UNIVERSITY OF CALIFORNIA SANTA CRUZ

INVESTIGATING THE ROLE OF SNRP-27 AND SNU66 IN 5' SPLICE MANTINANCE

A dissertation submitted in partial satisfaction of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY AND BIOCHEMISTRY

by

Kenna K. Sarka

March 2023

The Dissertation of Kenna K. Sarka is approved:

_____
Alan Zahler, Ph.D., chair

_____
Melissa Jurica, Ph.D.

_____
Seth Rubin Ph.D.

_____
Peter Biehl

Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

ABSTRACT: INVESTIGATING THE ROLE OF SNRP-27 AND SNU66 IN 5' SPLICE MANTINANCE

Kenna Sarka

Structural data suggest that human SNRP27 M141 and SNU66 H734 have a stacking interaction at the base of the U6 snRNA ACAGAGA box in the pre-B complex. Previously, we found that mutations in C. *elegans* at SNRP-27 M141 promote changes in cryptic 5' splice site choice and alternative 5'ss usage at native alternative splicing events [1]. We wanted to better understand if the potential interaction between SNRP-27 M141 and SNU-66 H765 (the *C. elegans* equivalent position to human SNU66 H734) contributes to maintaining the 5' splice site identity during spliceosome assembly. Mutations at SNU-66 H765 change alternative 5' splice sites usage genome wide. SNU-66(H765G) is a suppressor of *unc-73(e936)* 5' splice site cryptic splicing, similar to SNRP-27 (M141T). Many of the alternative 5' splicing events affected by SNU-66 (H765G) and (H765L) overlapped with those affected SNRP-

27 (M141T). Double mutants of *snrp-27(M141T)* with *snu-66(H765G)* are homozygous lethal. Our findings show a new role for SNU-66 in 5' splice site selection in a similar way as SNRP-27 M141T. Mutations at SNRP-27 M141 and SNU-66 H765 allow the spliceosome to choose alternative 5' splice sites with weaker consensus in the proximity of a stronger consensus 5'ss. While we observed alternative 5' splicing events that could be promoted either upstream or downstream of the stronger 5'ss in these mutants, the mechanism by which SNRP-27 M141 and SNU-66 H765 mutants affect 5' splice sites is dependent on both the presence of a weaker consensus 5'ss in close proximity and other factors (perhaps splicing factor binding sites nearby) that are important for determining the directionality of alternative splicing.

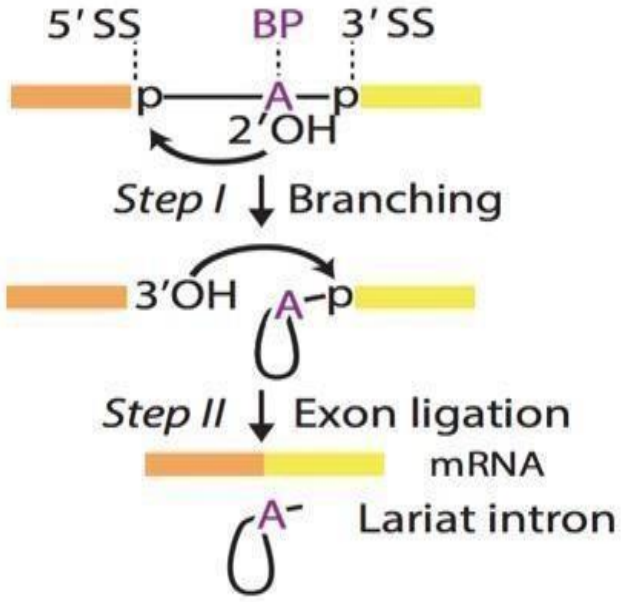
Acknowledgments

To my incredibly supportive mom, my inspiring dad, and my talented and loving partner Martin Kinisu.

Chapter One

The vast majority of eukaryotic protein-coding genes contain sections of non-coding sequences named introns, and the entire gene including the introns are transcribed into precursor mRNA (pre-mRNA), which must be processed before being translated into protein. Splicing is the process of removing introns from pre-mRNA and ligating together the coding exons. Sounds simple, but it is a highly sophisticated and complicated process with many unanswered questions still to explore. Understanding these mechanisms are significant, since splicing is essential for mRNA to code functional proteins. It is understood that 35% of all diseasecausing mutations disrupt splicing, and splicing factors expressed incorrectly have oncogenic properties and are involved in diseases[2]. Splicing also allows a single gene to make multiple protein isoforms through alternative splicing; this is when exons are included, skipped, or an alternative exon or splice site is chosen. This allows a wider number of proteins to be expressed than the number of genes in the genome, contributing to the complexity and evolution of higher organisms[3].

Figure 1



RNA Splicing by the Spliceosome
(Wilkinson et. al, 2020)

Figure 1: Splicing Reaction

(A)Structure of human precatalytic spliceosome showing the quasi pseudo knot and U4/U6helix at the base of the ACAGAGA box (red dotted line), and SNU66 histidine 734 next to SNRP27 methionine 141 and PRP8 histidine 1580. Coordinates are from structure 6QX9 in the PDB.

(B)Close up of the proposed binding interactions of SNU66 histidine 734, SNRP27 methionine 141, and PRP8 histidine 1580. SNRP27 methionine 141 packs against PRP8 histidine 1580 and histidine 734 of SNU66 which itself stacks on the quasi-pseudoknot. This stacking is hypothesized to stabilize the SNRP27 C-terminal loop, which buttresses the ACGAGA loop as it projects off U4/U6 stem III.

Splicing consists of two transesterification reactions to remove the introns and ligate the exons (Figure 1). In the first step, named branching, the 2' hydroxyl of the branchpoint (BP) adenosine attacks the phosphodiester at the 5' splice site (SS). This leaves a cleaved 5' exon and a lariat-intron-3' exon as intermediates. For the second step, named exon ligation, the 3' hydroxyl from the cleaved exon attacks the 3'SS phosphodiester on the lariatintron-3' exon. This leaves the final products of splicing - the ligated exons, and a lariat intron.

Introns contain key marker sequences for their processing; the 5' splice site (5'SS), the branch point (BP), and the 3' splice site (3'SS) (Figure 1). These sites are defined by short consensus sequences which can differ by species. In the model organism *Caenorhabditis elegans*, a simple roundworm that is the experimental model eukaryotic system used in this thesis, the consensus 5'ss sequence is found at the start of the intron, GUAAGU, with 99% of introns beginning with GU. The nucleotides surrounding the BP adenosine are not obviously conserved, and in fact only two branchpoints have been mapped in this organism[4]. The 3'SS consensus sequence at the end of the intron is a strong match to UUUCAG[5]. For the yeast *Saccharomyces cerevisiae*, a genetic system in which the splicing mechanism has been studied extensively, the 5'ss consensus sequence at the start of the intron is a strong match to

GUAUGU, while the 3'SS consensus at the end of the intron is YAG where Y is a pyrimidine[6]. The BP adenosine in yeast is 18-40 nucleotides upstream from the 3'ss with the highly conserved sequence, UACUA**A**C, with the bold A being the BP adenosine[6]). In humans, consensus 5'ss sequence is GURAGU (R = purine), the branchpoint is YNYUR**A**Y), and the 3'SS YAG sequence is preceded by a poly-pyrimidine tract sequence[7]. Humans having a higher number of splicing factors and more complex splicing than in *S. cerevisiae*, allowing the human consensus intron sequences to be less conserved without loss of splicing efficiency.

Figure 2

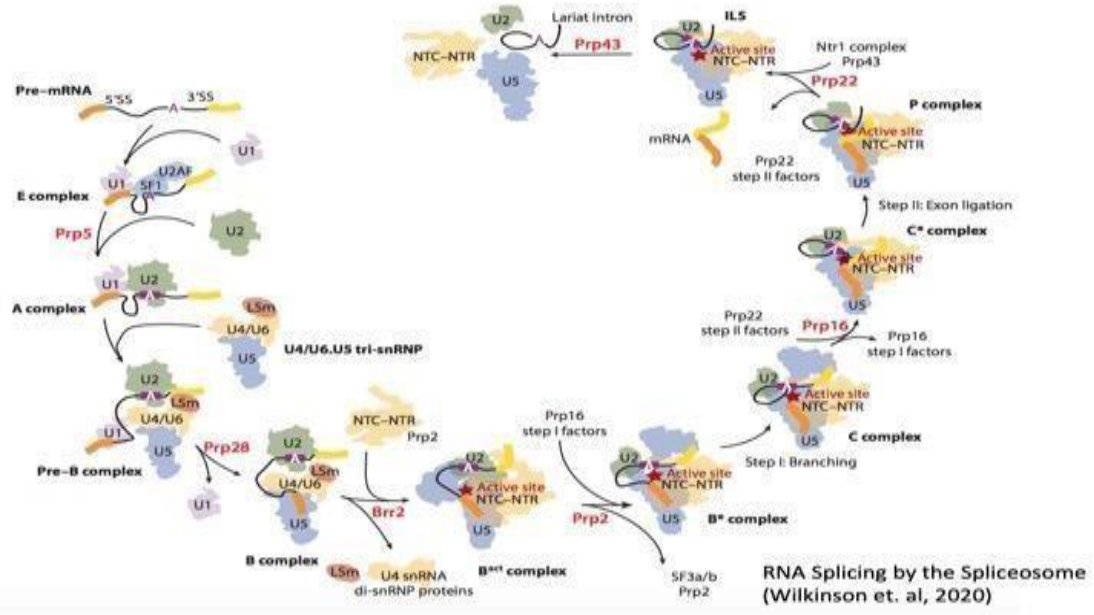


Figure 2: Splicing Cycle

A) CLUSTAL W Sequence alignment of SNU66 from S. cerevisae, C. elegans and humans. The bold H with the arrow is the histidine thought to stack with SNRP-27 M141.

B) Cy3 RT-PCR products showing alternative splicing for mab-10 and Y71H2AM.2. Black arrows indicate the 5'ss whose usage is promoted by the mutants, and the white arrows show the 5'ss whose usage is reduce by the mutants. The small black arrow for Y712AM.2 represents the alternative splice site shown by the doublet on the upper band. The sequence covering both 5' splice sites for mab-10 and Y71H2AM.2 are shown above the gels. The percent that each strain used the upstream 5'ss with standard deviation are shown below.

Splicing is performed by large RNA-protein assembly called the spliceosome. The spliceosome is not one single enzyme, instead the RNA and protein components are entering, interacting and leaving throughout the splicing cycle (Figure 2). The main components are

five types of small nuclear RNAs (snRNAs) which each bind a specific set of proteins to form a small nuclear ribonucleoprotein (snRNP). The 3' end of each U1, U2 U4, and U5 snRNA contain a U-rich sequence named the Sm site with seven homologous Sm proteins assembled around it in a ring. U6 snRNA also contains a Urich sequence at the 3' end, except the preassembled ring contains seven paralogous LSm proteins[7].

In first step of splicing, E complex formation, U1 snRNP recognizes the 5'SS through base paring with the 5'end of U1 snRNA, initiating the assembly of the spliceosome [8]. The U1 specific protein U1-C, stabilizes the pre-mRNA/U1 snRNA complex through interaction with a zinc finger domain, and by forming hydrogen bonds and electrostatic interactions with the RNA backbone and splice junction[9]. The 3' end of mammalian introns are marked by SF1, with the U2AF65-U2AF35 heterodimer bound to an adjacent sequence. U2AF65 interacts with SF1 to cooperatively facilitates branch point sequence selection[10]. Other U1 snRNP associated proteins can help the U1 snRNP bind to weaker 5'SS acting as alternative splicing factors, such as SR proteins which bind to enhancer sequences to direct U1 snRNP to alternative 5'SS[11]. U2AF65 also has a RS domain

which is used to bind the branch point sequence to facilitate annealing of U2 snRNA.

Next the prespliceosome enters A-complex. It is initiated when the DEAD-box helicases Prp5 and Sub2 replace SF1 and U2AF recruiting U2 snRNP to the BP [7]. The ATP-dependent binding of U2 to the BP sequence allows the 5' end of U2 snRNA to be freed [12] and connect with an unpaired sequence on the U6 snRNA 3' end forming U6/U2 snRNA stem II, and recruiting the pre-bound U4/U6.U5 trisnRNP to form the pre-B complex [13]. U4 snRNA is base paired with U6 snRNA in the tri-snRNP to maintain the catalytic core in an inactive state while the U1 snRNA is still bound to the 5'SS. Once the 5'SS and U1 interaction is dissociated, the 5'SS is transferred from U1 snRNA to U6 snRNA forming the catalytic center[14]. This marks the formation of B complex and initiation of spliceosome activation.

The largest preassembled splicing factor is the U4/U6.U5 tri-snRNP. Named "tri-snRNP" since it is comprised of the U4, U5, and U6 snRNA. It contains the heart of the spliceosome since the U6 snRNA will fold to form the active site. In fully assembled pre-B complex, the pre-spliceosome has recruited the tri-snRNP and affiliated proteins fully integrating the pre-mRNA into the spliceosome. U6 snRNA is initially base paired with U4 snRNA, which maintains an inactive

conformation. The ACAGAGA box of U6, which will bind with the first several nucleotides of the intron in B complex, is looped out in the tri-snRNP between a U4/U6 structure called the quasipseudoknot, and U4/U6 stem I and stem III[15]. When the 5'SS/U1 snRNA bond is disrupted by the DEAD-box helicase Prp28[16], the 5'SS base pairs with U6 snRNA, which folds to form the active site. The 5' exon is transferred to

U5 snRNA loop 1, and once the 5'SS transfer is completed the spliceosome is in B complex. From B complex, the spliceosome is ready to undergo largescale remodeling to prepare for catalysis.

U2 and U6 snRNAs form two helices named Ia and Ib, which are separated by a bulge and are adjacent to the U6 internal stem loop (ISL) causing the ISL to have a highly twisted backbone of

bulged

nucleotides [17,18]. Between the bulged U and the U6 ISL is a base triple which guides two phosphates from the ISL close to two phosphates from the U6 lb helix. From the four U6 phosphates, five nonbridging oxygens coordinate the two catalytic metal ions that activate the nucleophiles, stabilizing the leaving groups to allow the transesterification reactions of splicing to occur[19]. The U6 ACAGAGA upstream of U2 snRNA helix Ia positions the 5'SS at the catalytic metal ions. The final GA of the ACAGAGA box forms two

base pairs with helix Ib and this complex forms a catalytic triplex with the base triple docking the 5'SS into the active site[7]. Downstream of the U2 snRNA helix Ib, U2 binds the BP sequence forcing the BP adenosine to budge out of the branch helix. The 5' exon is held in place from before branching through exon ligation by U5 snRNA loop 1 which extends into the active site to bind

to it.

The whole catalytic RNA core is surrounded by the Prp8 large domain, Prp19 associated complex (NTC) proteins, and Prp19-related complex (NTR) protein, and its structure remains unchanged from B^{act} complex until ILS complex formation at the end of the splicing cycle. Although the catalytic core is formed and the 5'SS is loaded in the active site during B^{act}, the spliceosome is still in an inhibited state. The 2'OH group of the BP adenosine which is the nucleophile for the branching reaction cannot reach the active site since the branch helix is blocked by SF3b of U2 snRNP[20]. The 5'SS is also inhibited by SF3a protein Prp11 and B^{act} complex-specific factor Cwc24[21]. This inhibited conformation is further stabilized by Brr2 binding SF3b to retain it on the spliceosome[22]. These inhibiting interactions are disrupted by the DEAH-box ATPase Prp2, allowing for the branch helix to dock next to the

5’SS displacing Cwc24, and for the BP adenosine to dock in the active site.

Once the active site has all the components to catalyze the branching reaction the spliceosome is in B* complex. In this conformation the nucleophile 2’OH group of the BP adenosine attacks the 5’SS leaving a cleaved 5’ exon and a lariat-3’ exon intermediate. When the branching reaction is catalyzed the spliceosome is in C complex. The structure of the spliceosome remains unchanged from B* to C complex.

The same active site is used for the exon ligation step. Prp16, a DEAH-box ATPase binds downstream of the BP and then displaces the branching factors and rotates the branch helix to make space in the active site for the 3’SS to dock [23]. The branch helix is stabilized in the new position by

Prp17, a exon ligation factor pre-bound to the U6 snRNA 5’ stemloop. Prp17 braces the branch helix against the NTC and the RNaseH domain of Prp8 rotates to clasps the branch helix[7]. Once the branch helix is undocked it returns back to an A-form helix.

The spliceosome is in C* complex right before the exons are ligated.

The 3'SS will not dock into the active site until assisted by exonligation factors. Prp18 binds the Prp8 RNaseH domain and inserts a loop between the

Prp8 Large and RNaseH domains to insure that the 3'SS remains docked in the active site. Once the 3'SS docks into the core of C* complex, it is recognized through non-Watson-Crock base-paring with the 5'SS and BP adenosine. The 3'SS G^{-1} pairs with the G^{+1} of 5'SS, while the 3'SS A^{-2} pairs with the BP adenosine. The branched lariat is formed once the 5'SS and BP adenosine covalently bond during the branching reaction. When the exons are ligated, the spliceosome is in P complex.

The release of the ligated exon is catalyzed by Prp22 which is a DEAH-box ATPase, 3' 5' helicase. It remodels contacts between the mRNA and U5 snRNP[23]. In metazoans, the released mRNA contains the exonjunction complex (EJC) upstream of the exon junction which will perform a role in nonsense-mediated mRNA decay. The EJC is recruited to the 5' exon by Cwc22 and remains bound to Cwc22 until ligated-exon release.

Finally, at the end of the splicing cycle the spliceosome must be disassembled so the intron lariat can be decayed and the snRNAs and splicing factors can be recycled. Prp43 is DEAH-box ATPase that translocates along

U6 from the 3' end to disassemble the spliceosome and release low quality substrates or excised introns[24]. It breaks down the spliceosome into the intron-lariat, U6 snRNA, U2 snRNP core, U5 snRNP, and NTC proteins[25]. The exact structure Prp43 targets is still unknown, just another mystery of the spliceosome.

My focus on splicing is in the steps from A complex through pre-B complex to B complex. U1 snRNA binds the 5' ss until the U4/U6.U5 trisnRNP arrives, then the 5'ss is transferred to U6 snRNA and U1 snRNA dissociates from the spliceosome. How the 5'ss is maintained throughout this handoff from U1 to U6 needs further studies to fully understand how the spliceosome ultimately determines the 5' ss. In the Zahler lab we use a forward genetic screen which identifies suppressors of cryptic 5' ss usage to determine specific splicing factor residues contributing to 5'ss maintenance [26]. The screen uses an allele, *e936,* of the *C. elegans unc-73* gene*,* which encodes a factor important for axon guidance. This allele contains a G→U mutation at the invariable intronic first nucleotide G of the 5'ss of intron 16. This mutation at the +1G drives splice site choice to cryptic splice sites at the -1 and +23 positions. These cryptic splice sites both have the canonical GU sequence and are frameshift mutations which result in a loss of functional protein, and a dramatic

movement phenotype. We also observe splicing at the wildtype splice site, even though the intron now begins with UU. The *unc73(e936)* screen allows us to find extragenic suppressor alleles that change alternative 5'ss choice first by looking for suppression of the *unc-73* movement defect and observing increases in the relative usage of the wildtype splice site. We use a labeled reverse transcription-PCR assay (RT-PCR) of the *unc*-73 transcript to quantify the increase of inframe messages. These cryptic splicing events are not subject to nonsense mediated decay, so the relative amount of splice site usage by RT-PCR reflects the output of the spliceosome [1]. Analysis of splice products confirms that the phenotype is due to a suppressor mutation that acts through splicing. We then do genetic mapping and genomic DNA sequencing to identify potential mutants that cause the suppression. We use CRISPR technology to recreate the mutation *de novo*, and if this CRISPR-mimic allele suppresses the movement defect and changes cryptic splicing, this confirms the genetic identity of the transacting suppressor of splice site choice [27] [28,29].

Before I joined the Zahler lab the *unc73(e936)* screen identified the *C. elegans* homolog of the SR protein, SNRP-27, as a suppressor of cryptic splicing [26]. SNRP-27 has an N-terminal RS domain [30]. This domain containing alternating arginines and phosphoserines is found

at the C-terminus in the members of the SR protein family [31]. SR proteins simultaneously interact with RNA and other proteins through an RNA recognition motif (RRM) and an arginine and serine rich RS domain , and are known to play roles in determining and maintaining the 5' and 3' ss[32]. SNRP-27 contains a RS domain and a highly conserved C-terminal domain, but no RRM. The human homolog was first found as a component of the U4/U6-U5 tri-snRNP and can be phosphorylated by a spliceosome associated kinase[30]. When we further studied the conserved region of SNRP-27, alignment of the terminal 20 amino acids showed conservation in *C. elegans,* human, *Drosphila melanogaster,* and *Schizosaccharomyces pombe.* Note that SNRP-27 is not found in the *S. cerevisiae* yeast which is an important genetic model system. The genetic screen identified a dominant single point mutation, M141T, in the highly conserved C-terminal region of SNRP-27 as a suppressor of cryptic 5' ss usage. Using highthroughput mRNA sequencing to compare alternative splice site usage, it was determined this mutation changed the splicing of 26 native genes [1]. Using the CRISPR-cas9 system to randomize the amino acid at this position, we showed 8 amino acid substitutions were viable, three were homozygous lethal, and all viable substitutions at M141 lead to changes in alternative 5' splice site choice. Although we had shown SNRP-27 is essential in splicing, little else was known about its role

and interactions in splicing. I began to explore these open questions through experiments such as Yeast two hybrids, CoImmunoprecipitations, and randomizing amino acids adjacent to SNRP-27 M141 to study the effect on 5' splice sites.

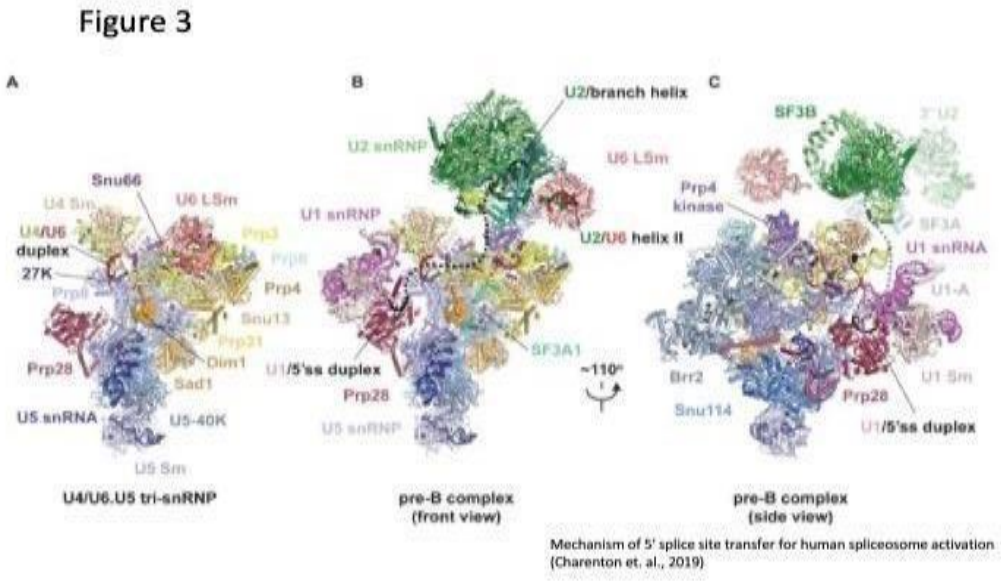


Figure 3 : Structure of tri-snRNP and pre-B complex

Structures of human tri-snRNP and pre-B complex. The dashed line represents the theoretical path of the pre-mRNA intron.

The project was able to pivot in a more precise direction in 2019 when the Nagai lab published the cryoEM structure of the human tri-snRNP and pre-B complex with the density of SNRP-27 conserved C-terminus mapped in

[15] (Figure 3). The conserved C-terminus of SNRP-27 is at the base of the U6 snRNA ACAGAGA box which binds with the 5'ss of the intron during Bcomplex. The ACAGAGA box is looped out between the U4/U6 stem III and the U4/U6 quasi-pseudoknot, which caps U4/U6 stem I and II. Stem III formation obscures the U4 snRNA nucleotides which are loaded in the active site of Brr2 to unwind the U4/U6 duplex for catalysis. The quasi-pseudoknot structure positions the ACAGAGA sequence to loop out and project towards the U1 snRNP to accept the 5'ss. SNRP-27 M141 packs against residues from two proteins, prp8 H1580 and snu66 H734. These three amino acids stack on one another, and in turn all stack on the quasipseudoknot. From this structural model, they suggested the packing may stabilize the SNRP-27 Cterminal loops which buttresses the

ACAGAGA loop.

Figure 4

Figure 4: Snu66 and SNRP-27 Sequence Alignments

Sequence alignments for S. cervisiae C. elegans, and H. sapiens for snu66 and SNRP- 27 . The bold H and M represent the amino acid that was mutated for snu66 and SNRP -27.

The structure gave new insights into where exactly in the splicing cycle SNRP-27 is present during pre-B complex, as well potential interactions with other protein residues such as snu66 H734. Snu66 was first identified as a gene encoding antigentic peptides of human squamous cell carcinomas recognized by human histocompatibility leukocyte antigens [33]. It was named SART1 (squamous cell carcinoma antigen recognized by T cells 1), although the name was later changed to mean spliceosome associated factor 1. It was first associated with splicing in yeast when identified by mass spectrometry in an analysis of U4/U6.U5 tri-snRNP proteins[34]. The human orthologue of snu66 was also shown to associate with the trisnRNP, and is required for its recruitment[35]. Aligning the *S. cerevisiae*, *C. elegan* and *H. sapien* snu66 sequences I found H734 and the surrounding region to be conserved, although not as conserved as the region surrounding M141 of SNRP-27 in humans and worms (Figure 4).

With minor previous knowledge on the specific role of SNRP27 and snu66 in splicing, I wanted to study their possible interaction and determine if snu66 H734 contributes to 5' splice site maintenance. With a similar strategy used to study SNRP-27 M141 [1]. I mutated snu66 H765 (*C. elegans* numbering) to random amino acids using CRISPR-cas9. I studied the effect of the mutant snu66 H765

alleles on 5' splice sites, and how those events overlapped with the results of SNRP-27 M141T to test if these residues contribute to splicing in similar ways, and a possible novel role for snu66 in 5' splice site selection.

Chapter Two

Abstract

Structural data suggest that human SNRP27 M141 and SNU66 H734 have a stacking interaction at the base of the U6 snRNA ACAGAGA box in the pre-B complex. Previously, we found that mutations in C. *elegans* at

SNRP-27 M141 promote changes in cryptic 5' splice site choice and alternative 5'ss usage at native alternative splicing events [1]. We wanted to better understand if the potential interaction between SNRP27 M141 and SNU-66 H765 (the *C. elegans* equivalent position to human SNU66 H734) contributes to maintaining the 5' splice site identity during spliceosome assembly. Mutations at SNU-66 H765 change alternative 5' splice sites usage genome wide. SNU66(H765G) is a suppressor of *unc-73(e936)* 5' splice site cryptic splicing, similar to SNRP-27 (M141T). Many of the alternative 5' splicing events

affected by SNU-66 (H765G) and (H765L) overlapped with those affected SNRP-27 (M141T). Double mutants of *snrp27(M141T)* with *snu-66(H765G)* are homozygous lethal. Our findings show a new role for

SNU-66 in 5' splice site selection in a similar way as SNRP-27 M141T. Mutations at SNRP-27 M141 and SNU-66 H765 allow the spliceosome to choose alternative 5' splice sites with weaker consensus in the proximity of a stronger consensus 5'ss. While we observed alternative 5' splicing events that could be promoted either upstream or downstream of the stronger 5'ss in these mutants, the mechanism by which SNRP-27 M141 and SNU-66 H765 mutants affect 5' splice sites is dependent on both the presence of a weaker consensus 5'ss in close proximity and other factors (perhaps splicing factor binding sites nearby) that are important for determining the directionality of alternative splicing.

Introduction

Early in spliceosome assembly, the U1 snRNP recognizes the 5' splice site through base pairing with the U1 snRNA [8] and the branch point is recognized by base pairing with the U2 snRNA . This is followed by recruitment of the U4/U6.U5 tri-snRNP to form the

pre-B complex . U4 snRNA base pairs with U6 snRNA to maintain its inactive state; when this interaction is dissociated, the 5' splice site is transferred from U1 snRNA to the ACAGAGA box of U6 snRNA, and the 5' exon to U5 snRNA loop 1[15]. U6 snRNA refolds and pairs with U2 snRNA forming the catalytic center, while U4 and U1 are removed from the spliceosome. This marks the formation of B complex and the initiation of spliceosome activation. This process of transferring the 5' splice site from U1snRNA to U6snRNA is significant, as U1snRNA, which initially identifies the 5' splice site in E complex, is not present in the spliceosome when catalysis occurs. Determining and maintaining the correct 5' splice site is essential for mRNA to code for functional proteins[36]. However, the mechanism of how 5' splice site fidelity is retained during the hand-off from the U1 snRNA to the catalytic core, and the splicing factors involved in this process, are still not well understood.

Previously, we demonstrated a role for the *C. elegans* homolog of the tri-SNRP-specific SNRNP27K protein in 5' splice site maintenance during spliceosome assembly. Using a genetic screen designed to identify suppressors of cryptic 5' splice site usage, we identified a dominant single point mutation, M141T, in the highly conserved C-terminal region of SNRP27 as a potent suppressor [26].

Using the CRISPR-cas9 system to randomize the amino acid at position 141, we showed 8 amino acid substitutions were viable, three were homozygous lethal, and all viable substitutions at M141 lead to changes in alternative 5'splice site usage. Then, high-throughput mRNA sequencing was used to compare alternative splice site usage, it was determined this mutation specifically changed alternative

5'splice site choice of 26 native genes. Further experiments showed SNRP-27 is an essential protein [1]. However, other than its association with the tri-snRNP, we had no information of the potential interactions of SNRP-27 in the splicing machinery.

Figure 1

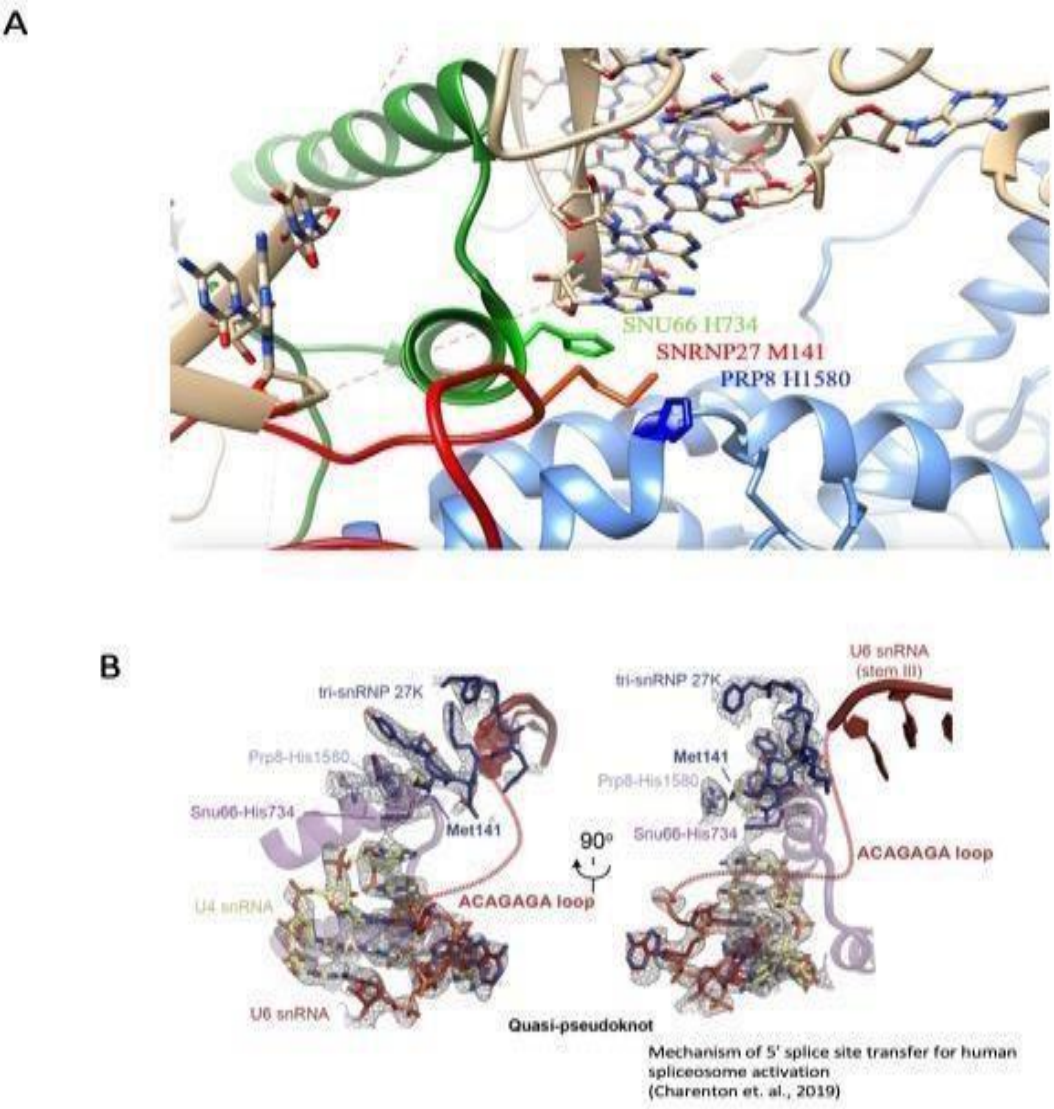


Figure 1: Structure of pre-B complex including snu66, SNRP-27 and PRP8 stacking

(A)Structure of human precatalytic spliceosome showing the quasi pseudo knot and U4/U6 helix at the base of the ACAGAGA box (red dotted line), and SNU66 histidine 734 next to SNRP27 methionine 141 and PRP8 histidine 1580. Coordinates are from structure 6QX9 in the PDB.

(B)Close up of the proposed binding interactions of SNU66 histidine 734, SNRP27 methionine 141, and PRP8 histidine 1580. SNRP27 methionine 141 packs against PRP8 histidine 1580 and histidine 734 of SNU66 which itself stacks on the quasi-pseudoknot. This stacking is hypothesized to stabilize the SNRP27 C-terminal loop, which buttresses the ACGAGA loop as it projects off U4/U6 stem III.

In 2019, the Nagai lab published cryo-EM structural models of the human tri-snRNP complex, and the human pre-B complex prior to U1 snRNP dissociation[15]. These were the first to include modeling of SNRNP27K into a spliceosomal structure. They were able to model in the very C-terminal amino acids of the protein, including the M141 position in which we found the semi-dominant mutation that affected cryptic 5' splice site usage. The structure modeled potential protein interactions for M141 of SNRP-27 with H734 of SNU66 and H1580 of PRP8. In this cryo-EM structure, the U4 snRNA nucleotides 63-67 form a structure referred to as a quasi-pseudoknot. The formation of the quasi-pseudoknot anchors one end of the U6 ACAGAGA sequence loop (Figure 1A). SNRNP27K M141 is sandwiched between H1580 of PRP8 and H734 of SNU66, which in turn stacks on the quasi-pseudoknot (Figure 1A,B). The residues around SNRP-27 M141 appear to be part of a core that ties together multiple parts of the complex and stabilizes the U4/U6 quasipseudoknot. The structure suggests a role in SNRNP27 and SNU66 helping the U6 snRNA ACAGAGA loop stay in the correct position before the ACAGAGA loop accepts the 5'splice site from the U1 snRNA.

SNU66 is also known as SART1 and was identified in human cells as an antigen recognized by cytotoxic T lymphocytes and named SART1 (squamous cell carcinoma antigen recognized by T cells 1)[33]. The *SART1* gene encodes two proteins – one 125kD and one 43kD. Both of the proteins are involved with regulation of proliferation and have tumor-rejection antigens, but no published role in splicing. The yeast homolog snu66 (SNUrp associated) was identified as a 66kD protein isolated from U4/U6.U5 trisnRNP[37]. During B complex, two prominent surface loops of snu66 act as a scaffold through stabilizing the Switch loop and β-finger, which are two catalytic motifs of Prp8[38]. Nothing else was known about snu66's role in splicing before its density was mapped in the pre-B complex[15].

In this study, we set out to test the model of SNU66SNRNP27K interaction and a potential novel role of the C-terminus of SNU66 in 5' splice site maintenance during spliceosome assembly. Using CRISPR mutagenesis in worms to alter SNU66 at the histidine residue modeled to interact with SNRNP27K M141, we demonstrate that altering H765 in *C. elegans* changed global alternative 5' splice site choice. The *snu-66(H765G)* mutation is a suppressor of *unc73(e936)* locomotion and splicing defects, similar to the *snrp27(M141T)* mutation that we found in the forward genetic suppressor screen. High throughput mRNA-Seq analysis of SNRP-27

and SNU66 mutant strains indicate that the *snu-66* allele primarily leads to changes in alternative 5' splice site usage genome-wide with the majority overlapping with the SNRP-27 alternative 5' splice sites events. We show the *snu-66* alleles promote alternative 5' splice site usage, and that the sites chosen appear to rely on additional clues in the pre-mRNA in addition to the sequence content and direction of the alternative splice site. Overall, we have demonstrated the importance of maintaining the SNU-66:SNRP-27 interaction in promoting 5' splice site fidelity during the handoff from U1 to U6 snRNA.

Figure 2

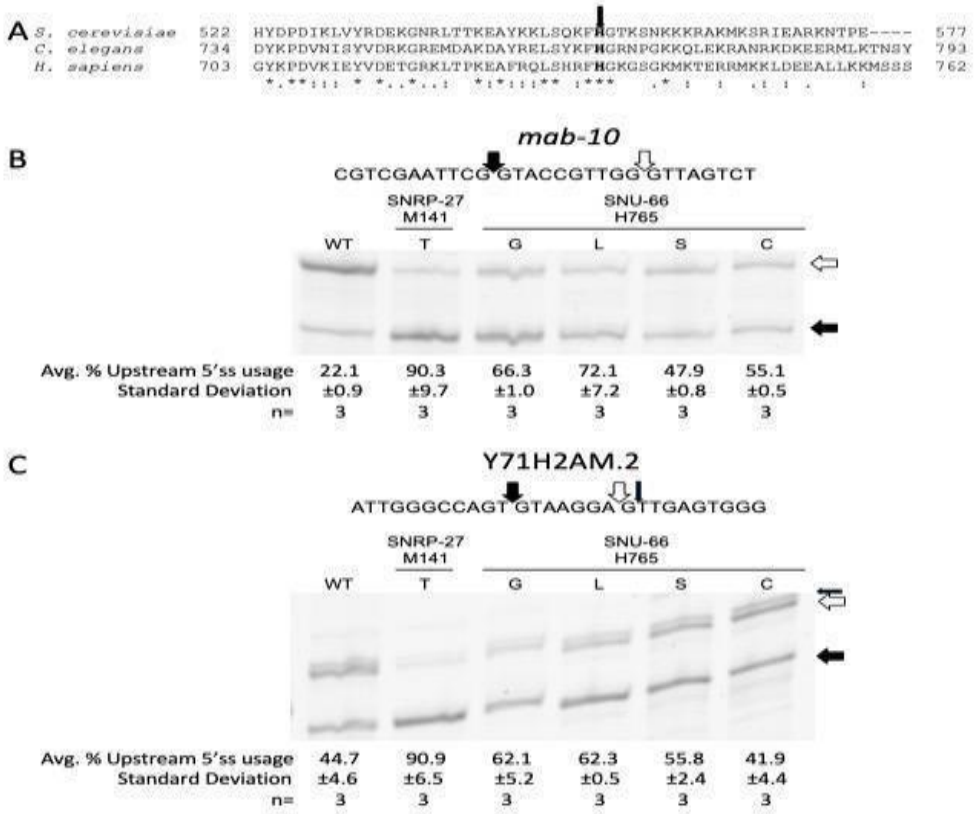


Figure 2: mab-10 and Y71H2AM.2 alternative splicing

A) CLUSTAL W Sequence alignment of SNU66 from S. cerevisae, C. elegans and humans. The bold H with the arrow is the histidine thought to stack with SNRP-27 M141. B) Cy3 RT-PCR products showing alternative splicing for mab-10 and Y71H2AM.2. Black arrows indicate the 5'ss whose usage is promoted by the mutants, and the white arrows show the 5'ss whose usage is reduce by the mutants. The small black arrow for Y712AM.2 represents the alternative splice site shown by the doublet on the upper band. The sequence covering both 5' splice sites for mab-10 and Y71H2AM.2 are shown above the gels. The percent that each strain used the upstream 5'ss with standard deviation are shown below.

Generating new *snu-66* mutants:

We wanted to test the model that SNU66 H734 (H765 in *C elegans*) and PRP8 H1580 (H 1573 in *C. elegans*) directly interact with M141 of SNRP27. We hypothesize that interruptions to that

interaction would have a similar effect on 5' splice site choice as the SNRP-27 M141T mutation [39]. The two domains of SNU66 known to have a role during B complex formation are in the N-terminal and center of the protein, while H765 is on the C-terminal end of the protein and has no previously known function in splicing. The Cterminus of SNU66 is not well conserved, however the histidine interacting with SNRP-27 as well as the two amino acids adjacent are conserved from *S. cerevisiae* to *C. elegans* to humans (Figure 2A), a sign of an important role at this sequence. To test this, we used a CRISPR-cas9 system coupled with oligonucleotide-directed recombination repair-based screen to randomize *snu-66* at position H765 and *prp-8* at position H1573 and screened for viable mutants that could be recovered.

We initially recovered nine independent strains, with five different amino acids substitutions at *snu-66* position H765 – G, L, S, C, and T. These strains grow and look like wild-type *C. elegans*. To test whether these *snu-66* H765 mutations affect splicing, we chose to study two out of the 26 alternative 5' splicing events found to be affected in the SNRP-M141T mutant, *mab-10* and Y71H2AM.2 [1]. We performed reverse-transcription followed by PCR with cy-3-labeled primers for the five new *snu-66* mutant alleles on alternative 5' splicing events in the *mab-10* and Y71H2AM.2 genes, and ran the

PCR products on polyacrylamide gels to visualize products (Figure 2B,C). For *mab-10*, the wild-type strain showed strong usage of the downstream 5' splice site, while in *snrp-27(M141T)*, splicing strongly shifted to the upstream splice site. The five *snu-66* strains with mutated H765 all shifted splicing towards the upstream 5'ss, but none as strongly as the *snrp27(M141T)*. Note that we are not showing the H765T in either experiment in

Figure 2B, but it behaved similarly to H765S which is shown. For Y71H2AM.2, wild-type was split almost evenly between the upstream and downstream alternative 5' splice sites, with *snrp-27(M141T)* almost exclusively choosing the upstream splice site (Figure. 2C). None of the *snu66* mutants preferred the upstream splice site to the extent of *snrp-27(M141T)*, however *snu-66* H765G, H765L and H765S promote increased usage of the upstream splice site. Consistent with the hypothesis that SNU-66 H765 interacting with SNRP-27 M141 is required for proper 5' splice site choice, a single point mutation in the unstudied C-terminal of SNU-66 showed changes in two alternative 5' splicing events in a manner similar to SNRP-27 M141T.

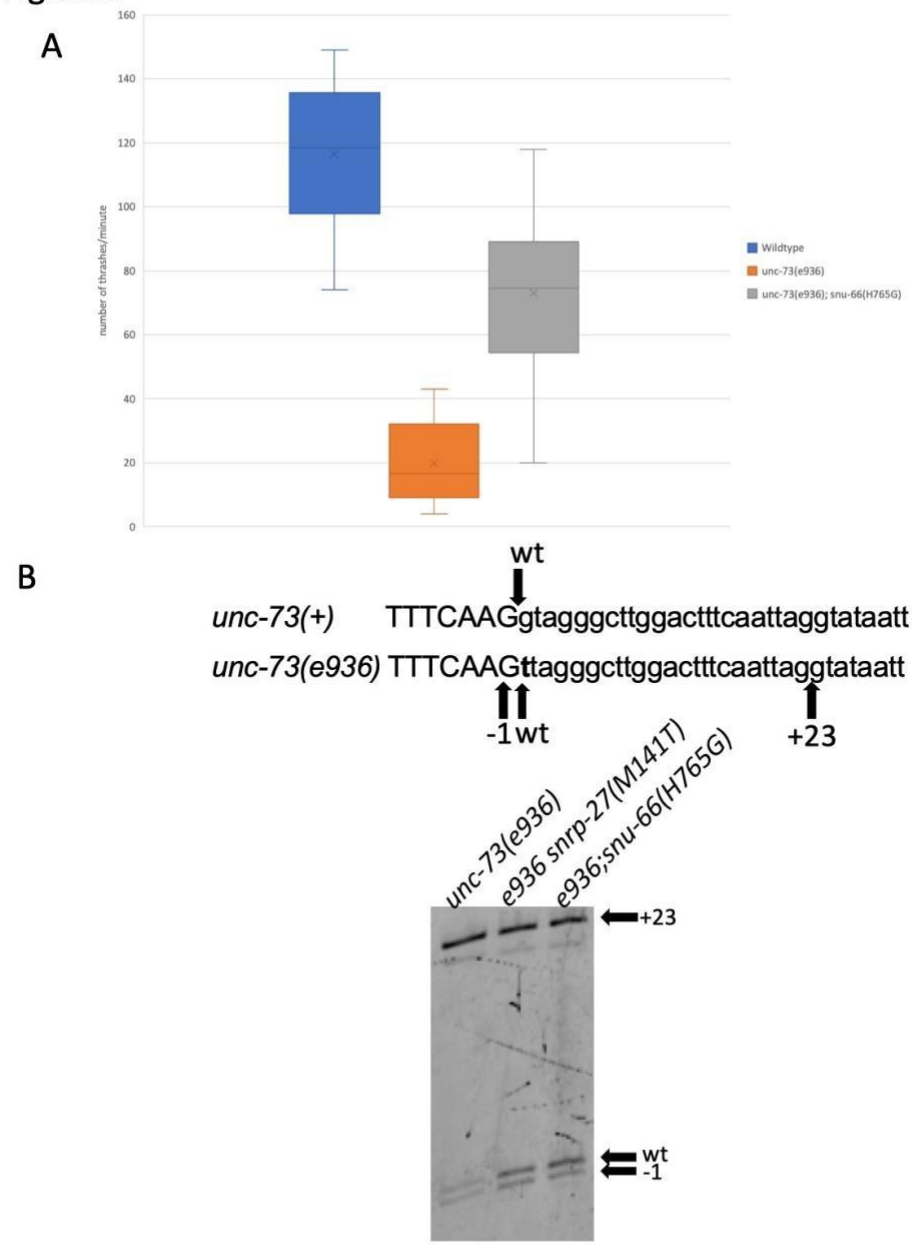
For PRP-8 H1573, we recovered five new alleles, three of which (C, F, S, N) we could maintain as heterozygotes over the wildtype allele, but could not be made homozygous viable. One new

PRP-88 H1573R allele was recovered and viable as a homozygous allele, however it did not show a change of splicing for the alternative 5' splicing events in *mab-10* and Y71H2AM.2 (data not shown). The homozygous lethal disruptions in PRP-8 H1573 may stem from PRP-8 having many roles throughout the splicing cycle, as opposed to SNRP-27 and SNU-66 which exit in the transition from pre-B to B complex. PRP-8 H1573 may play an essential role at a later time in the splicing cycle, so we cannot know if these disruptions are specific to the SNRP-27 interaction.

To study the combined genetic effects of the *snrp-27(M141T)* and *snu66 (H765G)* we attempted to create a *snrp-27(M141T)I ; snu66(H765G)V* double mutant strain. Worms were viable if homozygous for one allele and heterozygous for the other, however we determined that the double homozygous mutation was lethal. This is consistent with a hypothesis that either mutation weakens the SNRP-27:SNU-66 interaction, but that the double mutant eliminates the interaction. The sensitivity of the disruption of this interaction is consistent with our previous finding that SNRP-27 M141G, F, and E mutant alleles are homozygous lethal, along with our finding that *snrp-27* is an essential gene in worms[1].

snu-66(H765G) is a suppressor of cryptic 5' splice site usage

Figure 3



Genotype	-1	wt	+23	wt STDEV
unc-73(e936)I	16.0%	**19.6%**	64.4%	1.3% n=5
unc-73(e936) snrp-27(M141T)I	20.8%	**29.6%**	49.7%	0.7% n=4
unc-73(e936)I; snu-66(H765G)V	21.1%	**32.6%**	46.3%	2.4% n=6

Figure 3: Suppression of unc-73(e936)

(A) Box and whisker chart representing the thrash test data to demonstrate phenotypic suppression of unc73(e936) uncoordination. The X-axis represents the the average number of times 20 worms bent across their body axis in one minute.

(B) Sequences wild-type and e936 allele of the unc-73 gene. The arrows point to the-1, wild-type, and +23 splice sites.

(C) Crytpic splicing of the unc-73 gene using cy3 RT-PCR, showing the +23, wild-type,and-1 splice sites. The percent each splice site was used is shown below.

A screen for suppressors of *unc-73(e936)* locomotion defects identified *snrp-27(M141T)* as a suppressor that affects cryptic 5'ss choice[26]. We tested whether the new s*nu-66(H765G)* is also a suppressor of *unc-73(e936)* cryptic splicing. We performed genetic crosses to create a new double-mutant strain with *unc-73(e936); snu-66(H765G)* and tested it for changes to *unc-73(e936)* splicing. These worms looked and moved similar to wildtype. To quantify whether the *snu66(H765G)* suppressed the *unc73(e936)* locomotion defect we performed a thrash assay. *unc73(e936*) *C. elegans* are uncoordinated and struggle move in the typical sinusoidal way. Twenty *C. elegans* per strain were put individually in drops of M9 buffer on microscope slides, then they were observed for a minute and the number of times the *C. elegans* head crossed the X-axis was recorded (Figure 3A). The wild-type strains averaged 116 thrashes/min, while *unc73(e936*) strains averaged only 20 thrashes/min. The *unc-*

*73(e936);*snu66(H765G) double mutant showed increased movement averaging 73 thrashes/min, confirming the suppression of the *unc73(e936)* phenotype. Note that *snrp-27(M141T)* suppression of *unc73(e936)* leads to a reported thrash rate of 104 thrashes/min[26].

The *e936* allele of *unc-73* mutates the G at the beginning of intron 16 to a T, and this leads to activation of three cryptic splice sites, +23 into the intron, -1 upstream into the exon (the G➐T mutation leads to a new GT dinucleotide) and also splicing is detected at the wt splice site, even though that intron now begins with TT (Figure 3B). Both the +23 and -1 cryptic sites are out of frame, and we found that suppressors increase the usage of the wt splice site, leading to improved locomotion[26,27-29,40]. Even though the +23 and -1 sites are out of frame, transcripts that use that site are not subject to nonsense-mediated decay, indicating that relative splice site usage seen in this assay is consistent with the output of the spliceosome [41]. To confirm that *snu66(H765G)* increases splicing at the wt cryptic 5'ss, we performed reverse transcription followed by cy3 labeled PCR and ran the products on a denaturing polyacrylamide gel with single nucleotide resolution (Fig. 3B). In an almost identical way to *snrp-27(M141T)*, the *snu-66(H765G)* allele drove splicing away from the +23 and towards the -1 and wild-type splice site, providing a molecular explanation for the suppression of the

locomotion defect. Compared to *unc-73(e936)* worms which used the +23 64.4%, the *unc-73(e936); snrp-27(M141T)* and *unc73(e936);snu66(H765G)* double mutants decreased their +23 splice site usage to 49.7% and 46.3%, while the in-frame wildtype usage increased from 19.9% in *unc-73(e936)* to 29.6% and 32,6% in the *snrp-27(M141T)* and *snu-66(H765G)* suppressor strains respectively. These results indicate that *snu-66(H765G)* is indeed a locomotion defect suppressor of *unc-73(e936)* at the level of splicing, similar to *snrp-27(M141T)*. We have yet to saturate the suppressor screen [29] so it is not surprising that a specific point mutation is an undiscovered suppressor.

Alternative splicing analysis through high-throughput mRNAsequencing of SNRP-27 M141T, SNU-66 H765G and H765L strains

Figure 4

A

pairSum = 9 ΔPSI = 0.15	*snrp-27(M141T)* vs N2	*snu-66(H765G)* vs N2	*snu-66(H765L)* vs N2
Alt 5'ss	57	88	66
Alt 3'ss	0	2	1
Alt First Exon	2	9	6
Alt Last Exon	0	0	1
Retained Intron	2	35	16
Skipped Exon	0	2	2
Mutually Exclusive	0	0	2
Multiple Skip Exon	0	0	0

B

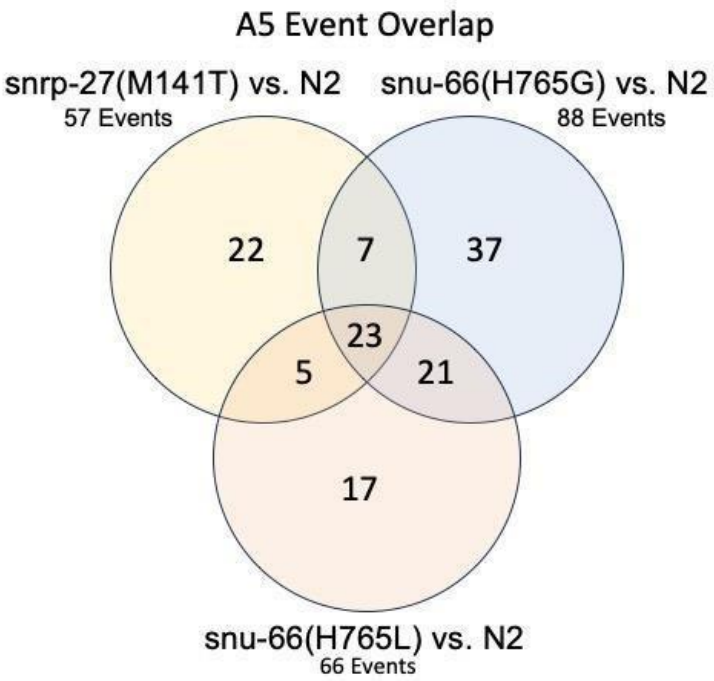


Figure 4: pre-mRNA High-throughput Sequencing

A) *The number of alternative splicing events of each class found for each strain relative to wild type. Each event must have a minimum of 15 junction spanning reads in each library, a ΔPSI > 0.15 for all nine pairwise comparisons, and a mean ΔPSI > 0.20 for the nine pairwise comparisons.*

B) *Venn diagram showing the overlapping and unique alternative 5'ss events found for all three strains.*

Next we wanted to determine if the *snu-66(H765G)* or *(H765L)* alleles lead to global changes in splicing on native genes and see how this compares to global splicing changes by *snrp-27(M141T)*. We compared highthroughput mRNA sequencing results of the three mutant strains with wildtype. Synchronized L3 animals were used in these experiments to control against splicing changes due to different developmental states that might be present in mixed populations [28]. RNA from three biologically independent RNA isolations from each strain were used to prepare libraries for high throughput sequencing. 150 x 150bp paired end reads from the libraries were mapped to the *C. elegans* genome (release ce10) using STAR[42]. Alternative 5'ss (A5) and alternative 3'ss (A3) events were identified *de novo* across all the libraries, and the percent spliced in (PSI) for each event in each library was calculated. We also examined Ensembl archived (ensArch65) retained intron, skipped exon, multiple skipped exon, mutually exclusive exon, alternative first exon and alternative last exon events and determined the PSI for each of these in each library. We then determined ΔPSI in pairwise comparisons for each event in a mutant

library vs. N2 (wild type) strain libraries. Nine pairwise comparisons (3 mutant libraries vs. 3 N2 libraries) were done for each event in each mutant strain. Alternative splicing events that showed ≥15% ΔPSI in all 9 pairwise comparisons (pairSum=9) were investigated further. Sequencing reads for individual pairSum=9 events were examined on the UCSC Genome Browser in order to disregard falsely predicted events.

The predominant change in alternative splicing that we found in the three mutant strains were in alternative 5' splice site usage. We found 57 alternative 5' splice site events for *snrp-27(M141T)*, 90 alternative 5' splice site events for *snu-66(H765G)*, and 66 alternative 5' splice site events for *snu66(H765L)* that met these stringent criteria (Figure 4A). We compared these
5' splice site events and found 23 were overlapping between all three mutant strain experiments (Figure 4B). This amount of overlap in 5' splice site events from point mutants in two different proteins is consistent with a connected role for these two residues in 5' splice site choice. The alternative
5' events from these tables are listed in supplemental table 1. It should be noted that this experiment increased the number of previously identified *snrp27(M141T)* native alternative splicing targets from 26 to

57. This can be explained as the current experiment has much greater depth of sequencing than the previous study[1].

Our mRNA seq analysis also found that the only other strong category of splicing event changes was in intron retention in the *snu66(H765G)* allele, 35 examples, and the snu-66(H765L) allele, 16 examples; snrp-27(M141T) only led to 2 examples of retained introns. (Figure 4A). Intron retention (IR) is when introns are retained in mature mRNAs instead of being spliced. It does play a role regulating gene expression, as these were events from the Ensembl alternative splicing archive, but it may also be an indication that splicing is inefficient in some way. Intron retention is associated with Alzheimer's disease and cancer[43]. These IR events could mean the H765G allele of *snu-66* may have an overall effect on splicing efficiency in addition to changing alternative 5'ss usage.

Figure 5

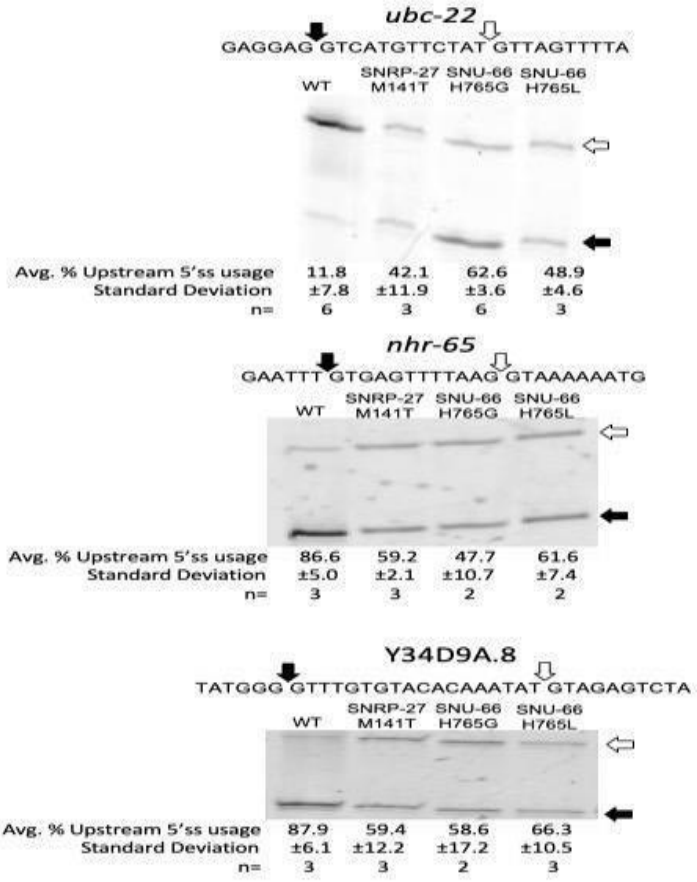


Figure 5: pre-mRNA Sequencing Validation

Confirmation of alternative 5'ss splicing predicted by mRNA sequencing using cy3 RT-PCR and separation on 6% polyacrylamide denaturing gels. Sequence of alternative splicing event and location of splice sites are indicated above the sample. Quantification of gel band intensity to determine relative alternative splice site usage is indicated below the gels.

We confirmed several of the new alternative 5' splicing events from the *snu-66* mutant sequencing data experimentally. The validated genes were chosen based off a high ΔPSI , and insuring they included cryptic 5' splice sites that were either upstream or downstream from the splice site used predominantly in the wildtype

strain. Three 5' splicing events were chosen, then reversetranscription followed by PCR with cy3 5' end-labeled primers was performed and the products separated by gel electrophoresis. These

RT-

PCR reactions confirm the alternative 5' splicing changes found through the high-throughput sequencing analysis (Figure 5).

Analyzing the alternative 5' splicing events identified by mRNA-Seq

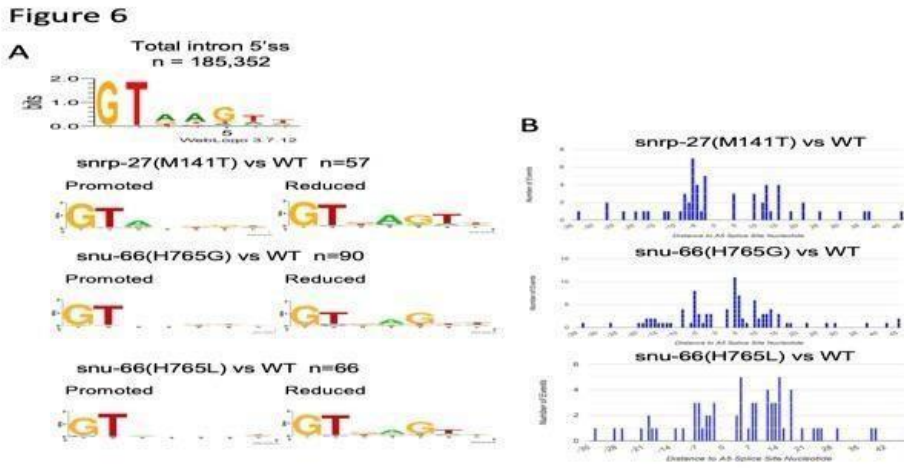


Figure 6: pre-mRNA Sequencing Analysis

A) Pictograms of the consensus sequence for the first 7 nucleotides of the intron for the alternative 5'splice sites of 185,352 wild-type C. elegans (top), and for the alternative 5' splice sites usage whose was promoted or reduced for the SNU-66 and SNRP-27 mutant srains. . https://weblogo.threeplusone.com/create.cgi

B) Histogram showing distance of alternative 5' splice sites events whose usage is increased in the mutant strains relative to the position of the alternative 5'ss whose usage is decreased.

The consensus sequence for the first seven nucleotides of the

5’ splice site intronic regions for *C. elegans* is GTAAGTT (Figure 6A)[5]. This sequence interacts with the U1 snRNA and U6 snRNA during the splicing cycle, with U1 identifying the 5’ splice site by base pairing initially in the E complex, and it is handed off to the U6 snRNA in the B complex. We analyzed the sequence context of the 5’ splice sites for the alternative isoforms

(Figure 6A). In the pairs of alternative 5’ splice sites, we refer to the one whose usage is increased in the presence of the mutant allele relative to wild type as “promoted” in the mutant strain, and the other member of the pair whose usage is reduced in the mutant strain relative to wild type as “reduced” in the mutant strain. For the 5’ splice sites sequences promoted in the *snrp-27* and *snu-66* mutants, besides the GT in the +1 and +2 positions of the intron, the other positions have no consensus, with the exception of the A at the +3 position for *snrp-27(M141T)*, which was observed previously[1]. The 5’ splice sites promoted in all three mutant strains are less conserved relative to total introns and relative to the 5’ss whose usage is reduced (Figure 6A). This indicates that, outside of the GT nucleotide to start the intron, the mutations in SNRP-27 and SNU-66 that we tested allow the spliceosome to be less stringent in consensus when selecting an alternative 5’ splice site sequence.

For the 5' splice site sequences whose usage is reduced in these strains, for all three mutants the site whose usage is reduced tended to exclude A residues at the 3+ position, with the other positions in agreement with the consensus 5' splice site (AGT at positions +4, +5 and +6 respectively). Our previous studies also saw exclusion of A residues at the 3+ position of alternative 5' splice sites reduced in SNRP-27 M141T[1].

We next examined the distance in nucleotides from the alternative 5' splice site whose usage is reduced in the mutant strains to the alternative splice site promoted in the mutant strains. These are displayed as histograms in figure 6B. The site whose usage is reduced in the mutants is set at zero nucleotides. Alternative 5' splice sites promoted by the mutants are then recorded in the histogram based on distance; promoted splice sites upstream are to the left of 0 and promoted splice sites downstream are shown to the right of 0. The number of nucleotides from the wild-type splice site ranged from -34 to +47, with a roughly even distribution of upstream and downstream sites for all three mutants. This result is consistent with the idea that the promotion of alternative 5' splice site usage by the *snrp-27* and *snu66* alleles is not directional, but is dependent on the presence of a sub-optimal alternative 5' splice site within a reasonable distance, ≤33nt upstream or ≤45nt downstream.

Exploring the potential mechanism for the SNRP-27 M141T, SNU66 H765G and H765L mutants on 5' splice site choice

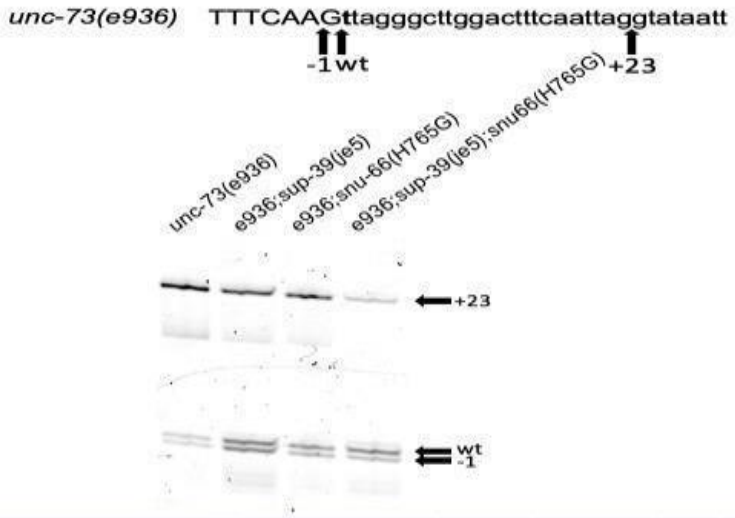
Based on a cryo-EM structure of the pre-B spliceosome, we hypothesized H765 of SNU-66 plays a role in SNRP-27 function. We had previously shown a role for SNRP-27 in maintenance of 5' splice site identity as the spliceosome assembles its active site. Studying the SNU-66 H765 mutants effect on global splicing and its similarity to the SNRP-27(M141T) effects, it is clear that SNU-66 H765 is also important for 5' splice site maintenance. We wanted to study this further, and determine a possible mechanism. SNRP-27 SNU-66 enter the spliceosome with the tri-snRNP, after U1 snRNA has bound to the 5' splice site. The cryo-EM structure shows interactions between SNRP27 M141 and SNU-66 H765 that stabilize the U4/U6 quasi-pseudo knot. One hypothesis to explain the alternative 5' splice site usage in the presence of these mutations is that these SNRP27:SNU-66 interactions are keeping the U6 ACAGAGA box sequestered from interacting with an alternative 5'ss while U1 snRNA is bound to the initial 5' splice site. Disrupting the stacking interactions through mutating these amino acids could allow the unstructured ACAGAGA box loop to prematurely interact with the pre-mRNA while the U1 snRNA is still bound to the 5' splice site.

This could allow the U6 ACAGAGA box to interact with an alternative 5' splice site while the correct one is occupied by U1

snRNA.

Figure 7

A



B

Strain	Genotype	-1	wt	+23	wt STDEV n=3
SZ181	*unc-73(e936)I*	16.0%	**19.6%**	64.4%	1.3% n=5
JW101	*unc-73(e936)I; sup-39(je5)II*	30.5%	**32.1%**	37.3%	0.2% n=2
SZ427	*unc-73(e936)I; snu-66(H765G)V*	21.1%	**32.6%**	46.3%	2.4% n=6
SZ440	*unc-73(e936)I; sup-39(je5)II; snu-66(H765G)V*	29.0%	**45.3%**	25.7%	3.9% n=3

C

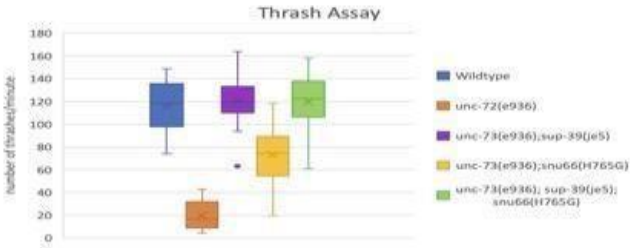


Figure 7: Effects of the unc-73(e936); sup-39(je5); snu66(H765G) triple mutant

(A) Above is the sequence of the unc -73 gene with the -1, wildtype and +23 splice sites shown by the black arrows. The bold T is the residue mutated from G to make the unc-73(e936) allele. Below shows the RT-PCR analysis of unc -73(e936) splicing for each strain.

(B) Quantification of RT-PCR samples for unc-73(e936) for each strain.

(C) Box and whisker chart representing the thrash test data to demonstrate phenotypic suppression of unc-73(e936) uncoordination. The X-axis represents the the average number of times 20 worms bent across their body axis in one minute.

Previously, we identified one *unc-73(e936)* suppressor, *sup39(je5)*, which is a U1 snRNA gene with a compensatory mutation that allows for U1 base pairing with introns beginning with UU[39]. It causes the spliceosome to prefer the *unc-73(e936)* wild-type cryptic site UU, leading to increases in splicing at the -1 and wild-type sites, and a decrease in +23 splice site usage (Figure 7). If *snu-66* mutant alleles are promoting alternative 5'ss usage by promoting a different 5'ss than is occupied by U1 snRNA, and if direction is not a factor, we hypothesized that by changing U1 snRNA 5'ss occupation with the mutant *sup-39* allele, we should see the *snu-66* mutant allele promote splicing in the opposite direction. In order to test this hypothesis we created a strain with *unc-73(e936)*, *sup-39(je5)* and *snu-66(H765G)*. We analyzed the usage of the *e936* cryptic splice sites in various strains carrying these alleles (Figure 7A,B). In the presence of *sup39(je5)* and *snu-66(H765G)*, we observe a dramatic additive effect on *unc-73(e936)* cryptic splicing away from usage of the +23 cryptic splice site, and stronger than either suppressor allele alone (Figure 7B). We also tested the *sup-39; snu-66* double mutant strain for its effect on unc-73(e936) locomotion defect suppression. The double

mutant suppresses unc-73(e936) to the same level as sup-39 alone, equivalent to wild type (Figure 7C). If our initial hypothesis was correct that *snu-66(H765G)* promotes splicing of the alternative 5'ss not occupied by U1snRNA, we would have expected *snu-66(H765G)* to promote splicing in the opposite direction of *sup-39(je5)* instead of the additive effect we see. This indicates that something in addition to a lack of U1 occupancy at the alternative 5' splice site determines the site promoted by the *snu-66(H765G)* allele.

Figure 8

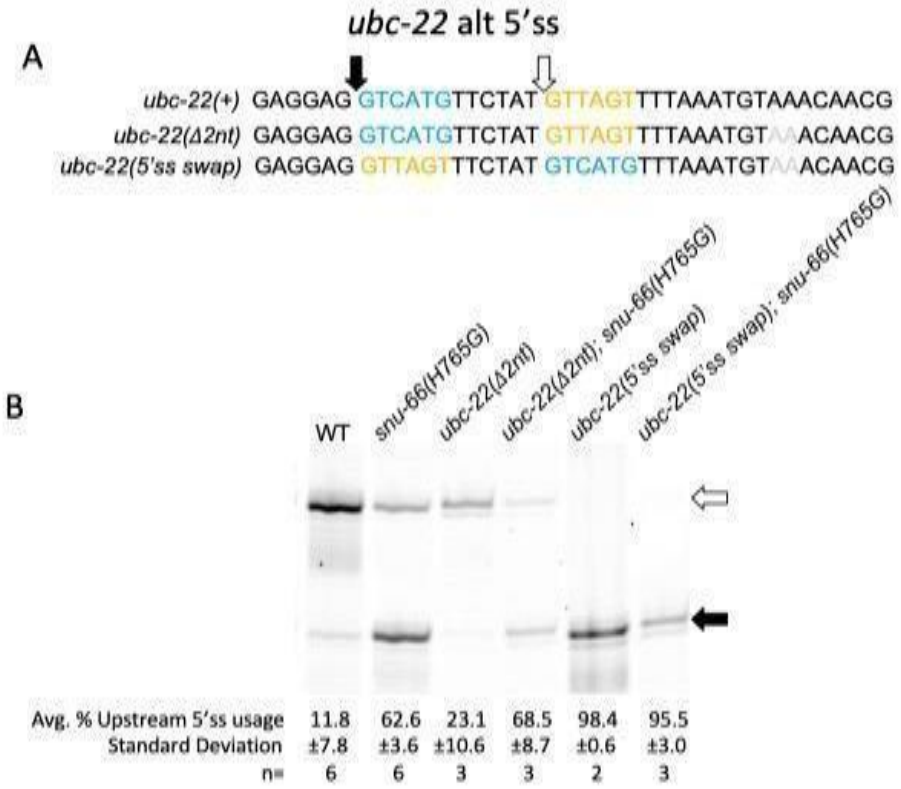


Figure 8: ubc-22 5'ss Swap

A) Top is the wildtype sequence for ubc-22. Next two sequences show ubc-22 with the two AA nucleotide deletion, and then ubc-22 with the 5'ss swapped and the

two AA nucleotide deletion. The blue sequence and black arrow represents the wildtype upstream splice site, while the yellow sequence and white arrow show the wildtype downstream sequence. The grey AA are the two nucleotides that were deleted.

B) Cy3 RT-PCR products for the alternatively spliced region of ubc-22. Quantitation is shown below the gel image.

We next asked whether the alternative 5' splice site promoted by the snu66 H765G mutants was based on the sequence of the 5'ss itself or the relative locations of the splice sites. In the gene *ubc-22*, there is an alternative 5' splicing event in which *snrp-27(M141T)* and *snu-66(H765G* and *H765L)* alleles promote an upstream 5' splice site (Figure 5). We designed a new allele of *ubc-22* in which the first six intronic bases of the pair of alternative splice sites, separated by 12nt, were swapped (Figure 8A).

We combined this new *ubc-22* allele (along with a control allele with a 2nt deletion 16nt downstream of the downstream 5'ss) with the *snu66(H7665G)* allele and determined the effect on splicing. As shown previously, the wild-type strains almost exclusively chose the downstream splice site in wildtype *ubc-22*, and the *snu-66(H765G)* mutation promoted the upstream splice site. This same effect held true in the 2nt intronic deletion control allele of *ubc-22*. In the *ubc22* 5'swap allele in a wildtype background, the upstream 5'ss was used almost exclusively, and in the presence of *snu-66(H765G)*, there was no switching to the weaker downstream 5'ss. If U1 snRNA still being

bound to the wild-type downstream splice site caused the shift toward the upstream splice site in the *snu-66(H765G)* background, then we might have predicted the opposite splice site usage in the *ubc-22* 5'ss swap event in the presence of *snu-66(H765G)*, where U1 is being directed to the stronger consensus upstream splice site. This would have been consistent with our observation that *snu-66(H765G)* promotes alternative 5'ss usage of native substrates in both directions, depending on the splicing event. Our results on both *unc-73(e936)* in the presence of *sup-39* (Figure 7) and the 5'ss swap of *ubc-22* (Figure 8) indicate that there must be a component outside of the 5'ss interaction sequence that also contributes to the location of the *snu-66(H7665G)*-promoted alternative splicing event.

Discussion

SNRP-27 was initially identified as a component of purified trisnRNPs that could undergo phosphorylation in its N-terminal RS domain[30]. Our genetic study identified the dominant mutation M141T that suppresses cryptic splicing [26] and changes 5' splice site use in *C. elegans* for a number of native alternative 5' splicing events[1]. Our results indicated an essential role of the highly conserved C-terminus of SNRP-27, but it was still a poorly understood splicing factor.

Subsequently, the cryoEM structure of the human tri-snRNP and pre-B complex gave a hint to possible SNRP-27-protien interactions, and to when in the splicing cycle it performs its role[1]. SNRP-27 M141 was modeled to interact with H734 of SNU66 and H1580 of PRP8 and to stabilize the U4 snRNA quasi-pseudoknot. The structure suggested the stacking interactions of the residues around SNRP-27 M141 could be stabilizing the U4/U6 quasi-pseudoknot and positioning the ACAGAGA loop correctly before accepting the 5'splice site from the U1 snRNA. SNU66 was only known in splicing as a component of U4/U6.U5 tri-snRNP, with a role during B complex formation in stabilizing PRP8. The potential of a SNU66 residue interacting with SNRP-27 M141 and quasi-pseudoknot indicated the previously unstudied SNU66 C-terminal region could have a role with SNRP27 in maintaining 5' splice site identity during the handoff from U1 snRNA to

U6 snRNA.

Based on modeling from the structural data of human pre-B complex

---

[1] we further studied SNU66 and determined H734 has a role in splicing. Analysis of the *C. elegans* SNU-66 (H765G/L) mRNA-seq showed alternative 5' splice sites promoted by the mutation are less conserved than the splice sites in the alternative pair whose usage are reduced (Figure 6A). This suggests that mutations at this position in SNU-66

disrupt the transfer of the optimal 5' splice site from U1 snRNA to U6 snRNA, and promotes usage of weaker nearby sites, similar to the SNRP-27(M141T) mutation. Mutating this SNU-66 position also suppresses the *unc-73(e936)* phenotype, meaning the spliceosome is more likely to splice at the non-canonical UU 5' splice site. Our data suggest that *snu-66(H765)* mutations allow the spliceosome to splice at positions with less stringent 5'ss consensus sequences.

The results also suggest SNU-66 and SNRP-27 work together and in additive ways to maintain the 5' splice site. We initially showed mutating *snu-66*(H765G/L) altered 5' splice site choice in *mab-10* and Y71H2AM.2, two events that saw large shifts in alternative 5'ss usage in *snrp-27(M141T)*. The alternative 5' splicing events found through mRNA sequencing for the *snu-66* and *snrp-27* mutants saw large overlaps of the events shared between them (Figure 4B). Although *snu-66(H765G)* and *snrp-27(M141T) C. elegans* strains are viable with no significant phenotypical changes, the double mutant is lethal. Perhaps each mutant destabilizes the protein-protein interaction between SNRP-27 and SNU-66 slightly, leading to similar alternative splicing changes, while the double mutant abolishes the interaction, leading to lethality consistent with certain alleles of *snrp27(M141)* mutations and the essential nature of *snrp-27* for viability[1].

We are beginning to study the mechanism of how *snu66(H765G/L)* and *snrp-27(M141T)* alter alternative splicing. The alternative 5' splice sites promoted in the mutant strains were both upstream and downstream of the 5' splice site whose usage is reduced, demonstrating the mutations do not have a specific directional effect on alternative 5'ss choice across all events. This was further confirmed when we swapped the sequence of the *ubc-22* 5' splice sites in the presence of *snu-66 (H765G)* and saw no change in splicing promoted by the mutant (Figure 8). This seemed to contradict the possibility that 5' splice site preference is due to activation of the alternative site while U1 was still bound to the stronger 5' splice site. This hypothesis was further explored using a U1 snRNA mutant that promoted a 5'ss starting with UU (Figure 7). In that case, the *unc73(e936);sup-39(je5);snu-66(H765G)* triple mutant saw the strongest change from the +23 splice site to -1 and wildtype in *unc-73(e936)* cryptic splicing, even though the *sup-39(je5)* allele led U1 snRNA to prefer 5' splice sites beginning with UU. Ultimately, our experiments suggest SNU-66 and SNRP-27 mutant ability alter 5' splice site choice depends on additional components of the pre-mRNA than just the presence of a 5'ss not occupied by U1snRNA in the region of the initially assembled 5' splice site (Figure 7-8).

When U6 snRNA binds the 5' splice site a Watson-Crick base pair is made between the 5' splice site +4 U and the central adenosine of the U6

AC**A**GAGA box (in bold). If this adenosine is methylated at the N6 position

(m^6A) by METTL16, 5' splice site preferences for the +4 position of the intron becomes more flexible. A recent comparative genomics study found

SNRP-27 may be necessary in genomes where U6 snRNA is m^6Amodified, because its presence is evolutionarily correlated to more variance of the 5' splice site +4 position when the U6 snRNA m^6A modification is absent (note that SNRP-27 is not found in the *S. cerevisiae* genome)[44]. In addition, with SNRP-27 M141 location close to the U6 snRNA m^6A, this has been interpreted to imply SNRP-27 stabilizes the U6/5' splice site helix by chaperoning the central A in the U6 AC**A**GAGA sequence for 5' splice site docking (Parker et al. 2023). This is consistent with our findings that SNRP-27 (M141T) and SNU-66 (H765G/L) cause a shift in 5' splice toward 5' splice sites with less stringency at the +4 position and away from the +4A (Figure 6A). Future experiments could further explore the effect of the intron sequence, particularly the +4 position, on how SNU-66 H765 and

SNRP-27 M141 effect 5' splice site identity maintenance during spliceosomal active site assembly.

SNRP-27 has an N-terminal region strikingly rich in alternating phosphoserines and arginines, known as an RS domain[30]. These phosphoserine/arginine domains are found at the C-termini of the SR protein family splicing factors [31] and these proteins have a significant role in recruitment of spliceosomal components throughout the splicing cycle, especially in the early stages of initial spliceosome recruitment of U1 snRNA to the pre-mRNA[45]. The U1-70K protein component of U1 snRNP also has an RS domain. These RS domains serve as a phospho-regulated interaction domains[46]. Given that SNRP27 has an RS domain, it is possible that it may be recruited to spliceosomes through SR protein interactions independent of its binding to *snu-66* in the tri-snRNP. This may explain the directionality independence of 5'ss sequence and U1 snRNA occupation that we see in the *ubc-22* 5'ss swap and *unc-73(e936)*;*sup-39* experiments. The *ubc-22* alternative splicing event contains the sequence GAGGAG directly upstream of the 5' splice site, and G-A-R sequences, where R is a pyrimidine, are known to bind to a subset of SR binding sequences[47]. It is possible this sequence recruits an SR protein and its presence at the upstream 5' splice sites in either version of the *ubc-22*

swap experiment could contribute to the upstream 5' splice site preference of both of the *snu-66* mutants. Perhaps the *snrp-27(M141T)* and *snu-66(H765G)* alleles allow for alternative 5'ss usage and U6 interaction with alternative splice sites when there are other elements of the pre-mRNA directing components of the spliceosome towards the alternative sites. Studying the role of the RS domain of SNRP-27, and the position of SR protein binding sites near 5' splice sites at events whose splicing changes in *snrp-27* M141 and *snu-66* H765 mutants could lead to more insights to the mechanism of the role of these proteins in 5' splice site identity maintenance during active site formation.

Table 1 – Strains used in this study

Strain Name	Genotype	Notes
JW101	*unc-73(e936)I; sup-39(je5)II*	U1 snRNA suppressor
SZ118	*unc-73(e936) snrp27(az26)*	snrp-27(M141T)
SZ181	*unc-73(e936)I*	
SZ211	*snrp-27(az56)I*	*snrp-27(M141T)*
SZ326	*snu-66(az138)V*	snu-66(H765S)
SZ370	*snu-66(az160)V*	*snu-66(H765G)*
SZ371	*snu-66(az161)V*	*snu-66(H765L)*
SZ372	*snu-66(az162)V*	*snu-66(H765C)*
SZ427	*unc-73(e936)I; snu-66(az160)V*	*snu-66(H765G) e936* suppressor strain
SZ444	*ubc-22(az187)X*	*ubc-22* alt 5'ss swap
SZ445	*unc-73(e936)I; sup-39(je5)II; snu-66(az160)V*	*sup-39; snu-66(H765G)* double suppressor
SZ447	*ubc-22(az188)X*	*ubc-22* 2nt intron deletion control
SZ448	*snu-66(az160)V; ubc-22(az187)X*	*ubc-22* alt 5'ss swap with *snu-66(H765G)*
SZ457	*snu-66(az160)V; ubc-22(az188)X*	*ubc-22* 2nt intron del. control with *snu-66(H765G)*

Thrashing assay

A thrash test was performed to measure how well the suppressed worms move in comparison to the uncoordinated mutant *unc-73(e936)* worms[48]. Live L4 worms were transferred to a drop of M9 solution on an NGM agar plate (usually seven individual drops of M9 per plate) and observed at 20°C. For a period of 60 s, worms were observed and the number of times that they bent across their body axis was recorded. Twenty L4 animals were assayed for each strain.

CRISPR/Cas9 genome editing

Cas9 guides were chosen from the CRISPR guide track on the UCSC Genome Browser *C. elegans* reference assembly (WS220/ce10) ([49]and crRNAs were synthesized by Integrated DNA Technologies (www.idtdna.com). Cas9 CRISPR RNA guides were assembled with a standard tracrRNA; these RNAs were heated to 95°C and incubated at room temperature to allow for annealing. The full guides were then incubated with Cas9 protein to allow for assembly of the CRISPR RNA complex[50]. That mix, along with a single stranded repair guide oligonucleotide was then microinjected into the syncytial gonad of

young adult hermaphrodite animals. A *dpy-10(cn64)* co-CRISPR strategy was used to identify F1 animals showing homology-directed CRISPR repair in their genomes[51]. Silent restriction sites were incorporated into or deleted from the repair oligo design so that mutations could be easily tracked by restriction digestion of PCR products from DNA extracted from single worms. Injected animals were moved to plates using recovery buffer, allowed to recover for 4 hours, and surviving worms were plated individually. F1 offspring were screened for the *dpy-10(cn64)* dominant roller (Rol) co-injection marker phenotype. F1 animals expressing the co-CRISPR marker were plated individually, allowed to lay eggs, and then the adult was removed and checked for allele of interest by PCR and restriction enzyme digestion followed by gel electrophoresis. If an F1 worm showed the presence of a heterozygous DNA fragment matching the programmed restriction site, non-roll animals in the F2 progeny were screened by electrophoresis of digested PCR products. Individuals that had lost the co-injection marker but were homozygous for the allele of interest were retained and sequenced at the gene of interest to verify error-free insertion of sequences guided by the repair oligo.

CRISPR crRNA guide RNA sequences

(entered as DNA into IDTdna.com crRNA order form).

Strain	Sequence
snu-66 H765 randomized	CAATCCGGGTAAGAAACAGT
Prp8 H1573 randomized	TCTTTTGCCACAAGTGAGCA
ubc-22 5'ss swap ΔAA	AGTTTTAAATGTAAACAACG
ubc-22 ΔAA	AGTTTTAAATGTAAACAACG

CRISPR repair oligonucleotides

Strain	Sequence
snu-66 H765 randomized	TGAGTGTGTCTAGTGAATTTTGAAATTTACAACTTT TTTTTTTCAGGTTCNNNGGTCGCAATCCGGGTAAGA AACAG**TTcGAa**AAACGAGCTAATCGTAAGGA
Prp8 H1573 randomized	ACGGAAAGATCCCGACGCTCAAGATTTCTCTCATTC AAAT**a**TT**CgG**TGCTNNNTTGTGGCAAAAGATTCACG AGTCAGTAGTTATGGATCTGTGTCAAGTTTT
ubc-22 5'ss swap ΔAA	TATTGAAGATTTATTTCGATCAATCGAGGTAGACGA GGAGGTTAGTTTCTATGTCATGTT TAAA**TGTACA**ACGTGGCATTTTCTTTCTAAAAACTT AAGCATTTGACTATTAAAACTG TT
ubc-22 ΔAA	TATTGAAGATTTATTTCGATCAATCGAGGTAGACGA GGAGGTCATGTTCTATGTTAGTTT TAAA**TGTACA**ACGTGGCATTTTCTTTCTAAAAACTT AAGCATTTGACTATTAAAACTG TT

*NNN codon is replaced with randomly incorporated nucleotides at these positions. Lower case nucleotides are silent mutations (differ from wildtype). Underlined are the nucleotides that are mutated. Bold sequences are the restriction enzyme site used.

Primers for Cy3 RT-PCR

Event	Forward Primer	Cy3 Reverse Primer
snrp-27(M141T)	GAGTCGTTACAAAGTGGAGC	TTCGCCATGGTCAAATTCCC
snu-66 H765 randomized	GTGGACCAGTTATGCCATTC	CCACTGAGCACAAGATACGG
Prp8 H1573 randomized	TCTTTTGCCACAAGTGAGCA	CATAATCTCCCCAACGAAGC
unc-73	TCAACCAGAAGCTGTTGGTG	TCCCTTAAAGTAGGCTCGTG
ubc-22	TTCTTCTGCCAGTCATCGTCT	AATGTTTCGGAGGCACTGTC
Y34D9A.8	CAGCGGTTATCGTCGTTGTC	AATGGCGGATTCGCTTCTCTC

nhr-65	TTTCGTGCTCCAGTGTGACC	TACGTGAACCATTTGGTGGC
ubc-22 5'ss swap ΔAA	CTCACGGTACGCTGTCATTT	GATGTTCAGTATGTGGGCTTCA

Primers for CRISPR mutagenesis genomic screening

Strain	Forward Primer	Reverse Primer
snrp- 27(M1 41T)	GAGTCGTTACAAAG TGGAGC	TTCGCCATGGTCAAA TTCCC
snu-66 H765	GTGGACCAGTTATGCCATTC	CCACTGAGCACAAG ATACGG
rand mize1		
Prp8 H15 3 rand mizel	CTCTGGTGGTCTCCAACTATC	CATAATCTCCCCAACGAAGC
ubc22 5'ss swap ΔAA	CTCACGGTACGCTGTCATTT	GATGTTCAGTATGTGGGCTTCA
ubc2		

ΔAA 2	CTCACGGTACGCTG TCATTT	GATGTTCAGTATGTG GGCTTCA

RNA extraction, cDNA production, and PCR amplification

RNA from indicated strains was extracted from mixed staged or L3 populations of worms using TRIzol (Ivitrogen), before phase separation with $CHCl_3$. Then RNA was reverse transcribed with specific primers for each gene using AMV reverse transcriptase (Promega). cDNA was PCR-amplified for 25 cycles with 5′-Cy3labeled reverse primers (IDT) and unlabeled forward primers using e Phusion high-fidelity polymerase (NEB). PCR products were separated on 40 cm tall 6% polyacrylamide urea denaturing gels and then visualized using a Molecular Dynamics Typhoon Scanner. Band intensity quantitation was performed using ImageJ software (https://imagej.nih.gov/ij/). For quantitation, a box of the same size was drawn around each alternative splicing product on a gel in ImageJ, and a control background box of the same size was drawn between them in each lane (or just above the two if the bands were too close together). The background volume value was subtracted from each band's value within a lane and then the relative usage of the splice sites was calculated.

RNASeq

Total RNA isolations from three biological replicates were done for strains SZ211 (*snrp-27(M141T)*), SZ370 (*snu-66(H765G)*), SZ371 (*snu66(H765L)*), and N2 (wildtype). mRNA isolation and sequencing library preparation for each RNA isolation were performed by RealSeq Biosciences (Santa Cruz, CA). 150nt x 150nt paired-end reads were obtained. RNA-seq results were trimmed, subjected to quality control, and two-pass aligned to UCSC Genome Browser *C. elegans* reference assembly (WS220/ce10) (this earlier assembly release was used to facilitate comparison to previous RNAseq datasets obtained by our lab) using STAR[42].

High stringency ΔPSI analysis

Alternative 5′ (A5) and alternative 3′ (A3) splicing events found in the STAR mappings of all of the libraries were identified and filtered for those introns with at least 5 reads of support (total across all samples) and a maximum of 50 nucleotides between the alternative ends (either 5′ or 3′ respectively). In addition, alternative first exon (AF), alternative last exon (AL), skipped exon (SE), retained intron (RI), mutually exclusive exon (MX) and multiple skipped exon (MS) events were derived from the Ensembl gene predictions Archive 65 of WS220/ce10 (EnsArch65) using junctionCounts
'infer pairwise events' function

(https://github.com/ajw2329/junctionCounts). The percent spliced in (PSI) for each event in each sample was derived using junctionCounts with the option suppress_eij_use for A3 and A5 events. Each strain had 3 biological replicates, therefore between any two strains, there are a total of nine possible pairwise comparisons for each event. For each suppressor strain, only alternative splicing events with a minimum of 15 junction counts that showed a change in the same direction >15% ΔPSI compared to the N2 wildtype control in all nine pairwise comparisons

(pairSum = 9) were considered further.Those events with a mean ΔPSI >20% across the nine comparisons were chosen for examination. The reads supporting that alternative splice site choice event were then examined individually on the UCSC Genome Browser *C. elegans* reference assembly (WS220/ce10) to ensure that the algorithmically flagged events looked like real examples of alternative splice site choice.

Consensus motifs

Consensus motifs were created using

WebLogo[52]; https://weblogo.berkeley.edu/logo.cgi.

Multiple sequence alignments

Multiple sequence alignments were generated using BLAST at the online web interface; https://blast.ncbi.nlm.nih.gov/Blast.cgi.

References

1. Zahler, A.M., Rogel, L.E., Glover, M.L., Yitiz, S., Ragle, J.M., and Katzman, S. (2018). SNRP-27, the C. elegans homolog of the trisnRNP 27K protein, has a role in 5' splice site positioning in the spliceosome. Rna *24*, 1314-1325. 10.1261/rna.066878.118.

2. Sterne-Weiler, T., and Sanford, J.R. (2014). Exon identity crisis: diseasecausing mutations that disrupt the splicing code. Genome Biol *15*, 201. 10.1186/gb4150.

3. Ule, J., and Blencowe, B.J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. Molecular Cell *76*, 329-345. 10.1016/j.molcel.2019.09.017.

4. Ragle, J.M., Katzman, S., Akers, T.F., Barberan-Soler, S., and Zahler, A.M. (2015). Coordinated tissue-specific regulation of adjacent alternative 3' splice sites in C. elegans. Genome Res *25*, 982-994. 10.1101/gr.186783.114.

5. Li, Y., Xu, Y., and Ma, Z. (2017). Comparative Analysis of the ExonIntron Structure in Eukaryotic Genomes. Yangtze Medicine *01*, 50-64. 10.4236/ym.2017.11006.

6. Spingola, M., Grate, L., Haussler, D., and Ares, M. (1999). Genomewide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. RNA *5*, 221-234. 10.1017/s1355838299981682.

7. Wilkinson, M.E., Charenton, C., and Nagai, K. (2020). RNA Splicing by the Spliceosome. Annual Review of Biochemistry *89*, 359-388. 10.1146/annurevbiochem-091719-064225.

8. Pomeranz, D.A., Oubridge, C., Adelaine Kwun-Wai, L., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. Nature *458*, 475-480. 10.1038/nature07851.

9. Kondo, Y., Oubridge, C., Roon, v., and Nagai, K. (2015). Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. eLife.

10. Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes Dev *12*, 858-867. 10.1101/gad.12.6.858.

11. Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Lührmann, R., GarciaBlanco, M.A., and Manley, J.L. (1994). Protein–protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature *368*, 119-124. 10.1038/368119a0.

12. Perriman, R., and Ares, M. (2010). Invariant U2 snRNA Nucleotides Form a Stem Loop to Recognize the Intron Early in Splicing. Molecular Cell *38*, 416427. 10.1016/j.molcel.2010.02.036.

13. Nguyen, T.H.D., Galej, W.P., Bai, X.-c., Oubridge, C., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. Nature *530*, 298-302. 10.1038/nature16940.

14. Staley, J.P., and Guthrie, C. (1999). An RNA Switch at the 5′ Splice Site Requires ATP and the DEAD Box Protein Prp28p. Molecular Cell *3*, 55-64. 10.1016/s1097-2765(00)80174-4.

15. Clément, C., Wilkinson, M.E., and Nagai, K. (2019). Mechanism of 5′ splice site transfer for human spliceosome activation. Science *364*,

362-367.

10.1126/science.aax3289.

16. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa. Cell *125*, 287-300. 10.1016/j.cell.2006.01.054.

17. Madhani, H.D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell *71*, 803-817. 10.1016/0092-8674(92)90556-r.

18. Fica, S.M., Tuttle, N., Novak, T.J., Li, N.-S., Lu, J., Prakash, K., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyses nuclear premRNA splicing. *503*, 229-234. 10.1038/nature12734.

19. Fica, S.M., and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine.
Nature Structural & Molecular Biology *24*, 791-799. 10.1038/nsmb.3463.

20. Rauhut, R., Fabrizio, P., Olexandr, D., Klaus, H., Pena, V., Chari, A., Kumar, V.A., Lee, C.-T., Urlaub, H., Kastner, B., et al. (2016). Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science *353*, 13991405. 10.1126/science.aag1906.

21. Wu, N.-Y., Chung, C.-S., and Cheng, S.-C. (2017). Role of Cwc24 in the First Catalytic Step of Splicing and Fidelity of 5′ Splice Site Selection. Molecular and Cellular Biology *37*. 10.1128/mcb.00580-16.

22. Bao, P., Will, C.L., Urlaub, H., Boon, K.-L., and Reinhard, L. (2017). The RES complex is required for efficient transformation of the precatalytic B spliceosome into an activated B^{act} complex. *31*, 2416-2429. 10.1101/gad.308163.117.

23. Schwer, B., and Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. The EMBO Journal *11*, 5033-5039. 10.1002/j.1460-2075.1992.tb05610.x.

24. Toroney, R., Nielsen, K.H., and Staley, J.P. (2019). Termination of premRNA splicing requires that the ATPase and RNA unwindase Prp43p acts on the catalytic snRNA U6. Genes & Development *33*, 1555-1574. 10.1101/gad.328294.119.

25. Tanaka, N., Aronova, A., and Schwer, B. (2007). Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. Genes Dev *21*, 2312-2325. 10.1101/gad.1580507.

26. Dassah, M., Patzek, S., Hunt, V.M., Medina, P.E., and Zahler, A.M. (2009). A Genetic Screen for Suppressors of a Mutated 5′ Splice Site Identifies Factors Associated With Later Steps of Spliceosome

Assembly. Genetics *182*, 725734. 10.1534/genetics.109.103473.

27. Mayerle, M., Yitiz, S., Soulette, C., Rogel, L.E., Ramirez, A., Ragle, J.M., Katzman, S., Guthrie, C., and Zahler, A.M. (2019). Prp8 impacts cryptic but not alternative splicing frequency. Proc Natl Acad Sci U S A *116*, 2193-2199. 10.1073/pnas.1819020116.

28. Suzuki, J., Osterhoudt, K., Cartwright-Acar, C.H., Gomez, D.R., Katzman, S., and Zahler, A.M. (2022). A genetic screen in C. elegans reveals roles for KIN17 and PRCC in maintaining 5' splice site identity. PLoS Genet *18*, e1010028. 10.1371/journal.pgen.1010028.

29. Cartwright-Acar, C.H., Osterhoudt, K., Suzuki, J., Gomez, D.R., Katzman, S., and Zahler, A.M. (2022). A forward genetic screen in C. elegans identifies conserved residues of spliceosomal proteins PRP8 and SNRNP200/BRR2 with a role in maintaining 5' splice site identity. Nucleic Acids Res *50*, 1183411857. 10.1093/nar/gkac991.

30. Fetzer, S., Lauber, J., Will, C.L., and Lührmann, R. (1997). The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated protein kinase. RNA *3*, 344-355.

31. Zahler, A.M., Lane, W.S., Stolk, J.A., and Roth, M.B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev *6*, 837-847.

10.1101/gad.6.5.837.

32. Jeong, S. (2017). SR Proteins: Binders, Regulators, and Connectors of RNA. Molecules and Cells *40*, 1-9. 10.14348/molcells.2017.2319.

33. Shigeki, S., Nakao, M., Imai, Y., Hideo, T., Kawamoto, M., Fumihiko, N., Yang, D., Toh, Y., Hideaki, Y., and Itoh, K. (1998). A Gene Encoding Antigenic Peptides of Human Squamous Cell Carcinoma Recognized by Cytotoxic T Lymphocytes. Journal of Experimental Medicine *187*, 277-288. 10.1084/jem.187.3.277.

34. Gottschalk, A. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6middle dotU5] tri-snRNP. The EMBO Journal *18*, 4535-4548. 10.1093/emboj/18.16.4535.

35. Makarova, O. (2001). The 65 and 110 kDa SR-related proteins of the U4/U6U5 tri-snRNP are essential for the assembly of mature spliceosomes. The EMBO Journal *20*, 2553-2563. 10.1093/emboj/20.10.2553.

36. Roca, X., Olson, A.J., Rao, A.R., Enerly, E., Kristensen, V.N., Børresen-Dale, A.L., Andresen, B.S., Krainer, A.R., and Sachidanandam, R. (2008). Features of 5'-splice-site efficiency derived from disease-causing mutations and comparative genomics.

Genome Res *18*, 77-87. 10.1101/gr.6859308.

37. Stevens, S.W., and Abelson, J. (1999). Purification of the yeast U4/U6 U5 small nuclear ribonucleoprotein particle and identification of its proteins. Proceedings of the National Academy of Sciences of the United States of America *96*, 7226-7231. 10.1073/pnas.96.13.7226.

38. Zhan, X., Yan, C., Zhang, X., Lei, J., and Shi, Y. (2018). Structures of the human pre-catalytic spliceosome and its precursor spliceosome. Cell Research *28*, 1129-1140. 10.1038/s41422-018-0094-7.

39. Zahler, A.M., Tuttle, J.D., and Chisholm, A. (2004). Genetic Suppression of Intronic +1G Mutations by Compensatory U1 snRNA Changes in Caenorhabditis elegans. Genetics *167*, 1689-1696. 10.1534/genetics.104.028746.

40. Roller, A.B., Hoffman, D.C., and Zahler, A.M. (2000). The allelespecific suppressor sup-39 alters use of cryptic splice sites in Caenorhabditis elegans. Genetics *154*, 1169-1179. 10.1093/genetics/154.3.1169.

41. Kent, W.J., and Zahler, A.M. (2000). The intronerator: exploring introns and alternative splicing in Caenorhabditis elegans. Nucleic Acids Res *28*, 91-93. 10.1093/nar/28.1.91.

42. Dobin, A., Davis, C.A., Schlesinger, F., Jorg, D., Zaleski, C., Jha, S., Philippe, B., Chaisson, M., and Gingeras, T. (2012). STAR: ultrafast universal RNAseq aligner. Bioinformatics *29*, 15-21. 10.1093/bioinformatics/bts635.

43. Zheng, J., Lin, C.-X., Fang, Z.-Y., and Li, H. (2020). Intron Retention as a Mode for RNA-Seq Data Analysis. Frontiers in Genetics *11*. 10.3389/fgene.2020.00586.

44. Parker, M.T., Fica, S.M., Barton, G.J., and Simpson, G.G. (2023). Interspecies association mapping links splice site evolution to METTL16 and SNRNP27K. Elife *12*. 10.7554/eLife.91997.

45. Zahler, A.M., and Roth, M.B. (1995). Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. Proc Natl Acad Sci U S A *92*, 2642-2646. 10.1073/pnas.92.7.2642.

46. Shepard, P.J., and Hertel, K.J. (2009). The SR protein family. GenomeBiology.com (London. Print) *10*, 242-242. 10.1186/gb-2009-10-10242.

47. Nagel, R.J., Lancaster, A.M., and Zahler, A.M. (1998). Specific binding of an exonic splicing enhancer by the pre-mRNA splicing factor SRp55. Rna *4*, 1123.

48. Run, J., Steven, R., Ms, H., van Weeghel, R., Culotti, J., and Way, J. (1996).

Suppressors of the unc-73 Gene of Caenorhabditis elegans. Genetics *143*, 225-236. 10.1093/genetics/143.1.225. 49.

Haeussler, M., Shi, K., Eckert, H., Eschstruth, A., Joffrey, M., Renaud, J., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J.H., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool

CRISPOR. Genome Biology *17*. 10.1186/s13059-016-1012-2.

50. Paix, A., Folkmann, A., and Seydoux, G. (2017). Precision genome editing using CRISPR-Cas9 and linear repair templates in C. elegans. Methods *121122*, 86-93. 10.1016/j.ymeth.2017.03.023.

51. Arribere, J.A., Bell, R.T., Fu, B.X., Artiles, K.L., Hartman, P.S., and Fire, A.Z. (2014). Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. Genetics *198*, 837-846.

10.1534/genetics.114.169730.

52. Crooks, G.E., Hon, G.C., John-Marc, C., and Brenner, S.E. (2004). WebLogo:

A Sequence Logo Generator: Figure 1. Genome Research *14*, 1188-1190.

10.1101/gr.849004.

DocuSign

Certificate Of Completion

Envelope Id: 13C88DBBB62D4F32A3B356B16F7842AC Status: Completed
Subject: Complete with DocuSign: Determining the Role of SNRP-27 and Snu66 in 5' Splice Site Mantinance.docx Source Envelope:
Document Pages: 66 Signatures: 3 Envelope Originator:
Certificate Pages: 5 Initials: 0 Kenna Sarka
AutoNav: Enabled 1156 High Street
EnvelopeId Stamping: Enabled Santa Cruz, CA
Time Zone: (UTC-08:00) Pacific Time (US & Canada) 95064 ksarka@ucsc.edu
IP Address:

67.180.17.231

DocuSigned by:
1856D1409BD64C4...

Record Tracking

Status: Original 12/12/2023 1:25:18 PM
Holder: Kenna Sarka ksarka@ucsc.edu
Location: DocuSign

Signer Events	Signature	Timestamp
Alan Zahler Sent: 12/12/2023 1:28:54 PM zahler@ucsc.edu Viewed: 12/12/2023 1:29:49 PM		
University of California, Santa Cruz Security Level: Email, Account Authentication (Optional)	Signature Adoption: Uploaded Signature Image Using IP Address: 50.168.54.23	Signed: 12/12/2023 1:31:07 PM Freeform Signing DocuSigned by: B07AC47E39FB4D7...

Electronic Record and Signature Disclosure: Accepted: 4/15/2020 4:44:50 PM ID: 44435114-29da-4f3c-8429-82c7dc1c67ca

Melissa Jurica
mjurica@ucsc.edu
University of California, Santa Cruz
Security Level: Email, Account Authentication (Optional)
Signature Adoption: Uploaded Signature Image
Using IP Address: 128.114.151.194
Sent: 12/12/2023 1:31:09 PM
Viewed: 12/12/2023 1:39:06 PM
Signed: 12/12/2023 1:39:20 PM
Freeform Signing

Electronic Record and Signature Disclosure:
Accepted: 1/31/2020 6:38:59 PM ID: 9dd92bc9-48a0-401b-9956d55ae1141037

Electronic Record and Signature Disclosure:
Accepted: 1/8/2020 10:13:28 PM ID: b2485f12-1f0b-47f2-b89dcbd96d3a40b6

SETH RUBIN srubin@ucsc.edu
Seth Rubin
University of California, Santa Cruz Security Level: Email, Account Authentication (Optional)

DocuSigned by:
SETH RUBIN
300A6FD70D154C7...

Signature Adoption: Pre-selected Style Using IP	Address: 169.233.222.241	Sent: 12/12/2023 1:39:23 PM Viewed: 12/12/2023 2:50:36 PM Signed: 12/12/2023 2:50:52 PM Freeform Signing

In Person Signer Events	Signature	Timestamp

Editor Delivery Events	Status	Timestamp

Agent Delivery Events	Status	Timestamp

Intermediary Delivery Events	Status	Timestamp

Certified Delivery Events	Status	Timestamp

Carbon Copy Events	Status	Timestamp

Witness Events	Signature	Timestamp

Notary Events	Signature	Timestamp

Envelope Summary Events	Status	Timestamps
Envelope Sent	Hashed/Encrypted	12/12/2023 1:28:54 PM
Certified Delivered	Security Checked	12/12/2023 2:50:36 PM
Signing Complete	Security Checked	12/12/2023 2:50:52 PM
Completed	Security Checked	12/12/2023 2:50:52 PM

Payment Events	Status	Timestamps

Electronic Record and Signature Disclosure

DOCUSIGN ELECTRONIC RECORD AND SIGNATURE DISCLOSURE

From time to time, the Regents of the University of California, on behalf of its Santa Cruz campus (we, us, UCSC, or the University) may provide to you certain written forms, notices, or disclosures. Described below are the terms and conditions for providing to you such notices and disclosures electronically through the DocuSign system. Please read the information below carefully and thoroughly, and if you can access this information electronically to your satisfaction and agree to this Electronic Record and Signature Disclosure (ERSD), please confirm your agreement by selecting the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system.

Getting paper copies

At any time, you may request from us a paper copy of any record provided or made available electronically to you by us. You will have the ability to download and print documents we send to you through the DocuSign system during and immediately after the signing session and, if you elect to create a DocuSign account, you may access the documents for a limited period of time (usually 30 days) after such documents are first sent to you. After such time, if you wish for us to send you paper copies of any such documents from our office to you, you will be charged a reasonable per-page fee. You may request delivery of such paper copies from us by following the procedure described below.

Withdrawing your consent

If you decide to receive forms, notices and disclosures from us electronically, you may at any time change your mind and tell us that thereafter you want to receive required notices and disclosures only in paper format. How you must inform us of your decision to receive future forms, notices, and disclosure in paper format and withdraw your consent to receive forms, notices, and disclosures electronically is described below.

Consequences of changing your mind

If you elect to receive and/or return required forms, notices, and disclosures only in paper format, it will slow the speed at which we can complete certain steps in transactions with you and delivering services to you because we will need first to send the required notices or disclosures to you in paper format, and then wait until we receive back from you your acknowledgment of your receipt of such paper notices or disclosures or your completed forms. Further, you will no longer be able to use the DocuSign system to receive required forms, notices and consents electronically from us or to sign electronically documents from us. Forms, Notices, and Disclosures may be sent to you electronically Unless you tell us otherwise in accordance with the procedures described herein, we may provide electronically to you through the DocuSign system forms, notices, disclosures, authorizations, acknowledgments, and other documents that are required to be provided or made available to you during the course of our relationship with you. If you do not agree to receive a certain form, notice or disclosure electronically, please let us know as described below. Please also see the paragraph immediately

above that describes the consequences of your electing not to receive delivery of the notices and disclosures electronically from us or return completed forms electronically to us

How to contact us

You may contact us to let us know of your changes as to how we may contact you electronically, to request paper copies of certain information from us, and to withdraw your prior consent to receive notices and disclosures electronically as follows: To advise us of your new email address for DocuSign Usage: To let us know of a change in your email address where we should send certain forms, notices, and disclosures electronically to you, you must send an email message to your primary contact with the University of California, Santa Cruz regarding the documents at issue, and in the body of such email request you must state: your previous email address, your new email address. If you created a DocuSign account, you may update it with your new email address through your account preferences. To request paper copies of documents previously provided to you electronically through this DocuSign account from the University of California, Santa Cruz for routine business and operational transactions: To request delivery from us of paper copies of the notices and disclosures previously provided by us to you electronically, send an email to your primary contact with the University of California, Santa Cruz regarding the documents or transactions at issue and in the body of such request you must state your email address, full name, mailing address, and telephone number. You may be charged a reasonable per-page fee as well as any applicable postage. Note that to request any documents or copies of any documents other than documents previously provided to you through this DocuSign account for routine business and operational transactions, including requests made pursuant to the California Public Records Act (CPRA) and/or the California Information Practices Act (IPA), submit the appropriate Request for Records through the UCSC Information Practices Office, available at https://infopractices.ucsc.edu/

To withdraw your consent

To inform us that you no longer wish to receive certain forms, notices, and disclosures in electronic format you may: i. Decline to sign a document from within your signing session, and on the subsequent page, select the check-box indicating you wish to withdraw your consent;

OR

Send an email to your primary contact with the University of California, Santa Cruz regarding the documents or transactions at issue, and in the body of such request you must state your email, full name, mailing address, and telephone number. Note that if you withdraw your consent to receive documents from us electronically it will slow the speed at which we can complete certain steps in transactions with you and delivering services to you because we will need first to send the required forms, notices or disclosures to you in paper format, and then wait until we receive back from you your acknowledgment of your receipt of such paper notices, disclosures, and/or completed forms.

Required hardware and software

The minimum system requirements for using the DocuSign system may change over time. The current system requirements are found here: https://support.docusign.com/guides/signerguidesigningsystem-requirements.

Acknowledging your access and consent to receive and sign documents electronically To confirm to us that you can access this information electronically, which will be similar to other electronic forms, notices, and disclosures that we will provide to you, please confirm that you have read this ERSD, and (i) that you are able to print on paper or electronically save this ERSD for your future reference and access; or (ii) that you are able to email this ERSD to an email address where you will be able to print on paper or save it for your future reference and access. Further, if you consent to receiving and returning forms, notices, and disclosures in electronic format as described herein, then select the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system. By selecting the check-box next to 'I agree to use electronic records and signatures', you confirm that:

- You can access and read this Electronic Record and Signature Disclosure; and
- You can print on paper this Electronic Record and Signature Disclosure, or save or send this Electronic Record and Disclosure to a location where you can print it, for future reference and access; and
- Until or unless you notify The University of California, Santa Cruz as described above, you consent to receive through electronic means forms, notices, disclosures, authorizations, acknowledgments, and other documents that are required to be provided or made available to you by us during the course of your relationship with The University of California, Santa Cruz.