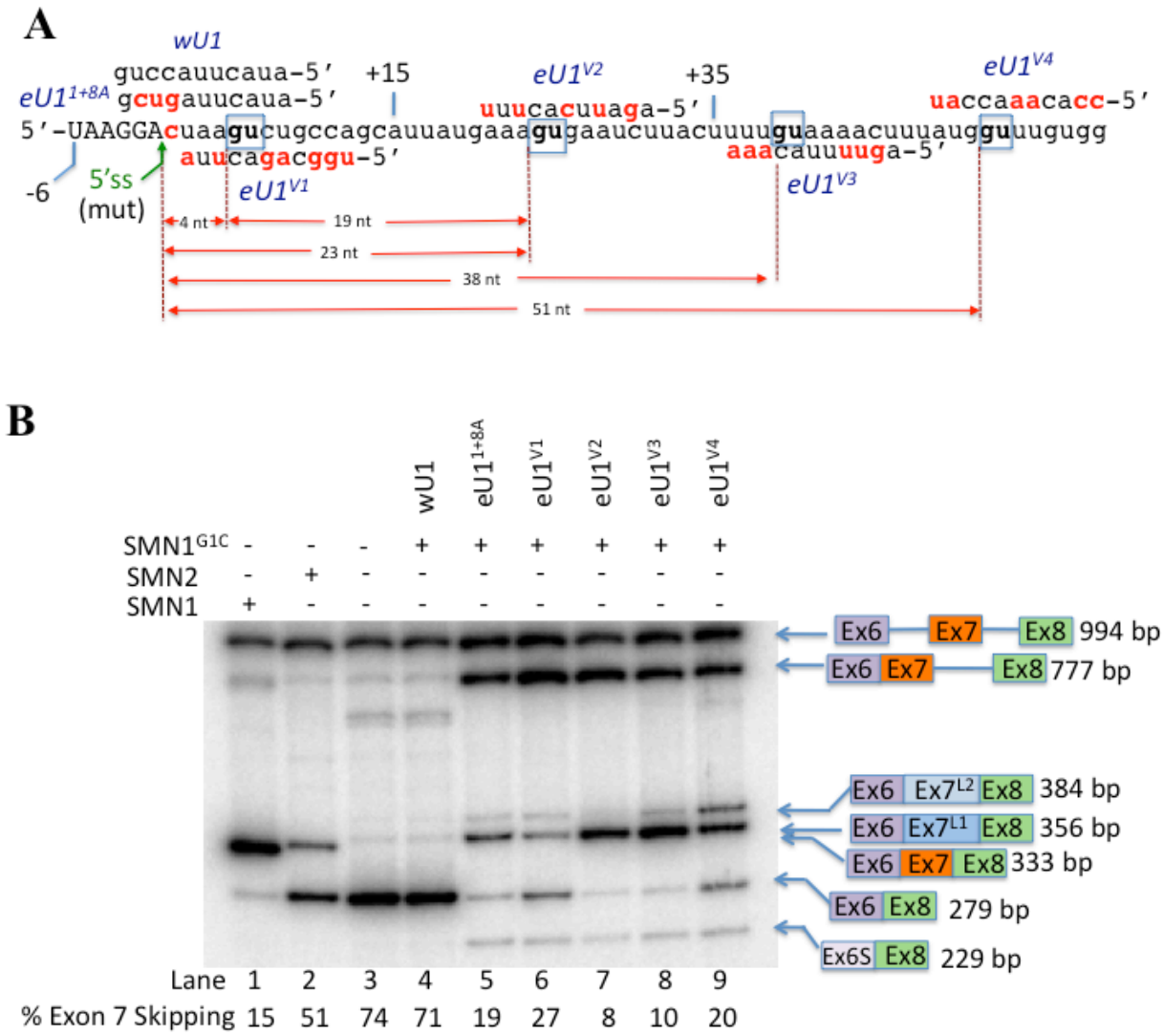


Figure 7. The effects of eU1 snRNAs targeting different splice donor sites in the *SMN1*^{G1C} minigene.

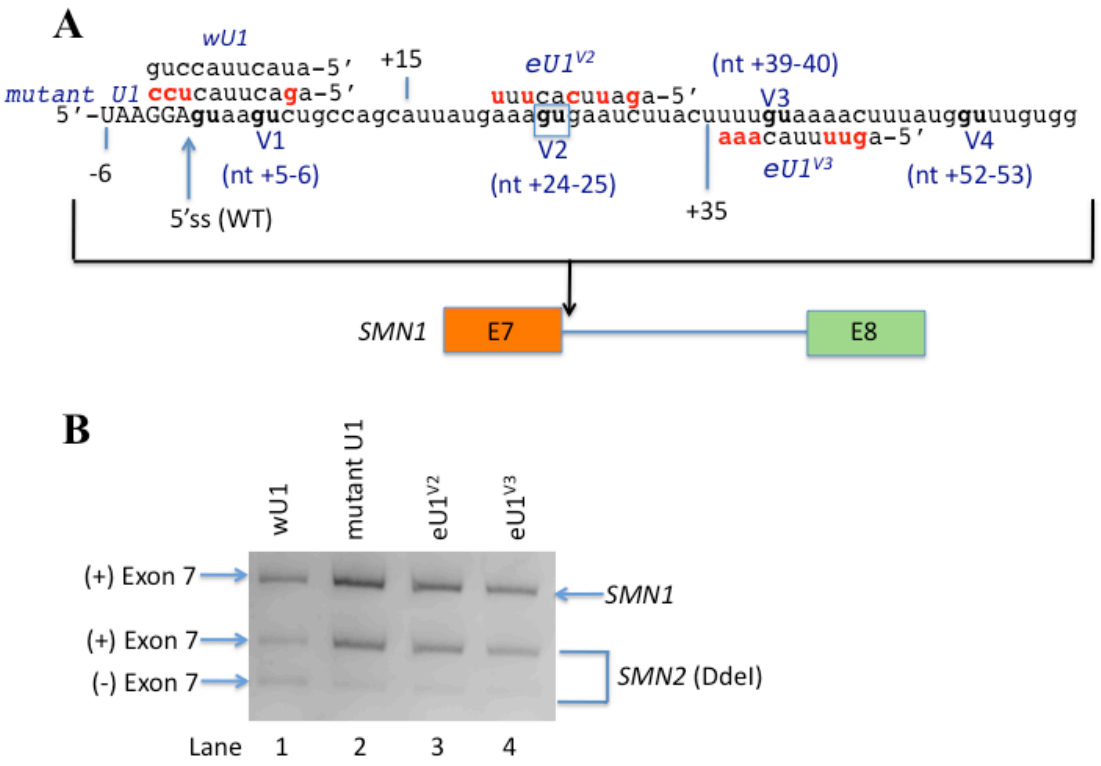
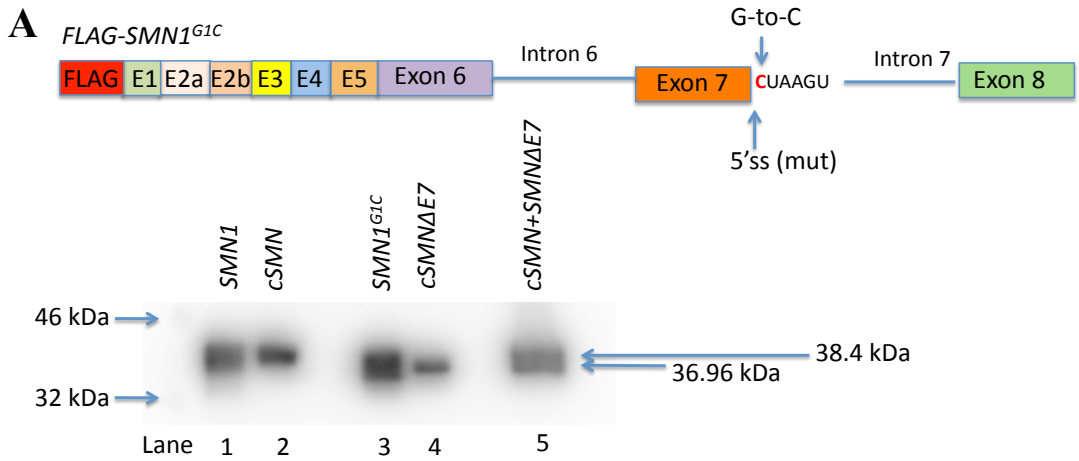
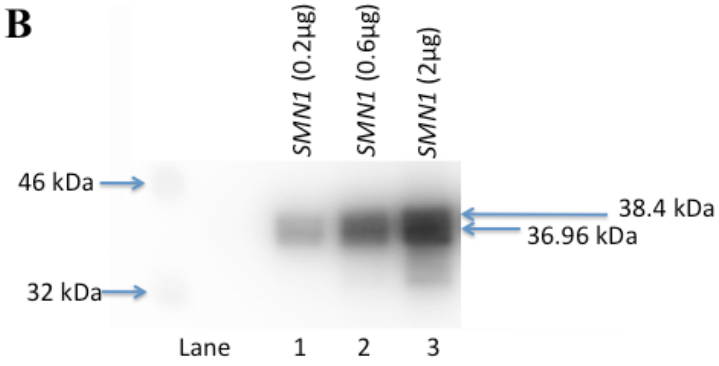
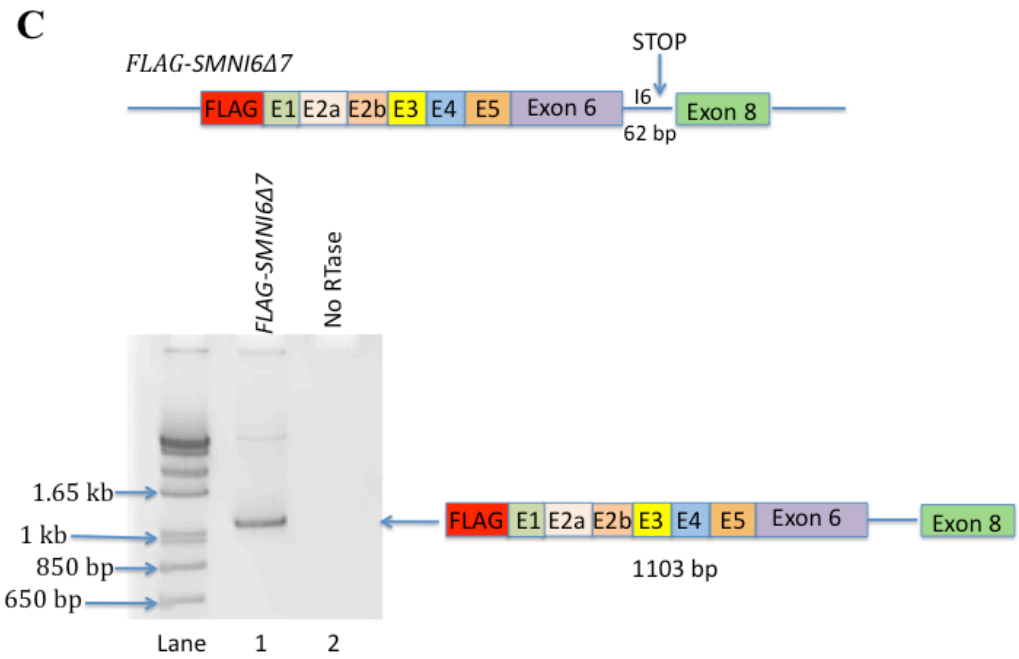


Figure 8. eU1 snRNA can alter splicing of endogenous *SMN1* and *SMN2*.







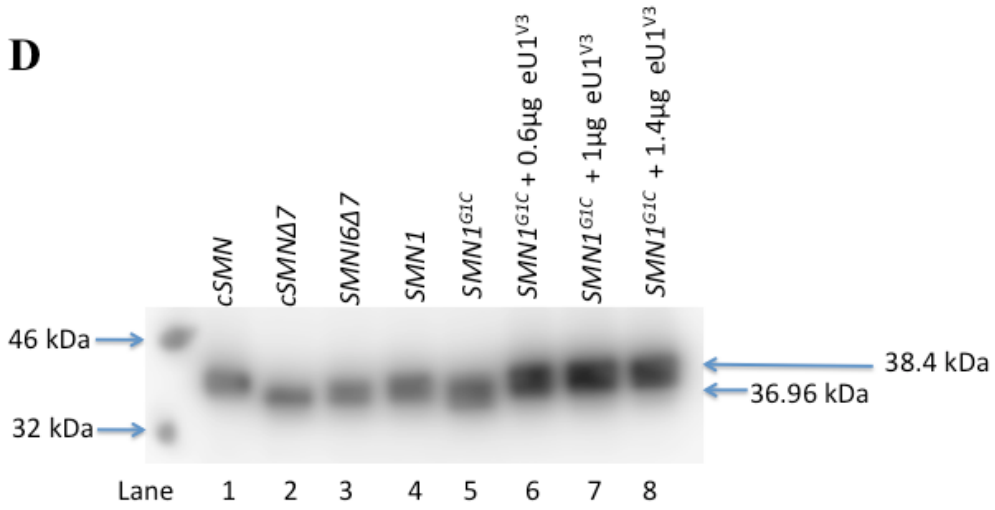


Figure 9. Protein expression in FLAG-tagged *SMN* minigenes.

CHAPTER 3. GENERAL CONCLUSIONS

Mutations that cause aberrant splicing are present in a variety of genetic diseases. In patients afflicted with Spinal muscular atrophy (SMA), a C-to-T mutation in exon 7 that weakens the 5′ss of the exon triggers exon 7 skipping and prevents *SMN2* from being able to compensate for the loss of *SMN1*. In deleterious *SMN1* alleles such as $SMN1^{G1C}$, mutations at the 5′ss of exon 7 induce skipping by preventing intron 7 from being spliced out. U1 snRNPs with snRNAs that have extended base pairing with the 5′ss or an alternative splice donor site could be used to correct splicing in deleterious *SMN1* alleles as well as *SMN2*.

In chapter 2, we demonstrated that expression vectors for engineered U1 snRNAs (eU1s) could be used to inhibit exon 7 skipping in both $SMN1^{G1C}$ and *SMN2*. eU1s that target the mutated 5′ss of exon 7 in $SMN1^{G1C}$ promote intron 7 retention and weakly activate splice site (V2) that can redefine exon 7. The eU1s designed to bind to potential splice sites downstream of the mutated exon 7 splice donor site can promote the usage of splice sites at the 23rd position and the 51st position of intron 7 more effectively than the eU1s directly targeting the mutated 5′ss of exon 7. eU1s targeting downstream of the 5′ss of exon 7 can also prevent exon 7 skipping in *SMN2*. Additionally, eU1s targeting alternative splice sites prevents exon 7 skipping to a great enough extent to promote full-length SMN protein in $SMN1^{G1C}$. Overall, the study indicates the identification of "novel" splice donor sites in *SMN* genes and that the eU1s targeting the alternative sites have potential to be a general treatment for SMA patients, and possibly other patients with similar genetic disorders.

APPENDIX A

PRIMERS

Name	Type	Sequence (5′ to 3′)	Tm (˚C)	Anneal site
5CTRL	Forward	GAATCTCGAGATGGACTACAAAGACCATGACG G	62.2	FLAG
3In7a	Reverse	CTTTAGTGGTGTCATTTACAAAACAAAGTC	55.3	I6
3′pUCBU1	Reverse	ATCCTCGAGCCTCCACTGTAGGATTAAC	60.9	PUCBU1
3SMN1E7I7	Reverse	CTTTCATAATGCTGGCAGACTTAGTCCTTAATTT AAGG	60	E7
3SMN1Ex8	Reverse	TCGAAGCGGCCGCCCGGGCACATACG	72.1	E8
5Ex8a	Forward	GACTTTGTTTTGTAAATGACACCACTAAAG	55.3	E8
5′pUCBU-mut1	Forward	CGAAGATCTCATACTTAGCTGGCAGGGGAGAT	63.2	PUCBU1
5′pUCBU-mut2	Forward	CGAAGATCTCAGACTTAGCTGGCAGGGGAGAT	64.6	PUCBU1
5′pUCBU-mut3	Forward	CGAAGATCTCAGACTTAGTTGGCAGGGGAGAT	63.3	PUCBU1
5′pUCBU-mut4	Forward	CGAAGATCTCAGACTTAGTCGGCAGGGGAGAT	64.4	PUCBU1
5′pUCBU-mutG1C	Forward	CGAAGATCTCAGACTTAGTCCGCAGGGGAGAT	64.4	PUCBU1
5Ex4	Forward	CCAAATCTGCTCCATGGAACTCTTTTC	58.1	E4
5SMN1E7I7	Forward	AAGGTGCTCACATTCCTTAAATTAAGGACTAAG TCTGC	62	I7
5SMN1Ex6	Forward	CTATGGCTAGCCTCGAGATAATTCC	56.4	E6
Ex8P2-2	Reverse	CTTCCTTTTTTCTTTCCCAACAC	52.8	E8
M2V2	Forward	CGAAGATCTCATACTTAGTTGGCAGGGGAGAT	61.8	PUCBU1
M3V2	Forward	CGAAGATCTCATACTTAGTCGGCAGGGGAGAT	62.9	PUCBU1
M3V3	Forward	CGAAGATCTCAGACTTAGCCGGCAGGGGAGAT	66.1	PUCBU1
M3V4	Forward	CGAAGATCTCAGACTTACTCGGCAGGGGAGAT	64.4	PUCBU1
MI7V1	Forward	CGAAGATCTCTGGCAGACTTAGCAGGGGAGAT	64.6	PUCBU1
MI7V2	Forward	CGAAGATCTCAGATTCACTTTGCAGGGGAGAT	62.4	PUCBU1
MI7V3	Forward	CGAAGATCTCAGTTTTACAAAGCAGGGGAGAT	61.1	PUCBU1
MI7V4	Forward	CGAAGATCTCCCACAAACCATGCAGGGGAGAT	65.6	PUCBU1
P1	Forward	CGACTCACTATAGGCTAGCC	54	PCI
P2	Reverse	GCATGCAAGCTTCCTTTTTTCTTTCCCAACAC	62.2	E8
P25	Reverse	CTCGAAGCGGCCGCAGCTCATAAAATTACCA	65	E8
P31	Forward	CATCAGTGGCTATCATACTG	50	PCI
PCI-DN	Reverse	AGCATCACAAATTTCACAAATAAA	50.2	PCI
PCI-UP	Forward	TGACATCCACTTTGCCTTTCTCTC	57.1	PCI

Supplementary Table 1. List of primers used for PCR. E and I are abbreviations for exon and intron respectively. Tm indicates the melting point of a primer in degrees Celsius.

APPENDIX B

MUTATIONS DISRUPTING RNA SPLICING

Supplementary Table 2. Documented Mutations Disrupting U1 snRNP Activity

Gene	Location[a]	Mutation[b]	Predicted Change[c]	Reference(s)
FUS	E14	c.1483C>T	p.R495X	(Calvo et al., 2014;Yu et al., 2015; Sun et al., 2015)
FUS	E15	c.1561C>T	p.R521G	(Calvo et al., 2014;Yu et al., 2015; Sun et al., 2015)
FUS	E15	c.1574C>T	p.P525L	(Calvo et al., 2014;Yu et al., 2015; Sun et al., 2015)
SNRPB	I2	c.164G>C	Mutation in intron 2 promotes alternative exon with PTC	(Lynch et al., 2014; Bacrot et al., 2015)
SNRPB	I2	c.164G>T	Mutation in intron 2 promotes alternative exon with PTC	(Lynch et al., 2014; Bacrot et al., 2015)
SNRPB	I2	c.165G>C	Mutation in intron 2 promotes alternative exon with PTC	(Lynch et al., 2014; Bacrot et al., 2015)
SNRPB	I2	c.166G>C	Mutation in intron 2 promotes alternative exon with PTC	(Lynch et al., 2014; Bacrot et al., 2015)
SNRPB	I2	c.213+57C>A	Mutation in intron 2 promotes alternative exon with PTC	(Bacrot et al., 2015)

[a] E refers to exon; I refers to intron.

[b] Nucleotide numbering refers to the coding sequence with A of ATG denoted as +1.

[c] Amino acids are indicated by their three letter codes; an X denotes a premature stop codon.

Supplementary Table 2. Documented Mutations Disrupting Splice Sites

Gene	Location[a]	Mutation[b]	Predicted Change	Reference(s)
IKBKAP	I20	c.2204+6T>C	Mutation in intron 20 splice donor site	(Slaugenhaupt et al., 2001; Ibrahim et al., 2007)
BBS1	E5	c.479G>A	Mutation in exon 5 splice donor site	(Schmid et al., 2011)
SPINK5	I2	c.81+2T>A	Mutation in conserved consensus intron 2 donor splice site; exon 2 skipped	(Bitoun et al., 2002)
SPINK5	E1	c.56G>A	Mutation in exon 1 splice donor site	(Bitoun et al., 2002)
SPINK5	I20	c.1888-1G>A	Mutation in conserved consensus intron 20 acceptor splice site; exon 21 skipped	(Bitoun et al., 2002)
SPINK5	I23	c.2240+1G>A	Mutation in conserved consensus intron 23 donor splice site; exon 23 skipped	(Bitoun et al., 2002)
SPINK5	I2	c.81+5G>A	Mutation in intron 2 splice donor site	(Bitoun et al., 2002)
SPINK5	I17	c.1698-1G>A	Mutation in conserved consensus intron 17 acceptor splice site; exon 18 skipped	(Bitoun et al., 2002)
SPINK5	E11	c.891C>T	Mutation in exon 11 promotes hnRNP binding; exon 11 skipped	(Dal Mas, Fortugno, et al., 2015)
CFTR	I5	c.711+1G>T	Mutation in conserved consensus intron 5 donor splice site; exon 5 skipped	(Fanen et al., 1992)
CFTR	I10	c.1717-1G>A	Mutation in conserved consensus intron 10 acceptor splice site; exon 11 skipped	(Fanen et al., 1992)
CFTR	I11	c.1811+1600A>G	Mutation in intron 11 produces a new exon	(Chillón et al., 1995)
CFTR	I14b	c.2789+5G>A	Mutation in intron 14b splice donor site	(Fanen et al., 1992)
CFTR	I17a	c.3272-26A>G	Mutation in intron 17a produces alternative acceptor site	(Fanen et al., 1992)
CFTR	I23	c.4374+1G>A	Mutation in conserved consensus intron 23 donor splice site; exon 23 skipped	(Fanen et al., 1992)
hSNF5	I7	c.986+1G>A	Mutation in conserved consensus intron 7 donor splice site; exon 7 skipped	(Venables, 2004; Eaton et al., 2011)
APC	I3	c.423-1G>T	Mutation in conserved consensus intron 3 acceptor splice site; exon 4 skipped	(Venables, 2004; Friedl & Aretz, 2005)
CCND1	E4	c.870G>A	Mutation in exon 4 splice donor site	(David & Manley, 2010)
SMN1	I7	c.888+1G>C	Mutation in conserved consensus intron 7 donor splice site; exon 7 skipped	(Ronchi et al., 2015)

[a] E refers to exon; I refers to intron

[b] Nucleotide numbering refers to the coding sequence with A of ATG denoted as +1.

REFERENCES

Bacrot S, Doyard M, Huber C, Alibeu O, Feldhahn N, Lehalle D, Lacombe D, Marlin S, Nitschke P, Petit F, Vazquez MP (2015). Mutations in SNRPB, Encoding Components of the Core Splicing Machinery, Cause Cerebro-Costo-Mandibular Syndrome. Human Mutation *36(2)*, 187-190. doi: 10.1002/humu.22729.

Bitoun E, Chavanas S, Irvine AD, Lonie L, Bodemer C, Paradisi M, Hamel-Teillac D, Ansai SI, Mitsuhashi Y, Taïeb A, de Prost Y (2002). Netherton syndrome: disease expression and spectrum of SPINK5 mutations in 21 families. Journal of investigative dermatology *118(2)*, 352-361. doi: 10.1046/j.1523-1747.2002.01603.x.

Calvo A, Moglia C, Canosa A, Brunetti M, Barberis M, Traynor BJ, Carrara G, Valentini C, Restagno G, Chiò A (2014). A de novo nonsense mutation of the FUS gene in an apparently familial amyotrophic lateral sclerosis case. Neurobiology of Aging *35(6)*, 1513-e7. doi:10.1016/j.neurobiolaging.2013.12.028.

Chillon M, Dörk T, Casals T, Gimenez J, Fonknechten N, Will K, Ramos D, Nunes V, Estivill X (1995). A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+ 1.6 kbA--> G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. American journal of human genetics *56(3)*, 623. PMCID: PMC1801150.

Dal Mas A, Fortugno P, Donadon I, Levati L, Castiglia D, Pagani F (2015). Exon-Specific U1s Correct SPINK5 Exon 11 Skipping Caused by a Synonymous Substitution that Affects a Bifunctional Splicing Regulatory Element. Human Mutation *36(5)*, 504-12. doi: 10.1002/humu.22762

David CJ, Manley JL (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev. *24*, 2343-2364. doi: 10.1101/gad.1973010.

Eaton KW, Tooke LS, Wainwright LM, Judkins AR, Biegel JA (2011). Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors. Pediatric blood & cancer *56(1)*, 7-15. doi: 10.1002/pbc.22831.

Fanen P, Ghanem N, Viduad M, Besmond C, Martin J, Costes B, Plassa F, Goossens M (1992). Molecular Characterization of Cystic Fibrosis: 16 Novel Mutations Identified by Analysis of the Whole Cystic Fibrosis Conductance Transmembrane Regulator (CFTR) Coding Regions and Splice Site Junctions. Genomics *13*, 770-776. doi: 10.1016/0888-7543(92)90152-I.

Friedl W, Aretz S (2005). Familial adenomatous polyposis: experience from a study of 1164 unrelated german polyposis patients. Hereditary Cancer in Clinical Practice *3(3)*, 95-114. doi: 10.1186/1897-4287-3-3-95.

Ibrahim EC, Hims MM, Shomron N, Burge CB, Slaugenhaupt SA, Reed R (2007). Weak definition of IKBKAP exon 20 leads to aberrant splicing in familial dysautonomia. Human Mutation *28(1)*, 41-53. doi: 10.1002/humu.20401.

Lynch DC, Revil T, Schwartzentruber J, et al (2014). Disrupted auto-regulation of the spliceosomal gene SNRPB causes cerebro–costo–mandibular syndrome. Nature Communications *5*, 4483. doi: 10.1038/ncomms5483.

Ronchi D, Previtali SC, Sora MG, Barera G, Del Menico B, Corti S, Bresolin N, Comi GP (2015). Novel splice-site mutation in SMN1 associated with a very severe SMA-I phenotype. J Mol Neurosci. *56(1)*, 212-215. doi: 10.1007/s12031-014-0483-4.

Schmid F, Glaus E, Barthelmes D, Fliegauf M, Gaspar H, Nürnberg G, Nürnberg P, Omran H, Berger W, Neidhardt J (2011). U1 snRNA-mediated gene therapeutic correction of splice defects caused by an exceptionally mild BBS mutation. Human mutation *32(7)*, 815-824. doi: 10.1002/humu.21509.

Slaugenhaupt SA, Blumenfeld A, Gill SP, Leyne M, Mull J, Cuajungco MP, Liebert CB, Chadwick B, Idelson M, Reznik L, Robbins CM (2001). Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. The American Journal of Human Genetics *68(3)*, 598-605. doi: 10.1086/318810.

Sun S, Ling SC, Qiu J, Albuquerque CP, Zhou Y, Tokunaga S, Li H, Qiu H, Bui A, Yeo GW, Huang EJ (2015). ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. Nature communications *6*, 6171. doi: 10.1038/ncomms7171.

Venables JP (2004). Aberrant and Alternative Splicing in Cancer. Cancer Research *64*, 7647-7654. doi: 10.1158/0008-5472.CAN-04-1910.

Yu Y, Chi B, Xia W, Gangopadhyay J, Yamazaki T, Winkelbauer-Hurt ME, Yin S, Eliasse Y, Adams E, Shaw CE, Reed R (2015). U1 snRNP is mislocalized in ALS patient fibroblasts bearing NLS mutations in FUS and is required for motor neuron outgrowth in zebrafish. Nucleic Acids Research *43(6)*, 3208-3218. doi: 10.1093/nar/gkv157.

APPENDIX C

SMN MUTATIONS AT EXON 7 AND INTRON 7 IN SMA PATIENTS

Supplementary Table 3. Documented SMN mutations at exon 7 and intron 7 in SMA patients.

Location[a]	Nucleotide Change[b]	Predicted Change[c]	SMA Patient Types	Reference(s)
Missense Mutations				
E7	c.835 G>T	p.Gly279Cys	II, III	(Wang et al., 1998)
E7	c.836 G>T	p.Gly279Cys	I	(Hahnen et al., 1997)
E7	c.859 G>C[d]	p.Gly287Arg	II, III	(Bernal et al., 2010; Verzain et al., 2010)
E7	c.863 G>T	p.Arg288Met	I, II	(Qu et al., 2015)
Nonsense Mutations				
E7	c.861_862insT	p.Arg288X	II	(Ganji et al., 2015)
Splice Site Mutations				
I7	c.888+1G>C	Mutation in conserved consensus intron 7 donor splice site; exon 7 skipped	I	(Ronchi et al., 2015)
I7	c.889+3delAGTC	Deletion at intron 7 splice donor site	II	(Lefebvre et al., 1995)
I7	c.889+6T>G	Mutation in intron 7 splice donor site	III	(Wirth et al., 1999)

[a] E refers to exon; I refers to intron

[b] Nucleotide numbering refers to the coding sequence with A of ATG denoted as +1.

[c] Amino acids are indicated by their three letter codes; an X denotes a premature stop codon

[d] Mutation reported only in *SMN2*

REFERENCES

Bernal S, Alías L, Barceló MJ, Also-Rallo E, Martínez-Hernández R, Gámez J, Guillén-Navarro E, Rosell J, Hernando I, Rodríguez-Alvarez FJ, Borrego S, Millán JM, Hernández-Chico C, Baiget M, Fuentes-Prior P, Tizzano EF (2010). The c.859G>C variant in the *SMN2* gene is associated with types II and III SMA and originates from a common ancestor. *J. Med. Genet.* 47(9), 640–642. doi: 10.1136/jmg.2010.079004.

Ganji H, Nouri N, Salehi M, Aryani O, Houshmand M, Basiri K, Fazel-Najafabadi E, Sedghi M (2015). Detection of intragenic *SMN1* mutations in spinal muscular atrophy patients with a single copy of *SMN1*. *Journal of Child Neurology 30*(5), 558-562. doi: 10.1177/0883073814521297.

Hahnen E, Schönling J, Rudnik-Schöneborn S, Raschke H, Zerres K, Wirth B (1997). Missense mutations in exon 6 of the survival motor neuron gene in patients with spinal muscular atrophy (SMA). *Hum. Mol. Genet.* 6(5), 821–825. doi: 10.1093/hmg/6.5.821.

Lefebvre S, Bürglen 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(1), 155-65. doi:10.1016/0092-8674(95)90460-3.

Qu YJ, Bai JL, Cao YY, Zhang WH, Wang H, Jin YW, Song F (2015). A rare variant (c. 863G> T) in exon 7 of SMN1 disrupts mRNA splicing and is responsible for spinal muscular atrophy. *European Journal of Human Genetics*. doi: 10.1038/ejhg.2015.213.

Ronchi D, Previtali SC, Sora MG, Barera G, Del Menico B, Corti S, Bresolin N, Comi GP (2015). Novel splice-site mutation in SMN1 associated with a very severe SMA-I phenotype. *J Mol Neurosci.* 56(1), 212-5. doi: 10.1007/s12031-014-0483-4.

Vezain M, Saugier-Veber P, Goina E, Touraine R, Manel V, Toutain A, Fehrenbach S, Frébourg T, Pagani F, Tosi M, Martins A (2010). A rare *SMN2* variant in a previously unrecognized composite splicing regulatory element induces exon 7 inclusion and reduces the clinical severity of spinal muscular atrophy. *Hum. Mutat.* 31, E1110–E1125. doi:10.1002/humu.21173.

Wang CH, Papendick BD, Bruinsma P, Day JK (1998). Identification of a novel missense mutation of the SMN(T) gene in two siblings with spinal muscular atrophy. *Neurogenetics* 1(4), 273–276. doi**:** 10.1007/s100480050040.

Wirth B, Herz M, Wetter A, Moskau S, Hahnen E, Rudnik-Schöneborn S, Wienker T, Zerres K (1999). Quantitative analysis of survival motor neuron copies: identification of subtle *SMN1* mutations in patients with spinal muscular atrophy, genotype-phenotype correlation, and implications for genetic counseling. *Am. J. Hum. Genet.* 64(5), 1340–1356. doi: 10.1086/302369.

APPENDIX D

SEQUENCES OF *SMN* MINIGENES

Key for minigenes:
FLAG: Red
Exon 1: Yellow
Exon 2a: Bright Green
Exon 2b: Bright Blue
Exon 3: Pink
Exon 4: Blue
Exon 5: Teal
Exon 6: Green
Delta Intron 6: Purple
Exon 7: Orange
Intron 7: Dark Yellow
Exon 8: Gray
XhoI site: **CTCGAG**
NotI site: **GCGGCCG**
NdeI site: **CATATG**
ApoI site: **RAATTY**

SMN1

CTCGAGATAATTCCCCCACCACCTCCCATATGTCCAGATTCTCTTGATGATGCTGAT
GCTTTGGGAAGTATGTTAATTTCATGGTACATGAGTGGCTATCATACTGGCTATTAT
ATGGTAAGTAATCACTCAGCATCTTTTCCTGACAATTTTTTTGTAGTTATGTGACTTT
GTTTTGTAAATTTATAAAATACTACTTGCTTCTCTCGGTACCCGGGTCTAGACGCGTG
TCTTGTGAAACAAAATGCTTTTTAACATCCATATAAAGCTATCTATATATAGCTATCT
ATGTCTATATAGCTATTTTTTTTAACTTCCTTTATTTTCCTTACAGGGTTTCAGACAAA
ATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAGGAGTAAGTCTGCCAGCA
TTATGAAAGTGAATCTTACTTTTGTAAAACTTTATGGTTTGTGGAAAACAAATGTTTT
TGAACATTTAAAAAGTTCAGATGTTAAAAAGTTGAAAGGTTAATGTAAAACAATCA
ATATTAAAGAATTTTGATGCCAAAACTATTAGATAAAAGGTTAATCTACATCCCTAC
TAGAATTCTCATACTTAACTGGTTGGTTATGTGGAAGAAACATACTTTCACAATAAA
GAGCTTTAGGATATGATGCCATTTTATATCACTAGTAGGCAGACCAGCAGACTTTTT
TTTATTGTGATATGGGATAACCTAGGCATACTGCACTGTACACTCTGACATATGAAG
TGCTCTAGTCAAGTTTAACTGGTGTCCACAGAGGACATGGTTTAACTGGAATTCGTC
AAGCCTCTGGTTCTAATTTCTCATTTGCAGGAAATGCTGGCATAGAGCAGCACTAAA
TGACACCACTAAAGAAACGATCAGACAGATCTGGAATGTGAAGCGTTATAGAAGAT
AACTGGCCTCATTTCTTCAAAATATCAAGTGTTGGGAAAGAAAAAAGGAAGTGGAA
TGGGTAACTCTTCTTGATTAAAAGTTATGTAATAACCAAATGCAATGTGAAATATTT
TACTGGACTCTATTTTGAAAAACCATCTGTAAAAGACTGGGGTGGGGGTGGGAGGC
CAGCACGGTGGTGAGGCAGTTGAGAAAATTTGAATGTGGATTAGATTTTGAATGAT
ATTGGATAATTATTGGTAATTTTATGAGCTGTGAGAAGGGTGTTGTAGTTTATAAAA
GACTGTCTTAATTTGCATACTTAAGCATTTAGGAATGAAGTGTTAGAGTGTCTTAAA
ATGTTTCAAATGGTTTAACAAAATGTATGTGAGGCGTATGTGCCCGG**GCGGCCGC**

SMN2

CTCGAGATAATTCCCCCACCACCTCCCATATGTCCAGATTCTCTTGATGATGCTGAT
GCTTTGGGAAGTATGTTAATTTCATGGTACATGAGTGGCTATCATACTGGCTATTAT
ATGGTAAGTAATCACTCAGCATCTTTTCCTGACAATTTTTTTGTAGTTATGTGACTTT
GTTTTGTAAATTTATAAAATACTACTTGCTTCTCTCGGTACCCGGGTCTAGACGCGTG
TCTTGTGAAACAAAATACTTTTTAACATCCATATAAAGCTATCTATATATAGCTATCT
ATGTCTATATAGCTATTTTTTTTAACTTCCTTTATTTTCCTTACAGGGTTTTAGACAAA
ATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAGGAGTAAGTCTGCCAGCA
TTATGAAAGTGAATCTTACTTTTGTAAAACTTTATGGTTTGTGGAAAACAAATGTTTT
TGAACATTTAAAAAGTTCAGATGTTAGAAAGTTGAAAGGTTAATGTAAAACAATCA
ATATTAAAGAATTTTGATGCCAAAACTATTAGATAAAAGGTTAATCTACATCCCTAC
TAGAATTCTCATACTTAACTGGTTGGTTGTGTGGAAGAAACATACTTTCACAATAAA
GAGCTTTAGGATATGATGCCATTTTATATCACTAGTAGGCAGACCAGCAGACTTTTT
TTTATTGTGATATGGGATAACCTAGGCATACTGCACTGTACACTCTGACATATGAAG
TGCTCTAGTCAAGTTTAACTGGTGTCCACAGAGGACATGGTTTAACTGGAATTCGTC
AAGCCTCTGGTTCTAATTTCTCATTTGCAGGAAATGCTGGCATAGAGCAGCACTAAA
TGACACCACTAAAGAAACGATCAGACAGATCTGGAATGTGAAGCGTTATAGACGAT
AACTGGCCTCATTTCTTCAAAATATCAAGTGTTGGGAAAGAAAAAAGGAAGTGGAA
TGGGTAACTCTTCTTGATTAAAGGTTATGTAATAACCAAATGCAATGTGAAATATTT
TACTGGACTCTATTTTGAAAAACCATCTGTAAAAGACTGGGGTGGGGGTGGGAGGC
CAGCACGGTGGTGAGGCAGTTGAGAAAATTTGAATGTGGATTAGATTTTGAATGAT
ATTGGATAATTATTGGTAATTTTATGAGCTGTGAGAAGGGTGTTGTAGTTTATAAAA
GACTGTCTTAATTTGCATACTTAAGCATTTAGGAATGAAGTGTTAGAGTGTCTTAAA
ATGTTTCAAATGGTTTAACAAAATGTATGTGAGGCGTATGTGCCCGG**GCGGCCGC**

SMN1^{G1C}

CTCGAGATAATTCCCCCACCACCTCCCATATGTCCAGATTCTCTTGATGATGCTGAT
GCTTTGGGAAGTATGTTAATTTCATGGTACATGAGTGGCTATCATACTGGCTATTAT
ATGGTAAGTAATCACTCAGCATCTTTTCCTGACAATTTTTTTGTAGTTATGTGACTTT
GTTTTGTAAATTTATAAAATACTACTTGCTTCTCTCGGTACCCGGGTCTAGACGCGTG
TCTTGTGAAACAAAATGCTTTTTAACATCCATATAAAGCTATCTATATATAGCTATCT
ATGTCTATATAGCTATTTTTTTTAACTTCCTTTATTTTCCTTACAGGGTTTCAGACAAA
ATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAGGACTAAGTCTGCCAGCA
TTATGAAAGTGAATCTTACTTTTGTAAAACTTTATGGTTTGTGGAAAACAAATGTTTT
TGAACATTTAAAAAGTTCAGATGTTAAAAAGTTGAAAGGTTAATGTAAAACAATCA
ATATTAAAGAATTTTGATGCCAAAACTATTAGATAAAAGGTTAATCTACATCCCTAC
TAGAATTCTCATACTTAACTGGTTGGTTATGTGGAAGAAACATACTTTCACAATAAA
GAGCTTTAGGATATGATGCCATTTTATATCACTAGTAGGCAGACCAGCAGACTTTTT
TTTATTGTGATATGGGATAACCTAGGCATACTGCACTGTACACTCTGACATATGAAG
TGCTCTAGTCAAGTTTAACTGGTGTCCACAGAGGACATGGTTTAACTGGAATTCGTC
AAGCCTCTGGTTCTAATTTCTCATTTGCAGGAAATGCTGGCATAGAGCAGCACTAAA
TGACACCACTAAAGAAACGATCAGACAGATCTGGAATGTGAAGCGTTATAGAAGAT
AACTGGCCTCATTTCTTCAAAATATCAAGTGTTGGGAAAGAAAAAAGGAAGTGGAA
TGGGTAACTCTTCTTGATTAAAAGTTATGTAATAACCAAATGCAATGTGAAATATTT
TACTGGACTCTATTTTGAAAAACCATCTGTAAAAGACTGGGGTGGGGGTGGGAGGC
CAGCACGGTGGTGAGGCAGTTGAGAAAATTTGAATGTGGATTAGATTTTGAATGAT
ATTGGATAATTATTGGTAATTTTATGAGCTGTGAGAAGGGTGTTGTAGTTTATAAAA
GACTGTCTTAATTTGCATACTTAAGCATTTAGGAATGAAGTGTTAGAGTGTCTTAAA
ATGTTTCAAATGGTTTAACAAAATGTATGTGAGGCGTATGTGCCCGG**GCGGCCGC**

3XFLAG-SMN

CTCGAGATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTA
CAAAGACGACGATGACAAGACGCGTTCTAGAGCGATGAGCAGCGGCGGCAGTGGTG
GCGGCGTCCCGGAGCAGGAAGATTCCGTGCTGTTCCGGCGCGGCACAGGCCAGAGC
GATGATTCTGACATTTGGGATGATACAGCACTGATAAAAG**CATATG**ATAAAGCTGT
GGCTTCATTTAAGCATGCTCTAAAGAATGGTGACATTTGTGAAACTTCGGGTAAACC
AAAAACCACACCTAAAAGAAAACCTGCTAAGAAGAATAAAAGCCAAAAGAAGAAT
ACTGCAGCTTCCTTACAACAGTGGAAAGTTGGGGACAAATGTTCTGCCATTTGGTCA
GAAGACGGTTGCATTTACCCAGCTACCATTGCTTCAATTGATTTTAAGAGAGAAACC
TGTGTTGTGGTTTACACTGGATATGGAAATAGAGAGGAGCAAAATCTGTCCGATCTA
CTTTCCCCAATCTGTGAAGTAGCTAATAATATAGAACAAAATGCTCAAGAGAATGA
AAATGAAAGCCAAGTTTCAACAGATGAAAGTGAGAACTCCAGGTCTCCTGGAAATA
AATCAGATAACATCAAGCCCAAATCTGCTCCATGGAACTCTTTTCTCCCTCCACCAC
CCCCCATGCCAGGGCCAAGACTGGGACCAGGAAAGCCAGGTCTAA**AATTC**AATGGC
CCACCACCGCCACCGCCACCACCACCACCCCACTTACTATCATGCTGGCTGCCTCCA
TTTCCTTCTGGACCACCAATAATTCCCCCACCACCTCC**CATATG**TCCAGATTCTCTTG
ATGATGCTGATGCTTTGGGAAGTATGTTAATTTCATGGTACATGAGTGGCTATCATA
CTGGCTATTATATGGGTTTCAGACAAAATCAAAAAGAAGGAAGGTGCTCACATTCCT
TAAATTAAGGA**GCGGCCGC**

FLAG-SMN1

CTCGAGATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTA
CAAAGACGACGATGACAAGACGCGTTCTAGAGCGATGAGCAGCGGCGGCAGTGGTG
GCGGCGTCCCGGAGCAGGAAGATTCCGTGCTGTTCCGGCGCGGCACAGGCCAGAGC
GATGATTCTGACATTTGGGATGATACAGCACTGATAAAAGCATATGATAAAGCTGTG
GCTTCATTTAAGCATGCTCTAAAGAATGGTGACATTTGTGAAACTTCGGGTAAACCA
AAAACCACACCTAAAAGAAAACCTGCTAAGAAGAATAAAAGCCAAAAGAAGAATA
CTGCAGCTTCCTTACAACAGTGGAAAGTTGGGGACAAATGTTCTGCCATTTGGTCAG
AAGACGGTTGCATTTACCCAGCTACCATTGCTTCAATTGATTTTAAGAGAGAAACCT
GTGTTGTGGTTTACACTGGATATGGAAATAGAGAGGAGCAAAATCTGTCCGATCTAC
TTTCCCCAATCTGTGAAGTAGCTAATAATATAGAACAAAATGCTCAAGAGAATGAA
AATGAAAGCCAAGTTTCAACAGATGAAAGTGAGAACTCCAGGTCTCCTGGAAATAA
ATCAGATAACATCAAGCCCAAATCTGCTCCATGGAACTCTTTTCTCCCTCCACCACC
CCCCATGCCAGGGCCAAGACTGGGACCAGGAAAGCCAGGTCTAAAATTCAATGGCC
CACCACCGCCACCGCCACCACCACCACCCCACTTACTATCATGCTGGCTGCCTCCAT
TTCCTTCTGGACCACCAATAATTCCCCCACCACCTCCCATATGTCCAGATTCTCTTGA
TGATGCTGATGCTTTGGGAAGTATGTTAATTTCATGGTACATGAGTGGCTATCATAC
TGGCTATTATATGGTAAGTAATCACTCAGCATCTTTTCCTGACAATTTTTTTGTAGTT
ATGTGACTTTGTTTTGTAAATTTATAAAATACTACTTGCTTCTCTCGGTACCCGGGTC
TAGACGCGTGTCTTGTGAAACAAAATGCTTTTTAACATCCATATAAAGCTATCTATA
TATAGCTATCTATGTCTATATAGCTATTTTTTTTAACTTCCTTTATTTTCCTTACAGGG
TTTCAGACAAAATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAGGAGTAA
GTCTGCCAGCATTATGAAAGTGAATCTTACTTTTGTAAAACTTTATGGTTTGTGGAA
AACAAATGTTTTTGAACATTTAAAAAGTTCAGATGTTAAAAAGTTGAAAGGTTAATG
TAAAACAATCAATATTAAAGAATTTTGATGCCAAAACTATTAGATAAAAGGTTAATC
TACATCCCTACTAGAATTCTCATACTTAACTGGTTGGTTATGTGGAAGAAACATACT
TTCACAATAAAGAGCTTTAGGATATGATGCCATTTTATATCACTAGTAGGCAGACCA
GCAGACTTTTTTTTATTGTGATATGGGATAACCTAGGCATACTGCACTGTACACTCTG
ACATATGAAGTGCTCTAGTCAAGTTTAACTGGTGTCCACAGAGGACATGGTTTAACT
GGAATTCGTCAAGCCTCTGGTTCTAATTTCTCATTTGCAGGAAATGCTGGCATAGAG
CAGCACTAAATGACACCACTAAAGAAACGATCAGACAGATCTGGAATGTGAAGCGT
TATAGAAGATAACTGGCCTCATTTCTTCAAAATATCAAGTGTTGGGAAAGAAAAAA
GGAAGTGGAATGGGTAACTCTTCTTGATTAAAAGTTATGTAATAACCAAATGCAATG
TGAAATATTTTACTGGACTCTATTTTGAAAAACCATCTGTAAAAGACTGGGGTGGGG
GTGGGAGGCCAGCACGGTGGTGAGGCAGTTGAGAAAATTTGAATGTGGATTAGATT
TTGAATGATATTGGATAATTATTGGTAATTTTATGAGCTGTGAGAAGGGTGTTGTAG
TTTATAAAAGACTGTCTTAATTTGCATACTTAAGCATTTAGGAATGAAGTGTTAGAG
TGTCTTAAAATGTTTCAAATGGTTTAACAAAATGTATGTGAGGCGTATGTGCCCGG**G**
CGGCCGC

FLAG-SMN^{G1C}

CTCGAGATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTA
CAAAGACGACGATGACAAGACGCGTTCTAGAGCGATGAGCAGCGGCGGCAGTGGTG
GCGGCGTCCCGGAGCAGGAAGATTCCGTGCTGTTCCGGCGCGGCACAGGCCAGAGC
GATGATTCTGACATTTGGGATGATACAGCACTGATAAAAGCATATGATAAAGCTGTG
GCTTCATTTAAGCATGCTCTAAAGAATGGTGACATTTGTGAAACTTCGGGTAAACCA
AAAACCACACCTAAAAGAAAACCTGCTAAGAAGAATAAAAGCCAAAAGAAGAATA
CTGCAGCTTCCTTACAACAGTGGAAAGTTGGGGACAAATGTTCTGCCATTTGGTCAG
AAGACGGTTGCATTTACCCAGCTACCATTGCTTCAATTGATTTTAAGAGAGAAACCT
GTGTTGTGGTTTACACTGGATATGGAAATAGAGAGGAGCAAAATCTGTCCGATCTAC
TTTCCCCAATCTGTGAAGTAGCTAATAATATAGAACAAAATGCTCAAGAGAATGAA
AATGAAAGCCAAGTTTCAACAGATGAAAGTGAGAACTCCAGGTCTCCTGGAAATAA
ATCAGATAACATCAAGCCCAAATCTGCTCCATGGAACTCTTTTCTCCCTCCACCACC
CCCCATGCCAGGGCCAAGACTGGGACCAGGAAAGCCAGGTCTAAAATTCAATGGCC
CACCACCGCCACCGCCACCACCACCCCACTTACTATCATGCTGGCTGCCTCCATT
TTCCTTCTGGACCACCAATAATTCCCCCACCACCTCCCATATGTCCAGATTCTCTTGA
TGATGCTGATGCTTTGGGAAGTATGTTAATTTCATGGTACATGAGTGGCTATCATAC
TGGCTATTATATGGTAAGTAATCACTCAGCATCTTTTCCTGACAATTTTTTTGTAGTT
ATGTGACTTTGTTTTGTAAATTTATAAAATACTACTTGCTTCTCTCGGTACCCGGGTC
TAGACGCGTGTCTTGTGAAACAAAATGCTTTTTAACATCCATATAAAGCTATCTATA
TATAGCTATCTATGTCTATATAGCTATTTTTTTTAACTTCCTTTATTTTCCTTACAGGG
TTTCAGACAAAATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAGGACTAA
GTCTGCCAGCATTATGAAAGTGAATCTTACTTTTGTAAAACTTTATGGTTTGTGGAA
AACAAATGTTTTTGAACATTTAAAAAGTTCAGATGTTAAAAAGTTGAAAGGTTAATG
TAAAACAATCAATATTAAAGAATTTTGATGCCAAAACTATTAGATAAAAGGTTAATC
TACATCCCTACTAGAATTCTCATACTTAACTGGTTGGTTATGTGGAAGAAACATACT
TTCACAATAAAGAGCTTTAGGATATGATGCCATTTTATATCACTAGTAGGCAGACCA
GCAGACTTTTTTTTATTGTGATATGGGATAACCTAGGCATACTGCACTGTACACTCTG
ACATATGAAGTGCTCTAGTCAAGTTTAACTGGTGTCCACAGAGGACATGGTTTAACT
GGAATTCGTCAAGCCTCTGGTTCTAATTTCTCATTTGCAGGAAATGCTGGCATAGAG
CAGCACTAAATGACACCACTAAAGAAACGATCAGACAGATCTGGAATGTGAAGCGT
TATAGAAGATAACTGGCCTCATTTCTTCAAAATATCAAGTGTTGGGAAAGAAAAAA
GGAAGTGGAATGGGTAACTCTTCTTGATTAAAAGTTATGTAATAACCAAATGCAATG
TGAAATATTTTACTGGACTCTATTTTGAAAAACCATCTGTAAAAGACTGGGGTGGGG
GTGGGAGGCCAGCACGGTGGTGAGGCAGTTGAGAAAATTTGAATGTGGATTAGATT
TTGAATGATATTGGATAATTATTGGTAATTTTATGAGCTGTGAGAAGGGTGTTGTAG
TTTATAAAAGACTGTCTTAATTTGCATACTTAAGCATTTAGGAATGAAGTGTTAGAG
TGTCTTAAAATGTTTCAAATGGTTTAACAAAATGTATGTGAGGCGTATGTGCCCGG**G**
CGGCCGC

FLAG-SMNI6Δ7

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAAGACGACGATG
ACAAGACGCGTTCTAGAGCGATGAGCAGCGGCGGCAGTGGTGGCGGCGTCCCGGAGCAGGAAGA
TTCCGTGCTGTTCCGGCGCGGCACAGGCCAGAGCGATGATTCTGACATTTGGGATGATACA
GCACTGATAAAAGCATATGATAAAGCTGTGGCTTCATTTAAGCATGCTCTAAAGAAT
GGTGACATTTGTGAAACTTCGGGTAAACCAAAAACCACACCTAAAAGAAAACCTGC
TAAGAAGAATAAAAGCCAAAAGAAGAATACTGCAGCTTCCTTACAACAGTGGAAAG
TTGGGGACAAATGTTCTGCCATTTGGTCAGAAGACGGTTGCATTTACCCAGCTACCA
TTGCTTCAATTGATTTTAAGAGAGAAACCTGTGTTGTGGTTTACACTGGATATGGAA
ATAGAGAGGAGCAAAATCTGTCCGATCTACTTTCCCCAATCTGTGAAGTAGCTAATA
ATATAGAACAAAATGCTCAAGAGAATGAAAATGAAAGCCAAGTTTCAACAGATGAA
AGTGAGAACTCCAGGTCTCCTGGAAATAAATCAGATAACATCAAGCCCAAATCTGC
TCCATGGAACTCTTTTCTCCCTCCACCACCCCCCATGCCAGGGCCAAGACTGGGACC
AGGAAAGCCAGGTCTAAAATTCAATGGCCCACCACCGCCACCGCCACCACCACCAC
CCCACTTACTATCATGCTGGCTGCCTCCATTTCCTTCTGGACCACCAATAATTCCCCC
ACCACCTCCCATATGTCCAGATTCTCTTGATGATGCTGATGCTTTGGGAAGTATGTTA
ATTTCATGGTACATGAGTGGCTATCATACTGGCTATTATATGGTAAGTAATCACTCA
GCATCTTTTCCTGACAATTTTTTTGTAGTTATGTGACTTTGTTTTGTAAATGACACCA
CTAAAGAAACGATCAGACAGATCTGGAATGTGAAGCGTTATAGAAGATAACTGGCC
TCATTTCTTCAAAATATCAAGTGTTGGGAAAGAAAAAAGGAAGTGGAATGGGTAAC
TCTTCTTGATTAAAAGTTATGTAATAACCAAATGCAATGTGAAATATTTTACTGGAC
TCTATTTTGAAAAACCATCTGTAAAAGACTGGGGTGGGGGTGGGAGGCCAGCACGG
TGGTGAGGCAGTTGAGAAAATTTGAATGTGGATTAGATTTTGAATGATATTGGATAA
TTATTGGTAATTTTATGAGCTGTGAGAAGGGTGTTGTAGTTTATAAAAGACTGTCTT
AATTTGCATACTTAAGCATTTAGGAATGAAGTGTTAGAGTGTCTTAAAATGTTTCAA
ATGGTTTAACAAAATGTATGTGAGGCGTATGTG

APPENDIX E
SEQUENCES OF U1 snRNA EXPRESSION VECTORS

Key for expression vectors:
U1 snRNA: Yellow
Mutations: Red text
XhoI site: **CTCGAG**
Bg1II site: **AGATCT**

wU1

AGATCTCATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

Mutant U1

AGATCTCAGACTTACTCCGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1[11]

AGATCTCAGACTTAGTCCGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1[10A]

AGATCTCAGACTTAGTCGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{10B}

AGATCTCAGACTTAGTTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{1+8C}

AGATCTCAGACTTAGCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{2+6}

AGATCTCATACTTAGCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{1+8D}

AGATCTCATACTTAGTTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{1+8A}

AGATCTCATACTTAGTCGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{1+8B}

AGATCTCAGACTTAGCCGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{2+7}

AGATCTCAGACTTACTCGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{V1}

AGATCTCTGGCAGACTTAGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{V2}

AGATCTCAGATTCACTTTGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{V3}

AGATCTCAGTTTTACAAAGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**

eU1^{V4}

AGATCTCCACAAACCATGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCA
GGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCTGCGATTTCCCCAAATG
TGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTTTCCCCT
GACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTT
GGTTTGTGTCTTGGTTGGCGTCTTAAATGTTAATCCTACAGTGGAGG**CTCGAG**