UNIVERSITY OF CALIFORNIA
SANTA CRUZ

SPLICEOSOMAL HELICASES DDX41/SACY-1 AND PRP22/MOG-5 BOTH CONTRIBUTE TO PROOFREADING AGAINST PROXIMAL 3' SPLICE SITE USAGE

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

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELL, AND DEVELOPMENTAL BIOLOGY

by

Kenneth Osterhoudt

December 2023

The Dissertation of Kenneth Osterhoudt
is approved:

_____
Professor Alan Zahler, Chair

_____
Professor Susan Strome

_____
Professor Manuel Ares, Jr.

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

TABLE OF FIGURES

Abstract

Spliceosomal helicases DDX41/SACY-1 and PRP22/MOG-5 both contribute to proofreading against proximal 3’ splice site usage

Kenneth Osterhoudt

RNA helicases drive necessary rearrangements and ensure fidelity during the pre-mRNA splicing cycle. DEAD-box helicase DDX41 has been linked to human disease and has recently been shown to interact with DEAH-box helicase PRP22 in the spliceosomal C* complex, yet its function in splicing remains unknown. Previous transcriptomic studies of C. elegans depleted of the DDX41 homolog SACY-1 uncovered predominantly changes in alternative 3’ splice site usage. We did a transcriptomic analysis of a viable sacy-1(G533R) allele in staged L3 animals; this allele causes alternative 3’ splicing in introns with pairs of 3’ splice sites separated by ≤ 18 nucleotides. We find that both SACY-1 depletion and the G533R allele lead to a striking unidirectional increase in the usage of proximal (upstream) 3’ splice sites. We have previously discovered a similar alternative splicing pattern between germline tissue and somatic tissue, in which there is a unidirectional increase in proximal 3’ splice site usage in the germline for ~200 events; many of the somatic SACY-1 alternative 3’ splicing events overlap with these developmentally regulated events. We generated a targeted mutant allele of the C. elegans homolog of PRP22, mog-5, in the region of MOG-5 that is predicted to interact with SACY-1 based on the human C* structural model. This viable allele also promotes usage of the proximal alternative adjacent 3’ splice sites in somatic cells. We show that mog-5 and sacy-1 have overlapping proofreading phenotypes against proximal alternative adjacent 3’ splice sites. This work demonstrates that C. elegans is tractable for the genetic study of 3’ss choice after early spliceosomal assembly. These findings are used to inform a program for future genetic research on 3’ss choice.

ACKNOWLEDGMENTS

I am extremely grateful for the excellent mentorship that Professor Alan Zahler provided me. He is very kind, always helpful, and he created a healthy lab environment in which I could thrive. He was a huge inspiration throughout my intellectual development as a scientist. My dissertation committee mentors, Profs Manny Ares and Susan Strome, were extremely helpful throughout my graduate career. My early lab mentors, Dr. Matt Ragle, Dr. Zach Neeb, and Taylor Akers, gave me a strong start as an experimentalist. A huge number of people in the UCSC MCD Biology Department provided help at critical times and were necessary for this work to be completed, with special thanks to Profs Josh Arribere, Melissa Jurica, and Angela Brooks. My mentees Amy Leslie, Diana Escalona, and Ozzie Bagno provided great assistance and made the workday feel more meaningful. I hope all my friends and family know how much their support makes my life whole.

Chapter 1

Introduction

Milestones in Splicing Research

The history of splicing research has been summarized from fascinating different perspectives (Sharp 2005; Mount and Wolin 2015). Here I will outline the major milestones in early splicing research most relevant to understanding the research presented in this thesis.

Splicing was first discovered in 1977 (Berget *et al.* 1977). When adenovirus RNA and DNA coding for the hexon protein were hybridized together, it was noted in electron micrographs that regions of single-stranded DNA looped out because they could not hybridize to the RNA. This indicated that some regions of DNA sequence that occur amidst the coding sequence were not included in the final transcript (Berget *et al.* 1977). This discovery was also made independently by another group the same year (Chow *et al.* 1977). These excised regions were soon termed introns. Before long, the sequences of many introns from a wide variety of species were found by comparing sequences DNA to the sequences of the corresponding messenger RNA (mRNA), and it was seen that the splice sites at both ends of the intron had conserved sequence motifs (Breathnach and Chambon 1981).

The presence of interruptions in the coding sequence of genes was very unexpected. Since introns are not found in prokaryotes, their existence highlighted that the gene structure of eukaryotic genes is significantly different from that of prokaryotes.

The catalytic driver of splicing was not immediately found. Uracil rich small nuclear RNAs (snRNAs) had been discovered in 1968 (Hodnett and Busch 1968; Weinberg and Penman 1968) but had no known function. Antibodies found in lupus patients were found to bind to ribonucleoprotein complexes that each contained different snRNAs. These snRNAs were U1, U2, U4, U5, and U6, and these complexes were called small nuclear ribonucleoproteins (snRNPs)(Lerner and Steitz 1979). It was hypothesized that snRNPs were involved in splicing because of many clues, including that the U1 snRNA was found to

have sequence complementarity with splice site motifs (Lerner *et al.* 1980). Several years after splicing was discovered, it was still unknown which specific proteins or RNAs were involved in catalyzing splicing, and the exact nature of the chemical reactions in the mRNA to remove the introns were unknown.

The development of *in vitro* splicing (Hernandez and Keller 1983) opened new corridors for biochemical characterization of the splicing process. Rather than simply removing the intron as a circular RNA in a single step, it was found that there is an intermediate step to splicing. Splicing creates a branchpoint (BP) adenosine, which is an intronic base that has its 5' and 3' hydroxyl groups still bound to their neighboring intronic nucleotides, but also has its 2' hydroxyl bound to the 5' end of the intron (Padgett *et al.* 1984; Ruskin *et al.* 1984). Since the 5' end of the intron is looped around to connect to the BP, this intermediate was called a lariat.

The pre-mRNA undergoes two transesterification reactions in splicing, reviewed in (Padgett *et al.* 1986). In the first reaction, the BP adenosine's 2' hydroxyl group undergoes nucleophilic attack on the 5' phosphate of the first base of the intron, with the 3' hydroxyl of the upstream exon as the leaving group. This splits the pre-mRNA into two pieces. The 5' portion is still linear, and the 3' portion, here called the splicing intermediate, contains a lariat at its 5' end. In the second reaction, the 3' hydroxyl of the upstream exon performs a nucleophilic attack on the 5' phosphate of the first base of the downstream exon, with the 3' hydroxyl of the last base of the intron as the leaving group. This creates the spliced product, which may have more introns yet to be removed, and an intron lariat, which is degraded.

Research uncovered early clues about the roles of the different snRNPs. U1 was proposed to bind to the 5' splice site (5'ss)(Mount *et al.* 1983). This interaction was found to be functionally important for splicing when a mutation in U1 that compensates for a mutation in a 5'ss to restore the base pairing interaction between them was found to allow splicing of the mutant pre-mRNA (Zhuang and Weiner 1986). This led to the hypothesis that

U1 is responsible for choosing where in the mRNA to use as the 5'ss. Similarly, compensatory mutations in U2 were found to restore splicing of introns with mutated BP sequences (Wu and Manley 1989). Since the BP is usually close to the 3'ss, this implicated U2 in choosing the rough location of the 3'ss. The binding of U1 and U2 snRNAs provided early insight into how the spliceosome first assembles onto the general location of an mRNA where it will splice out an intron.

snRNPs containing U1, U2, and a single snRNP with both U4 and U6 were found to be necessary for splicing *in vitro*, because antibodies against these snRNPs would inhibit splicing (Krainer and Maniatis 1985; Berget and Robberson 1986; Black and Steitz 1986). Splicing is unique in that it is carried out by so many different complexes. Unlike other molecular machines such as the ribosome, the spliceosome cannot be isolated as a single particle that can perform its function. *In vitro* splicing has not been achieved with only overexpressed and reconstituted components; nuclear extract must be included for *in vitro* splicing to occur. The spliceosome is a functional cooperative of macromolecules, so it should not be thought of as a unitary ribonucleoprotein complex.

Studying the snRNPs offers a bird eye view of splicing, but to understand more detailed mechanism, it is necessary to study the roles of individual proteins as well as individual amino acids. A major step towards a finer grain understanding of splicing was the use of yeast temperature sensitive mutations. By bringing known temperature sensitive mutants to their restrictive temperatures, making extracts for *in vitro* splicing, and measuring how extracts from these strains affected splicing, many new genes were linked to splicing (Lustig *et al.* 1986). By 1989, 26 genes that function in splicing had already been identified (Vijayraghavan *et al.* 1989). Many of the RNA helicases discussed below were first shown to function in splicing by this method.

An important example of a protein implicated in splicing through temperature sensitive mutation is Prp8. Prp8 is widely conserved, and it is very large, with its homologs

ranging from 230kd to 280kd, and it can crosslink to the 5'ss, the BP, the 3'ss, as well as U5 and U6 snRNAs, and mutations in Prp8 have led to a wide range of splicing effects, as reviewed in (Grainger and Beggs 2005).

Biochemical and genetic exploration revealed key aspects of the fundamental chemistry of splicing, and many of the macromolecular components of the spliceosome were identified through these approaches. These have led to the model of a splicing cycle (discussed in the next section), in which snRNPs and auxiliary protein factors dynamically assemble, activate, and disassemble to drive the splicing cycle. In recent years, structural biology has become a prominent tool in splicing research. The work of several key labs, including the labs of Luhrmann, Nagai, and Shi among others have led to a series of snapshots of yeast and mammalian spliceosomes at various stages of the splicing cycle and a more complete catalog of the protein components found associated with the spliceosome at each stage (Shi 2017; Kastner *et al.* 2019).

Since the spliceosome consists of so many different macromolecules, the exact nature of the catalytic chemistry of the spliceosome was long a mystery. Splicing was long suspected to be catalyzed by RNA (Steitz and Steitz 1993). The other huge macromolecular machine full of RNA, the ribosome, is a ribozyme (Noller *et al.* 1992). The existence of self-splicing introns, which are pieces of RNA that can catalyze their own excision from a larger piece of RNA, naturally leads to the hypothesis that spliceosomal splicing evolved from self-splicing introns with no need for protein (Kruger *et al.* 1982). The first direct evidence of catalysis by snRNAs came from an experiment with a protein free solution of U2, U6, and a small RNA substrate with a BP. In this experiment, the 2' OH of the BP was shown to covalently bind to U6 upon incubation with metal ions (Valadkhan and Manley 2001). Finally, in 2013 it was shown that both spliceosomal splicing reactions are catalyzed by a ribozyme that employs two magnesium atoms which are positioned and coordinated by the

U2-U6 complex, showing that the key catalytic element is RNA, even though proteins are still required for spliceosomal splicing (Fica *et al.* 2013).

As whole genomes began to be sequenced and assembled, and sequencing technologies continually improved to bring the cost of sequencing down, a new field of transcriptomics came to be. This brought a new scale to our understanding of introns and alternative splicing. By 2000, 28,478 introns and 844 alternatively spliced open reading frames were found in *C. elegans* (the model organism in which the experiments described in this thesis were performed), showing how widespread alternative splicing is, and providing a large library of 5' and 3' splice site sequences (Kent and Zahler 2000). Since the metazoan branchpoint sequence is not well conserved, it cannot be inferred just by viewing the sequence of an intron, so high-throughput sequencing of the intron lariat was required to find the locations of branchpoints genome-wide (Mercer *et al.* 2015). Transcriptomics allows for the observation of splicing phenotypes across thousands of introns, so we are no longer limited only to splicing phenotypes that change splicing of reporters.

RNA Helicases Involved in Spliceosomal Rearrangements and Proofreading

Two RNA helicases, PRP22 and DDX41 are centerpieces of this thesis. Here I will describe the splicing cycle, which is largely driven by RNA helicases, and discuss how proofreading by RNA helicases affects splice site choice.

The spliceosomal components cycle through many different complexes during assembly, catalysis, and disassembly (Figure 1, adapted from (Wilkinson *et al.* 2020)). All the different complexes have major differences in both their composition and in their conformations on the pre-mRNA. Only the complexes that have been experimentally isolated are known, but many unknown intermediate complexes surely exist, and new intermediates are still being discovered. Two important sources of free energy that drive the splicing cycle forward appear to be binding of spliceosomal components to the pre-mRNA

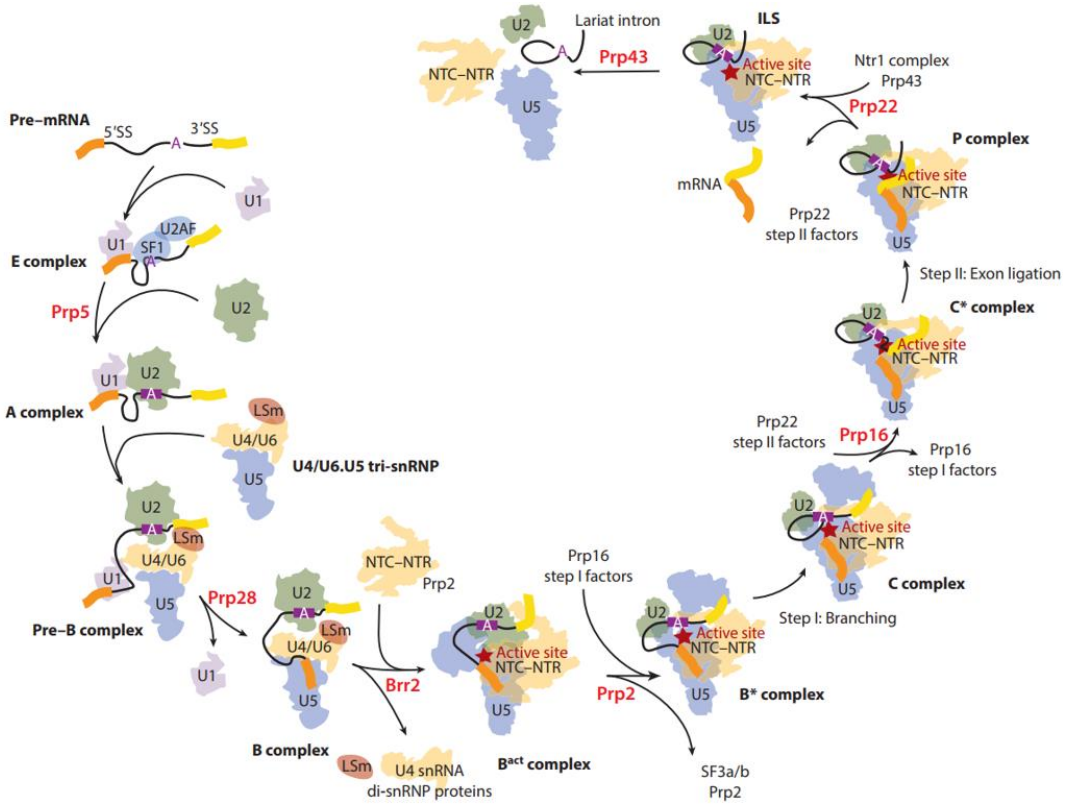

Figure 1. The splicing cycle, adapted from (Wilkinson *et al.* 2020). The spliceosomal complexes are arranged in a count-clockwise circle, showing the binding and disassembly of major components. Genes named in red, next to arrows, refer to splicing helicases required to move from the preceding complex to the following complex.

and to each other, and NTP hydrolysis by RNA helicases. The following is a summary of the roles of helicases in the human splicing cycle adapted from (Wilkinson *et al.* 2020). Splicing is most heavily studied in human and in the budding yeast *S. cerevisiae*. I will focus on the splicing cycle in humans because the *C. elegans* spliceosome is much closer to that of humans.

The first complex is the E complex, which consists of U1 snRNP bound to the 5'ss, the BP bound by the protein SF1, and the polypyrimidine track bound by the U2AF heterodimer. The transition from E complex to A complex requires DEAD-box helicases UAP56 and PRP5. The A complex has U2 snRNP bound to the region around the BP, with SF1 and U2AF removed. When the U4/U6.U5 tri-snRNP joins, the Pre-B complex is formed. This is the only complex with all 5 snRNPs. DEAD-box helicase PRP28 is required to transition to the B complex, which lacks U1 snRNP. BRR2, of the ski2-like helicase family, is required for the removal of snRNA U4 and snRNP U4/U6 proteins, but not U6 snRNA, and for the creation of the catalytic active site of the B^{act} complex. DEAH-box helicase PRP2 is needed to get to the B^{AQR} complex, a newly discovered intermediate, and then helicase activity by the CWF11-family RNA helicase Aquarius is required to form the catalytically active B* complex (Schmitzová *et al.* 2023). B* has the DEAH-box helicase PRP16 bound before catalysis occurs. Once the first transesterification reaction occurs, this complex is called the C complex, although the two complexes are otherwise identical. PRP16 is required to bring about the C* complex, in which PRP16 is replaced by DEAH-box helicase PRP22 in the same approximate location that PRP16 vacates (Fica *et al.* 2017). Once the second transesterification reaction occurs, the complex is called the P complex, but again, these two complexes are otherwise identical. PRP22 is required to remove the spliced mRNA, leaving the ILS (Intron Lariat Spliceosome) complex. DEAH-box helicase PRP43 is required to disassemble the ILS complex.

The biochemical mechanism of many of these RNA helicases is unknown. There is more knowledge of the DEAH-box helicases; during splicing they seem to translocate along single stranded RNA in the 3' to 5' direction (De Bortoli *et al.* 2021). They may not actually unwind RNA helices as their name implies. For the other helicases, even less is known about their biochemical activity in splicing.

Splicing helicases that drive the splicing cycle forward also have a fascinating tendency to be involved in proofreading. Five RNA helicases involved in splicing are currently known to perform proofreading function (De Bortoli *et al.* 2021). It is important to distinguish between alternative splice site choice caused by proofreading and regulated alternative splicing. Proofreading is done by core spliceosomal factors to avoid the usage of unfavorable substrates, while regulated alternative splicing is performed by RNA binding proteins to produce functional alternative isoforms. I will now discuss possible models of proofreading mechanism, and then review the proofreading functionality of RNA helicases.

In principle, there are several ways that proofreading could occur. It could occur actively, in which case a helicase uses energy to alter the mechanistic pathway when an unfavorable state occurs. It could also occur passively; in which case the presence of the protein alters the kinetics of the spliceosome so that it can distinguish between favored and unfavored states. It has mostly been the relatively small set of proteins in the spliceosome that have NTPase activity that have been found to have proofreading activity. This points to active proofreading as a likely hypothesis for at least some of the helicases. However, a passive role for PRP5 has been proposed (Zhang *et al.* 2021)(discussed below). If it occurs actively, two possible models have been proposed. The helicase could be a sensor, in which case it could sense an unfavorable state that causes the helicase to proofread, or it could act as a timer (De Bortoli *et al.* 2021). If it acts as a timer, the helicase only acts to proofread when the splicing pathway is slowed due to an unfavorable state, and under normal conditions the splicing pathway moves forward before the "timer" has time to go off.

Helicases could proofread by aborting the splicing cycle, or by creating a pause and facilitating conditions that allow the unfavorable state to resolve into a more favorable state before continuing onto the next step of the splicing cycle. A simple hypothesis is that at least some of these helicases only have one biochemical ability during splicing: translocating along a specific site of single-stranded RNA to cause rearrangement. If they translocate at the normal time, they drive the splicing cycle forwards, while if they translocate earlier in the cycle than usual, they drive proofreading (Horowitz 2011).

A cryo-electron microscopy (cryo-EM) structure of the spliceosome stalled in A complex by a BP sequence mutation provides insight into proofreading by *S. cerevisiae* Prp5 (Zhang *et al.* 2021). In this structure, the BP mutation has prevented the HEAT domain of Hsh155 (SF3B1 homolog) from closing. In this structure, Prp5 is blocking U2 from rotating into its A complex position. The authors hypothesize that closing of the HEAT domain of Hsh155 would destabilize Prp5, so that it would dissociate and allow U2 to rotate. A mutation which destabilizes Prp5 could cause it to dissociate even with the HEAT domain open, allowing an mRNA with an unfavorable BP sequence to continue onto the A complex, which could then downstream lead to splicing of a cryptic 3'ss. In this model, Prp5 is passively proofreading the BP sequence.

Prp28 is required for the 5'ss to be handed off from its base-pairing interaction with U1 snRNA to a new base-pairing interaction with U6 snRNA (Staley and Guthrie 1999). There are mutations in yeast and *Drosophila* Prp28 that increase usage of unfavorable 5' splice sites, showing that Prp28 is involved in proofreading (Yang *et al.* 2013). In humans, the 5'ss/U1 duplex is held directly in between the two recA domains of Prp28, so it is likely that Prp28 could sense a 5'ss that binds weakly to U1 (Charenton *et al.* 2019).

Prp16 is required for the second transesterification reaction (Schwer and Guthrie 1991). There are mutations in yeast Prp16 that allow usage of a mutated BP, showing that Prp16 is also involved in proofreading the BP (Burgess and Guthrie 1993). The fact that

Prp16 plays a role both before and after the first transesterification reaction is consistent with its presence in both the B* and C complexes. Single-molecule Förster resonance energy transfer (smFRET) experiments show that Prp16 disrupts the binding of a mutated BP within the spliceosome (Semlow *et al.* 2016). The authors hypothesized that Prp16 pulls the mRNA when an unfavorable BP is loaded into the catalytic core, and this pull allows the spliceosome to disengage from that BP and then to bring a new BP into the catalytic core. This same pull is hypothesized to move the splicing cycle forward if Prp16 instead pulls after the first transesterification reaction.

In many ways, Prp22 mirrors Prp16, often performing similar functions to PRP16, just one step later. Prp22 is required for the release of spliced mRNA (Company *et al.* 1991). There are mutations in yeast Prp22 that allow usage of a mutated 3'ss (Mayas *et al.* 2006). Prp22 is present in both C* and P complexes. smFRET experiments show that Prp22 disrupts the binding of a suboptimal 3' splice site (Semlow *et al.* 2016). The authors hypothesize that Prp22 pulling on the mRNA disengages a 3'ss if the pull occurs before the second transesterification reaction, but if the pull occurs after, it removes the mRNA from the spliceosome, parallel to their model for Prp16.

Splicing in *C. elegans*

All the experimental work for this thesis (described in chapter 2) was done using the nematode roundworm *Caenorhabditis elegans* as a model organism. The majority of research on the spliceosome has been performed with human or yeast models. I will now compare the benefits and challenges of studying splicing in human, *S. cerevisiae* and *C. elegans*. I will then review research into 3'ss choice in *C. elegans.*

The aspiration to cure disease and improve treatment drives much research on splicing in humans. Aberrant splicing causes a wide variety of developmental genetic disorders, and many oncogenic mutations are in splicing factors (Jiang and Chen 2021).

Disease alleles also provide insight into spliceosomal mechanism. Some disease alleles are in genes that are not conserved in *S. cerevisiae*, for example DDX41, which is a subject of chapter 2 of this thesis. Another benefit of splicing research in human systems is the many research tools available. Human cell culture model systems are well developed, and a plethora of experimental techniques have been developed. A huge amount of human transcriptomic data from many different backgrounds and conditions are available.

While human tissue culture is an excellent resource for studying splicing biochemistry and transcriptomics, genetic techniques are much more limited. Most tissue culture lines are polyploid, so studying missense alleles is often not possible. The development of stable haploid cell lines with programmable mutations through CRISPR has recently been employed to address this issue (Beusch *et al.* 2023). However, human cell lines are not very amenable to phenotypic assays.

S. cerevisiae is an excellent model for both biochemistry and genetics, and, as discussed above, has led to major insights in our understanding of the mechanisms of splicing. The primary drawback is that there are some aspects of splicing in metazoans that are not present in *S. cerevisiae. S. cerevisiae* does not have widespread alternative splicing (Schirman *et al.* 2021). In addition, the splicing signals at the 5'ss, BP, and 3'ss are more conserved across *S. cerevisiae* introns, while in metazoans these signals are generally weaker. These two features mean that the *S. cerevisiae* spliceosome has been evolving with less ambiguity about where in the mRNA to splice. Since mutations that majorly distort splicing would be lethal in humans, many disease alleles create subtle differences in splicing that only affect splice site choice in situations where the spliceosome is facing significant ambiguity (Scotti and Swanson 2016).

In this thesis, I utilize splicing ambiguity in *C. elegans* splice site choice to gain insight into a homolog of DDX41, a splicing factor linked to disease. I utilize the presence of widespread alternative adjacent splicing that occurs naturally in *C. elegans* (Ragle *et al.*

2015) to assay for splicing function in a way that would not be possible in humans or *S. cerevisiae.*

This unique perspective into ambiguous splice sites justifies our lab's use of this organism to study spliceosomal mechanism. Many techniques for studying splicing biochemistry and structural biology in other organisms have not been developed for *C. elegans*, so the research presented in this thesis consists of genetic and transcriptomic techniques. *C. elegans* does have the advantage of being able to use CRISPR technology to program in any mutant allele that is desired. This advantage is similar to classic *S. cerevisiae* reverse genetic techniques. The splicing factors we study are very highly conserved, so human structural biology and the study of *C. elegans* missense mutations can complement each other. For example, our lab studied how *snrp-27* was involved in 5'ss choice transcriptome-wide (Zahler *et al.* 2018), and soon after this was published, the human homolog was modeled into a structure for the first time, and found to interact with both U6 and U4 snRNAs right before U6 is remodeled to form the catalytic core (Charenton *et al.* 2019).

I will now discuss *C. elegans* introns and constitutive splice site choice. Since this work focuses on conserved elements of core splicing mechanism, I will not review the regulation of alternative splicing or trans-splicing, although there is an interesting body of research on those topics in *C. elegans.*

The introns of *C. elegans* have several interesting features. They are very rich in A and U bases. They are shorter than human introns, with a modal length of 47 and a median length of 65, although some very long *C. elegans* introns do exist (Spieth *et al.* 2014). The 5'ss consensus sequence is AG<u>G</u>UAAGU, with the first base of the intron underlined, and the 3'ss consensus sequence is UUUUCA<u>G</u>, with the last base of the intron underlined, although most introns feature some divergence from these sequences. The branchpoint consensus sequence is unknown, for very few branchpoints have been sequenced in *C.*

elegans. In fact, the only published branchpoint in *C. elegans* is upstream of a pair of alternative adjacent 3' splice sites, and this BP is 14 nucleotides from the proximal splice site and 20 nucleotides from the distance splice site, and has a UACACCA sequence (Ragle *et al.* 2015). *C. elegans* does not have a polypyrimidine tract *per se*, and C bases are quite rare in the 3' end of the intron compared to the purine A (U2AF binding in *C. elegans* is discussed in the next paragraph). There are about 110,000 introns in C. elegans (Spieth *et al.* 2014).

The 3' splice site is first recognized in E complex by the U2AF heterodimer in both humans and *C. elegans*. *uaf-1* (homolog of human U2AF65*)* and *uaf-2* (homolog of human U2AF35*)* crosslink to the *C. elegans* 3'ss (Zorio and Blumenthal 1999). The binding of oligos to *C. elegans* U2AF recapitulate the 3'ss consensus sequence: many oligos that match the *C. elegans* 3'ss except at one position were assayed for binding to U2AF, and oligos showed the largest drop in binding affinity when the one base that did not match was a base that is most strongly conserved in the *C. elegans* 3'ss (Hollins *et al.* 2005). The U bases upstream of the CAG motif are analogous to a short polypyrimidine track; however, it should be noted that substitution with pyrimidine base C significantly decreased binding affinity to U2AF, and in the -6 position an A decreased binding much less than a C. The -5 U the third most strongly conserved base, after the -1 A and -2 G. Mutation of a -5 T in the genome to a G causes a loss of function phenotype in the gene *daf-10,* highlighting how critical this position is for splicing (Itani *et al.* 2016). Mutation in *uaf-1* causes splicing of a cryptic splice site (Ma and Horvitz 2009). *C. elegans* sfa-1 (homolog of SF1) binds to human U2AF65 (Mazroui *et al.* 1999), and mutation in *sfa-1* causes exon skipping and intron retention (Ma *et al.* 2011), but *C. elegans sfa-1* has not been linked directly to branchpoint or 3'ss choice.

From human and yeast studies, we know U2AF leaves the spliceosome early in the splicing cycle. Many subsequent events in the splicing cycle are known to alter 3'ss choice.

The genetics of 3'ss choice in *C. elegans* after A complex have not been directly studied before this thesis.

Many *C. elegans* constitutive splicing factors have been depleted by RNAi, and many mutations in splicing factors have been found, by labs that are not primarily focused on splicing research. Disruption of these factors often lead to germline perturbations. RNAi depletion of many splicing genes disrupt germline proliferation, meiotic entry and germline sex determination (Kerins *et al.* 2010) and distal tip cell migration (Doherty *et al.* 2014). Screening for the masculinization of germline (Mog) phenotype found many splicing factors (Graham and Kimble 1993; Graham *et al.* 1993), as discussed in chapter 2 of this thesis. Several hypotheses have been proposed for this connection (Kerins *et al.* 2010), but why splicing perturbations affect germline more than somatic tissue remains an open question. One hypothesis is that the germline is simply very heavily studied in *C. elegans* with easily detectable phenotypes, so we have found what we are looking for.

The discovery of germline specific alternative adjacent 3'ss choice in *C. elegans* by Matthew Ragle (Ragle *et al.* 2015) laid the foundation for this thesis. In that study, sequenced RNA from gonads isolated by dissection was compared to RNA from worms with disrupted germline development due to temperature sensitive *glp-4* mutation. This comparison uncovered the major splicing difference between soma and germline to be in alternative adjacent (≤18 nucleotides apart) 3'ss pairs and found that the proximal splice site (closer to the 5' end of the intron) shows increased usage in germline tissue relative to somatic tissue. The proximal splice sites consisted of AG nucleotides without a close match to the usual UUUC sequence preceding them, while the distal sites closely matched the sequence which binds to U2AF. This study found that some of these alternative splicing events are conserved among closely related nematodes. This study also found that wildtype worms do regularly use non-AG dinucleotides for the 3'ss in a small number of transcripts (UG, AU, and GG dinucleotides were used) in the germline.

When I began work on this thesis, no mutations were known that altered splicing of these adjacent splice sites. I set out to use these intriguing splice sites as a new approach to study the genetics of 3' splice site choice.

Chapter 2

Spliceosomal helicases DDX41/SACY-1 and PRP22/MOG-5 both contribute to proofreading against proximal 3' splice site usage

Introduction

Introns are removed from precursor messenger RNA (pre-mRNA) by the spliceosome, a dynamic multi-megadalton ribonucleoprotein complex (Wilkinson *et al.* 2020). During the splicing cycle, five small nuclear RNAs (snRNAs) along with many protein splicing factors assemble onto the pre-mRNA and form spliceosomal complexes that transition through several rearrangements and composition changes to create the catalytically active spliceosome with a ribozyme core (Fica *et al.* 2013). The 5' splice site (5'ss) and the branchpoint (BP) are joined in the first trans-esterification reaction to form a lariat-containing splicing intermediate and a free 5' exon. The 5' exon is joined to the 3' exon in the second trans-esterification reaction, with the intron lariat released as a by-product. The spliceosome is then disassembled to start the cycle anew.

Many of the critical assembly, rearrangement, and disassembly steps of the spliceosome cycle require RNA helicases and ATP; eight helicases are conserved in eukaryotes, and five are found in metazoans but not in *S. cerevisiae* (De Bortoli *et al.* 2021). The DEAH-box helicases have been found to translocate along single-stranded RNA in the 3' to 5' direction to enact confirmation changes in the spliceosome, but, contrary to their name, do not necessarily unwind RNA helices during the splicing cycle (Semlow *et al.* 2016). The potential ATPase and helicase mechanism of action in splicing of the DEAD-box helicases is less understood. Five spliceosomal RNA helicases are currently known to proofread against unfavorable mRNA substrates or sequence features (De Bortoli *et al.* 2021).

Many different steps of the splicing cycle have been functionally linked to 3' splice site (3'ss) choice. During early assembly, the approximate location of the 3' end of the intron is initially bound by U2AF proteins which help recruit the U2snRNP. U2 snRNA base pairs with the branchpoint sequence (BPS) (Wu and Manley 1989). The candidate BP can then

be proofread by the DEAH helicase Prp16 (Burgess and Guthrie 1993). When the first trans-esterification reaction happens, BP choice is cemented. For the second step of splicing, the spliceosome must translocate to and load the 3'ss into its active site. A common model is that the spliceosome will scan for the first AG dinucleotide that occurs past a minimal distance downstream from the BP, and choose it for the 3'ss (Smith *et al.* 1989). However, there are examples of AGs found in some introns between the branchpoint and 3'ss that are bypassed by wildtype spliceosomes but become activated in mutant spliceosomes (Chua and Reed 1999). Hundreds of human introns feature a NAGNAG sequence motif in which either AG can be used as the 3'ss (Hiller *et al.* 2004), so scanning alone does not define all 3'ss choice. DEAH-box helicase Prp22 can proofread the 3'ss before splicing is completed (Mayas *et al.* 2006).

Disruption of factors involved in identifying the 3'ss contributes to human disease, and since most eukaryotic transcripts require splicing for mRNA export and translation, all cellular processes are potentially vulnerable to disruption. For example, oncogenic mutation of SF3b1 causes altered BP choice leading to altered 3'ss usage (Darman *et al.* 2015). The pleiotropic effects of disrupting spliceosomal components often make it difficult to untangle the different pathogenic mechanisms; SF3b1 also contributes to malignancy through R-loop formation and DNA damage (Singh *et al.* 2020). DEAD-box helicase DDX41 is another spliceosome-associated protein linked to cancers, and its disruption causes both changes in splicing, as well as R loop formation leading to replicative stress (Shinriki *et al.* 2022). *sacy-1,* the *C. elegans* homologue of DDX41, also has multiple roles; it was initially identified in a screen for oocyte meiotic maturation factors (the name *sacy-1* stands for "suppressor of *acy-1"*)(Kim *et al.* 2012), and was only later understood to be involved in splicing and to be associated with C complex proteins (Tsukamoto *et al.* 2020).

Several important splicing factors in *C. elegans* were initially discovered while studying germline development and were named after the masculinization of germline

(Mog) phenotype (Graham and Kimble 1993); these were only later found to be homologs of splicing proteins. *mog-1*(PRP16), *mog-4*(PRP2), and *mog-5(*PRP22) are all homologs of highly conserved DEAH-box helicases that proofread and then drive the splicing cycle forward. The *sacy-1(P222L)* allele also causes a Mog phenotype (Tsukamoto *et al.* 2020), while knockout is sterile but not Mog for both *sacy-1* (Kim et al. 2012) and *mog-5*(*C. elegans* Deletion Mutant Consortium 2012). Hypomorphic alleles of these factors may prove more useful for studying direct effects on splicing than null mutations in essential genes.

New cryo-electron microscopy (cryo-EM) structures have furthered our understanding of the portion of the splicing cycle that happens after PRP16 remodeling but before the second trans-esterification reaction. Several proteins were recently modeled into the human C* complex for the first time, including DDX41 (Dybkov *et al.* 2023). siRNA-mediated knockdown of some of these C* proteins was found to alter splicing of adjacent NAGNAG 3' splice sites. DDX41 and PRP22 interact in the C* model, with DDX41 modeled into the periphery of the spliceosome, on the opposite from the side of PRP22 that binds the 3' exon (Dybkov *et al.* 2023). In another study, three populations of human spliceosomes, termed pre-C*-I, pre-C*-2, and C* were modeled, providing new insight into the dynamic mechanism that prepares the spliceosome for the second trans-esterification reaction (Zhan *et al.* 2022). Between all three of these structures, PRP22 transitions through different conformations, and these transitions may be involved in PRP22 proofreading activity. These structures have many features that suggest specific, novel functions for specific residues in proteins whose roles in splicing have been unknown. Testing these novel functions will require applying novel functional splicing assays.

C. elegans genetics provides an opportunity to study splice site choice at sites with unusual features. These features allow us to assay the splicing effects of homozygous mutations of conserved amino acids. In Ragle et al., 2015, we found a set of 203 regulated alternative adjacent (≤18 nucleotides apart) 3'ss pairs that show tissue-specific splicing in

the germline; in all cases, the proximal or upstream 3'ss (closer to the 5' end of the intron) shows increased usage in germline tissue relative to somatic tissue (Ragle *et al.* 2015). Interestingly, the proximal splice sites do not have any sequence conservation besides an AG dinucleotide, while the distal or downstream (more towards the 3' end of the pre-mRNA) splice sites closely matched the *C. elegans* UUUCAG 3' splice site consensus. We hypothesized that the distal 3'ss matching the *C. elegans* consensus sequence is a binding site for the U2AF homologs UAF-1/UAF-2 (Zorio and Blumenthal 1999), while the proximal site represented an AG dinucleotide that enters the active site of the spliceosome due to altered translocation from the BP to the 3'ss. These splice sites provide an opportunity to study how a metazoan spliceosome chooses between adjacent 3' splice sites.

Although the function of DDX41 has been the focus of a growing intensity of research activity, and it is recognized as a C* complex protein, there is currently no functional understanding of its role in splicing. In this manuscript, we report that the DDX41 homolog *sacy-1,* and a portion of the PRP22 homolog *mog-5* that is predicted to interact with *sacy-1* both have overlapping phenotypes in 3'ss choice. Disruption of the putative interaction increases usage of proximal 3' splice sites. These results provide direct evidence that conserved residues of *sacy-1* and *mog-5* are required for a C*-linked proofreading mechanism.

Results

SACY-1 depletion causes both directional and sequence content changes in 3' splice site choice

RNA-seq analysis of SACY-1 protein depletion via the auxin-induced degron system leads predominantly to changes in alternative (alt.) 3' splice site choice (Tsukamoto *et al.* 2020). We first decided to determine whether these changes in alt. 3'ss usage showed

the directionality and differences in 3'ss sequence upon SACY-1 depletion that we first noted in the germline in (Ragle *et al.* 2015). We downloaded the data (accession number GSE144003)(Tsukamoto *et al.* 2020) and analyzed alt. splicing using our custom workflow (see methods)(Suzuki *et al.* 2022). We focused on the somatic cell 24-hour auxin-treated depletion samples. We compared their control strain CA1200, which has somatic TIR1 expression, vs DG4703, which has an auxin-induced degron tag on SACY-1 combined with somatic TIR1 expression. We expected to find fewer events than in (Tsukamoto *et al.* 2020), because our workflow is high stringency. We require 15% ΔPSI (change in percent spliced in) for each of the 6 pairwise comparison between replicates (three for DG4703 against two for CA1200) to call alt. splicing events, and then double-check each alt. event by inspecting .bam tracks on the UCSC Genome Browser (Nassar *et al.* 2023).

We called 122 alt. 3' (A3) events between somatic SACY-1-depleted vs non-depleted samples. We only detected five events of other classes of alt. splicing between the samples. Two interesting patterns emerged that were not previously reported for SACY-1 depletion. First, 121 of the events all saw an increase in usage of a proximal 3'ss upon SACY-1 depletion compared to control (hereafter called the directional effect). Second, the distal splice sites closely match the 3'ss consensus sequence, while the proximal splice sites have divergent sequence, usually consisting of an AG dinucleotide with no other consensus pattern (Figure 2) (hereafter called the sequence content effect). 36% of these alt. splicing events matched events that we previously showed are developmentally regulated with increased proximal 3'ss usage in the germline as compared to somatic tissue (Ragle *et al.* 2015). These results show that *sacy-1* performs a specific and ordered role in 3'ss choice that was not previously reported.

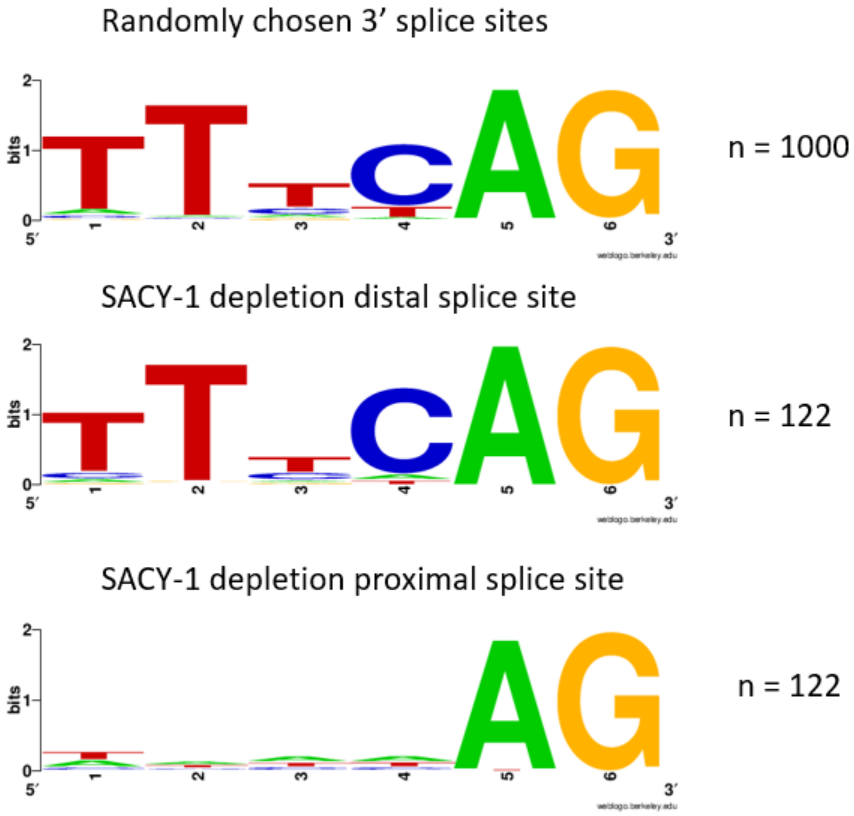

Figure 2. Only the distal splice site in SACY-1 depletion A3 events matches the *C. elegans* consensus sequence. Sequence logo. The height of the base represents significant enrichment of nt identity at each position over random chance, created on (https://weblogo.berkeley.edu/) (Crooks *et al.* 2004).

sacy-1 (G533R) increases proximal splice site usage of developmentally regulated alternative 3' splice sites

We chose to study splicing changes caused by the *sacy-1(tn1385)*(G533R) mutation because it causes a less severe phenotype than complete loss of function (Kim *et al.* 2012), so there are less potentially confounding pleiotropic effects. In addition, the mutation is in a region highly conserved with human DDX41 and adjacent to the human disease allele R525H (Figure 3B). We tested whether *sacy-1*(G533R) increases usage of proximal, developmentally regulated alt. 3' splice sites with divergent sequence by performing reverse transcription-polymerase chain reactions (RT-PCR) followed by polyacrylamide gel electrophoresis (PAGE). We chose alt. splicing events from (Ragle *et al.* 2015) in the genes *icd-2* and *lmd-1*. The event in *icd-2* includes an especially divergent proximal splice site without a canonical AG dinucleotide.

We have previously developed a method to study the effects of a splicing mutation while controlling for the germline specific splicing pattern (Suzuki *et al.* 2022) by collecting embryos by bleaching gravid adults, allowing 34 hours growth to reach the L3 larval stage, and then extracting RNA. The L3 stage is optimal in this assay as this stage is prior to germline expansion and these worms have minimal germline gene expression relative to somatic cells; therefore, we can distinguish between expected normal developmental changes in alt. 3'ss usage in the germline and mutant-induced changes in somatic splicing.

Comparing RNA samples extracted from WT animals at 34 vs 60 hours post-bleaching (the 60 hour samples are adults that have expanded their germline content dramatically relative to the L3 animals at 34 hours), the WT samples show that the germline-specific splicing pattern is not detectable at 34 hours (Figure 3C). For the event in *icd-2*, the 60hr WT sample shows increased proximal 3'ss usage. For the alt. splicing event in *lmd-1,* the 60-hour N2 control does not show increased proximal splicing. A possible reason for this is that most of the *lmd-1* mRNA in the samples may be from somatic tissue

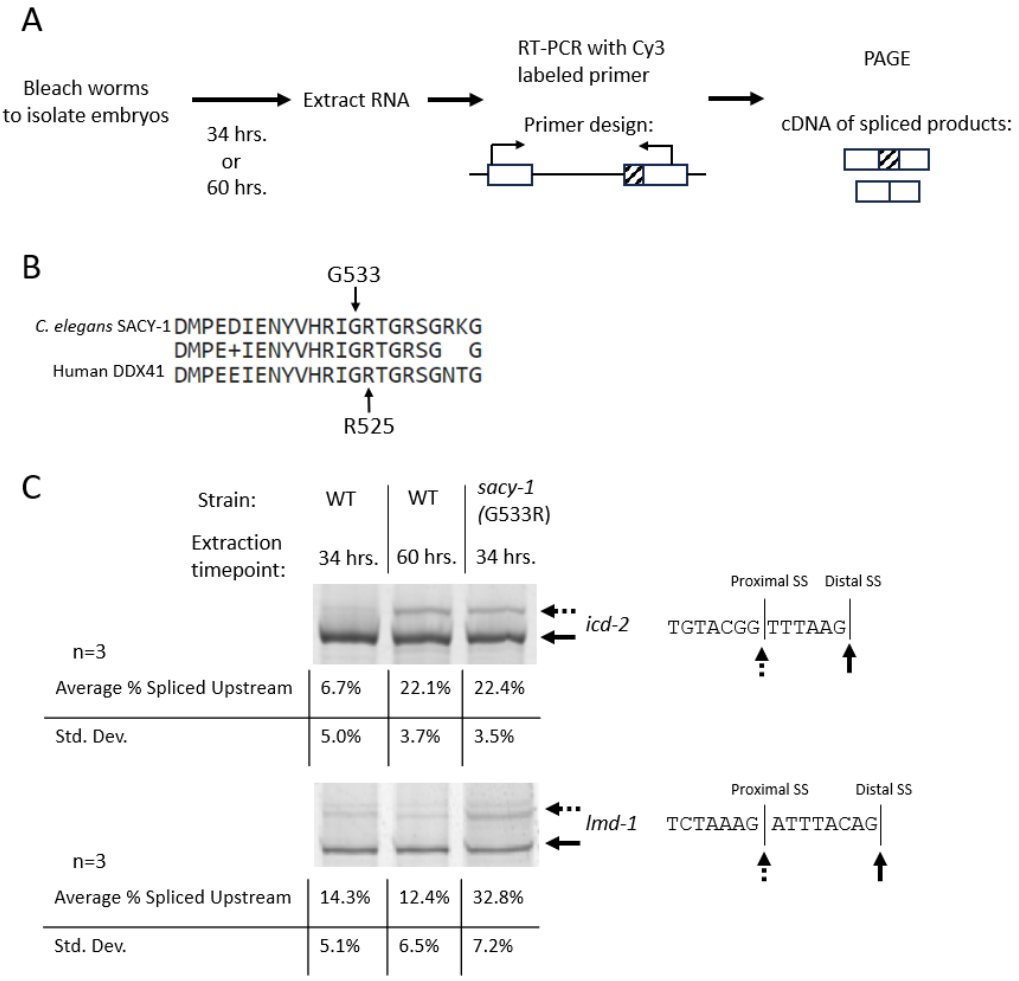

Figure 3. *sacy-1* (G533R) increases proximal splice site usage of developmentally regulated alt. 3' splice sites. (A) Flowchart of experimental procedure. The hatched portion is sequence for which inclusion requires proximal splice site choice. **(B).** BLASTp alignment of *C. elegans* SACY-1 vs human DDX41 at the region around *sacy-1* (G533). **(C).** Splicing assay of RNA from synchronized worms. RT-PCR products run on 6% polyacrylamide gels. The strain and timepoint of RNA extraction are shown above each lane. The sequence of the alt. spliced 3'ss is shown to the right of each gel. T was used rather than U because these splice site sequences are inferred from alignments with the genome rather than directly sequenced. Quantification and standard deviation are shown

below (see methods). Each sample shown is representative of 3 biological replicates; replicates are RNA extracted from independent worm samples with identical conditions.

due to tissue-specific expression patterns, so an increase in proximal splicing in this gene may only be visible using RNA from dissected gonads as was done in (Ragle *et al.* 2015).

For both introns, *sacy-1*(G533R) increases usage of the proximal 3'ss relative to WT for the 34-hour samples (Figure 3C). We find that the *sacy-1*(G533R) reduction-of-function allele affects alt. 3'ss choice in the soma.

Increasing proximal 3' splice site usage is the predominant splicing effect of *sacy-1* (G533R) on the transcriptome

To further study the effects of *sacy-1*(G533R) on splicing, we performed RNA-seq using mutant and WT RNA extracted at 34 hours. Three replicates each were analyzed to identify alt. splicing events. We called alt. 3' (A3) and alt. 5' (A5) splicing events *de novo*, which is helpful in detecting previously unannotated alt. events (Suzuki *et al.* 2022). We also looked for all the different classes of alt. splicing events using annotated datasets. Events that showed greater than 15% ΔPSI in all 9 pairwise comparisons were flagged for further analysis. After verifying called events by hand, we found 211 alt. splicing events between WT and *sacy-1*(G533R). All but 1 of these events were A3 events (Figure 4A), showing that the *sacy-1*(G533R) allele has a splicing effect very specific to 3' splice site choice, which is consistent with the presence of its human homolog DDX41 in the C* complex.

187 of the A3 events featured a pair of splice sites in which *sacy-1*(G533R) increases usage of the proximal site that matches the *C. elegans* 3'ss consensus less closely than the pair's distal site (Figure 4B). Two such splice sites randomly chosen from this group of 187 are shown (Figure 5A). Therefore, both the directional effect and sequence content effect are clearly associated with the *sacy-1*(G533R) mutation. We found 127 unused AG dinucleotides located ≤ 15 nucleotides (nt) downstream of a pair of used splice sites, so *sacy-1*(G533R) does not simply decrease sequence stringency to allow aberrant splicing in either direction, it only provides for an increase in upstream AG usage. There were 14 instances of 2 AG dinucleotides being present ≤ 18 nt upstream of a distal

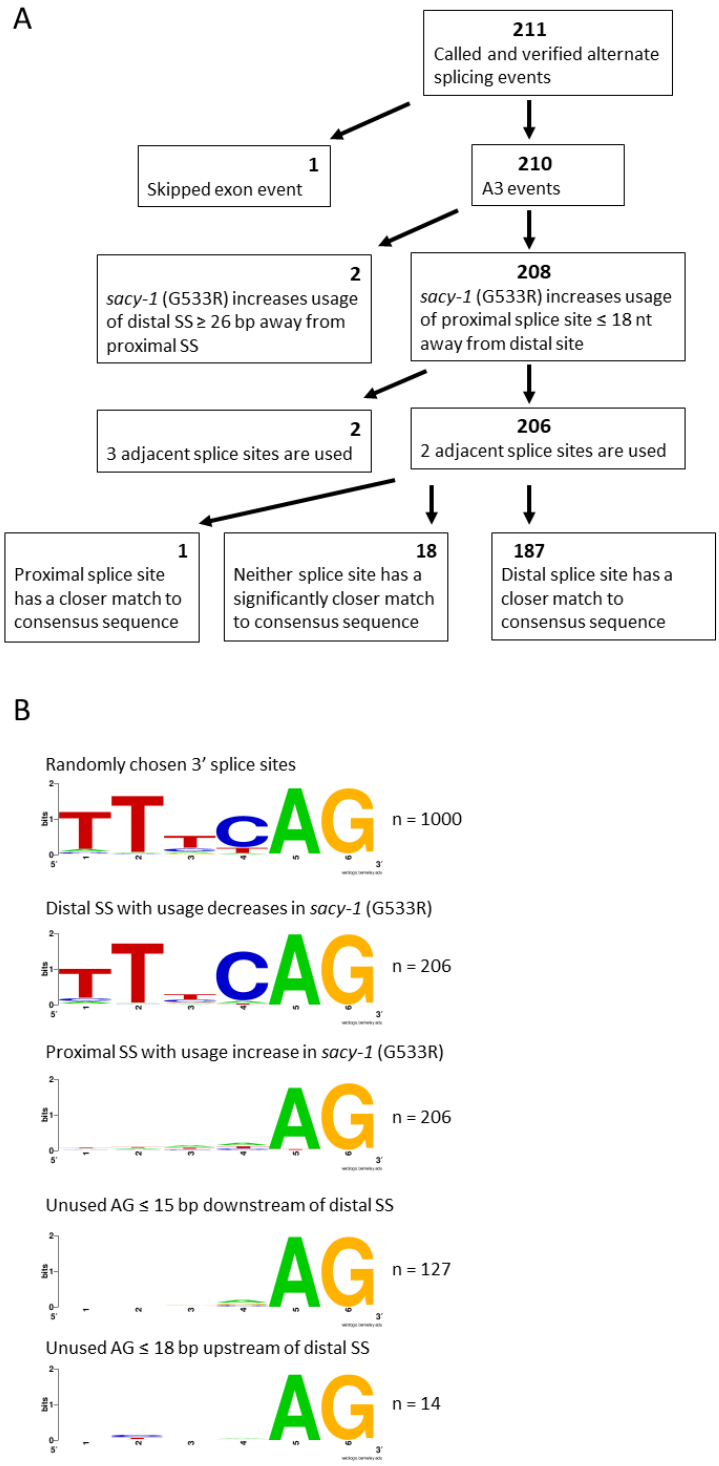

Figure 4. Overview of WT vs *sacy-1*(G533R) RNA-Seq comparative splicing analysis.

(A) Categories of Alt. splicing events. 211 events were divided and sub-divided into specific categories by the features described in the box (see methods). **(B)** Sequence logo. The height of the base represents significant enrichment of nt identity at each position over random chance, created on (https://weblogo.berkeley.edu/) (Crooks *et al.* 2004).

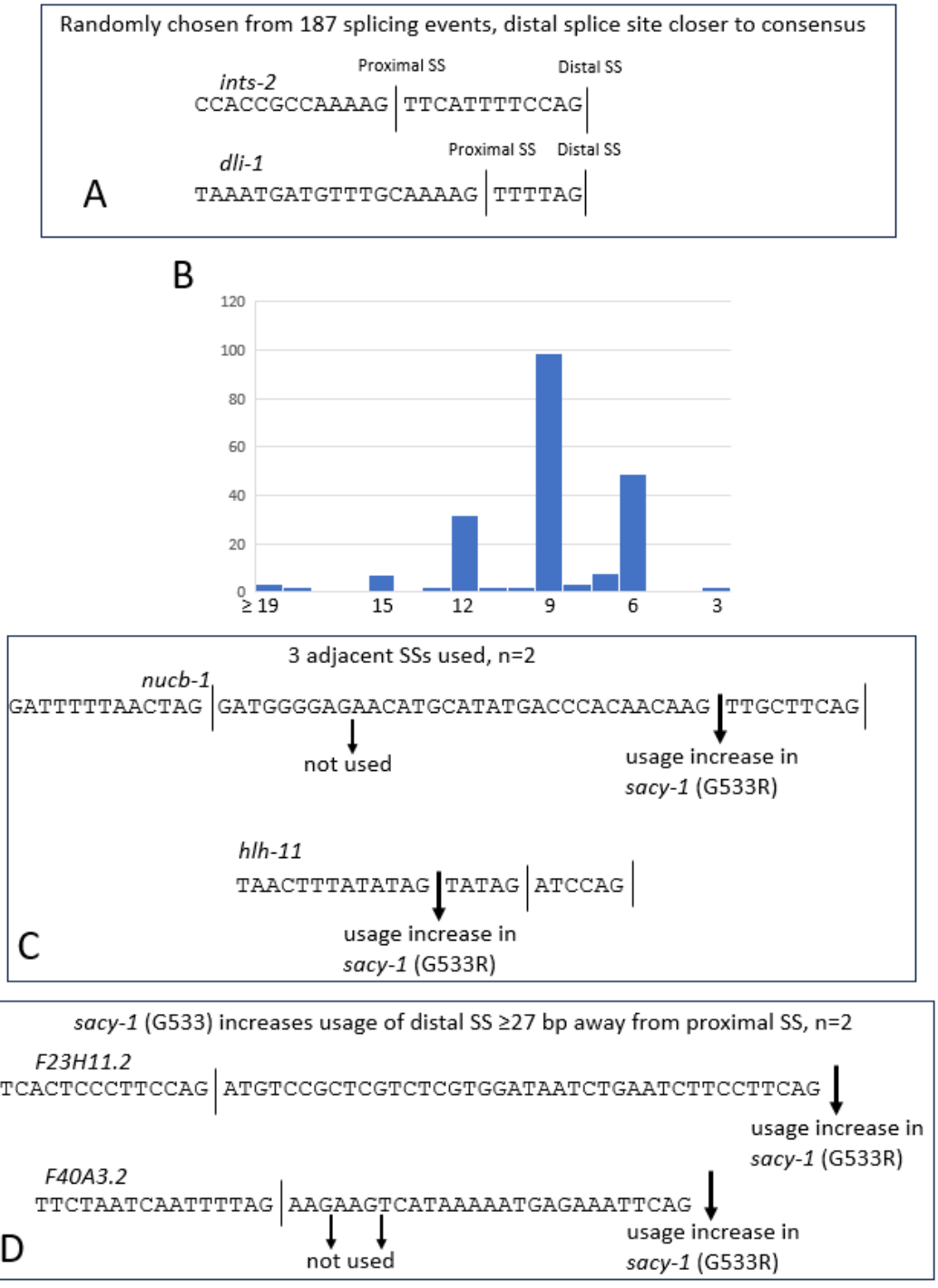

Figure 5. Features of WT vs *sacy-1*(G533R) A3 events. (A) Sequence of two splice site pairs randomly chosen from the 189 pairs in which the distal site has a closer match to the

consensus sequence. **(B)** Histogram of distance between the two splice sites in a pair, including all 210 A3 events **(C)** Sequences of the only two events found in which three adjacent splice sites are used. **(D)** Sequence of the only two events found in which *sacy-1* (G533R) increases distal splice site usage.

site, and in only one case (*hlh-11*, discussed below) were both used. The 13 other AG dinucleotides were further upstream than the *sacy-1*(G533R)-promoted proximal site; these were not used as splice sites, and had no sequence features to distinguish them. 206 of the events had ≤18 nt between a pair of used splice sites, and most of the distances were a multiple of 3 (Figure 5B). It is likely that some alt. splicing events were not captured because they created a frameshift, causing those isoforms to have premature stop codons and be targeted by nonsense-mediated decay (Losson and Lacroute 1979) and therefore not recovered for sequencing.

We studied whether the directional effect or the sequence content effect were stronger. For 208 of the 210 A3 events, *sacy-1*(G533R) increased proximal splice site usage. For 187 of the 206 events with just two splice sites, *sacy-1*(G533R) increased usage of the splice site with a weaker match to the *C. elegans* 3'ss consensus sequence (see methods). Interestingly, for all 18 adjacent alt. 3'ss pairs for which neither splice site is obviously closer to the consensus sequence, the directional effect still occurs. This is true even in the one instance in which the proximal splice site has a closer match to the consensus site than the distal splice site. This is a mechanistic clue that the *sacy-1*(G533R) mutation may be altering the scanning mechanism from the BP to the 3'ss after the first step of splicing. Furthermore, we found 106 pairs of used alt. 3'ss with one or more additional AGs in the vicinity that are not used, compared to only two instances of three adjacent splice sites being used, showing that *sacy-1*(G533R) does not increase usage of all nearby AG dinucleotides, but predominantly increases AGs directly upstream of a potential splice site. We conclude that *sacy-1*(G533R) causes the directional effect in alt. 3'ss usage and does not act primarily through sequence content. The sequence content difference may be due to evolutionary pressure on the 3'ss sequences of introns that results in a low number of introns with two 3' splice sites that match the consensus.

Four outlier events show that the *sacy-1*(G533R) directional effect is local

Four *sacy-1*(G533R) alt 3' events are outliers in the length between the used splice sites or in their splicing patten (Figure 5C,D). Splice sites in *nucb-1*, *F23H11.2*, and *F40A3.2* are separated by distances of 42, 40, and 27 nt between the most upstream and the most downstream splice sites, respectively, while the remaining 207 have a max of ≤18 nt between 3' splice sites. The event in *nucb-1* increases usage of a proximal splice site ≤ 18 nt away from the distal site, and also has an additional upstream splice site used in both WT and mutant. The A3 event in *hlh-11* is the only other event that has three 3' splice sites detected. The long-distance events in *F23H11.2* and *F40A3.2* are also outliers in that they are the only two events we found in which *sacy-1*(G533R) increases usage of a distal splice site. All this implies that the directional effect of *sacy-1*(G533R) has a distance limit, *i.e.* the effect is local.

In *nucb-1,* the middle and further downstream splice sites are close together, and if these two sites are looked at in isolation, their splicing pattern follows the directional effect in which *sacy-1*(G533R) increases usage of a nearby proximal site. Interestingly, there is another AG dinucleotide that is not used 24 nt upstream of the middle splice site, yet this AG cannot be too upstream to be useable, since an AG even further upstream is used. This indicates that the directional effect of *sacy-1*(G533R) is too local to increase usage of a theoretically usable AG 24 nt away. The same pattern occurs in the *F40A3.2* event. In the *F23H11.2* event, there are no intervening AGs, so if the directionality effect of *sacy-1*(G533R) had no distance limit, we would see increased usage of the upstream splice site, yet the opposite occurs. A possible explanation is that *nucb-1*, *F23H11.2*, and *F40A3.2* events have alt. BP choice, and then the directional effect occurs after BP choice. If an upstream BP is used, it may cause the upstream 3'ss to be used. If a downstream BP is used, the most proximal site would be upstream of the BP and therefore unusable. The event in *hlh-11* may feature a BP that is unusually far upstream, and the directionality effect of *sacy-1*(G533R) increases usage of the AG that is closest to the BP.

Disruption of the MOG-5 region predicted to interface with *SACY-1* causes a splicing effect phenotypically overlapping with that of sacy-1(G533R)

Dybkov et al. 2023 report a cryo-EM structural model of the human C* complex which includes the first modelling of DDX41 into the spliceosome. In this model, DDX41 and PRP22 bind to each other and DDX41's only contacts with the spliceosome are through PRP22 (Dybkov *et al.* 2023). Since PRP22 has been linked to proofreading the 3'ss (Mayas *et al.* 2006), and since *sacy-1* and *mog-5* both have a Mog phenotype-causing allele, we hypothesized that SACY-1 and MOG-5 cooperate to proofread against proximal splice sites. We observed that there are amino acids upstream of the recA domains on PRP22 that interact with DDX41 and are very near both PRP22 recA domains (Figure 6). These domains are well conserved between PRP22 and MOG-5. To test if this interaction is involved in proofreading, we performed CRISPR/cas9 genome editing to disrupt this interface in MOG-5. We obtained one *mog-5* allele*, az194*, that matched our repair template; this allele has two missense mutations, K522G and T524G. We also obtained an allele resulting from non-homologous end joining rather than the programmed homology-directed repair, *az192*. This allele replaces the amino acid sequence between K508-T524 in *mog-5*, which is KEMPEWLKHVTAGGKAT, with the 7 amino acids NIMEEIGSS (referred to as *mog-5*(Δ17+9)) creating a much stronger disruption of the interface. The homologous human residues for *mog-5*(K522G), *mog-5*(T524G) and *sacy-1*(G533R) are shown in yellow, and the 17 amino acids deleted in *mog-5*(Δ17+9) are in black (Figure 6B). Both these strains are viable and fertile and can be maintained as homozygotes. This contrasts with the two *mog-5* alleles available from the Caenorhabditis Genetics Center that require maintenance over a balancer; *mog-5*(*q449)* (E608K) that is sterile due to the Mog phenotype (Graham *et al.* 1993), and the *mog-5*(*ok1101)* knockout allele that causes developmental arrest before fertility (*C. elegans* Deletion Mutant Consortium 2012).

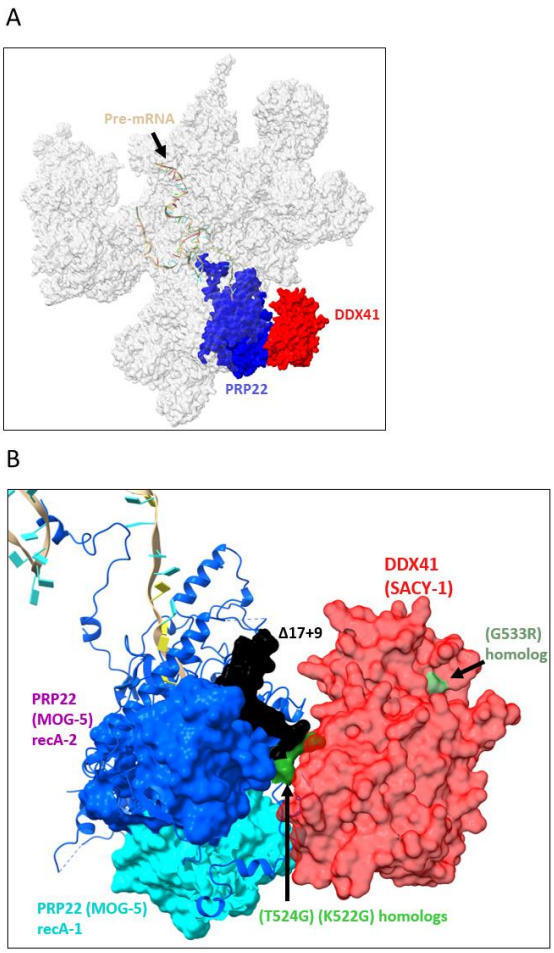

Figure 6. Diagram of *sacy-1* and *mog-5* mutations and the locations of their homologs in C*. (A) Model of C* structure based on coordinates from (Dybkov *et al.* 2023), PDB ID 8C6J, image generated with ChimeraX (Pettersen *et al.* 2021). Shown are the splicing intermediate RNA(tan), PRP22 (blue), and DDX41(red), and all other spliceosomal components are in transparent grey. **(B)** Closer view of PRP22/DDX41 interface showing homologs of mutations, same orientation as Figure 6 A. Shown are homologous positions to *mog-5*(K522G), *mog-5*(T524G) and *sacy-1*(G533R)(green), the region of amino acids that are altered in *mog-5(az192)* D17+9 (black), the RecA-1 domain of PRP22(light teal), and the RecA-2(purple).

We tested whether these two new targeted *mog-5* alleles cause changes to alt. 3'ss usage. We tested them by RT-PCR on 3 alt. splicing events with high ΔPSI in *sacy-1*(G533R). For all 3 introns tested, *mog-5*(Δ17+9) showed a strong increase in proximal splice site usage in L3 animals, but not quite to the same extent as *sacy*-1(G533R) (Figure 7). The *mog-5(*K522G + T524G) mutant increased proximal splice site usage, but to a much lower degree. Given the targeted nature of the new *mog-5* alleles at the interaction site with *sacy-1*, these results demonstrate that *mog-5* and *sacy-1* have a proofreading phenotypic overlap.

Global splicing analysis indicates a functional overlap between *sacy-1*(G533R) and *mog-5*(Δ17+9)

Since all three of the alt. splicing events tested showed functional overlap between *sacy-1*(G533R) and *mog-5*(Δ17+9), we hypothesized that both mutations have widespread overlapping specificity in exactly which 3' splice sites they affect. To test this directly, we performed high-throughput RNA sequencing on WT vs *mog-5*(Δ17+9) RNA extracted at 34 hours using the same pipeline we used with sacy-1(G533R). We identified 76 A3 events with >15% ΔPSI in all pairwise comparisons between WT and the *mog-5* mutant strain . In every called event, *mog-5*(Δ17+9) increased usage off the proximal splice site. The sequence content effect was also present for these A3 events (Figure 8A).

For the 76 *mog-5*(Δ17+9) A3 events that we identified, 53 (70%) also showed an increase in proximal splice usage with *sacy-1*(G533R) (Figure 8B). We conclude that *mog-5* and *sacy-1* have global overlap in proofreading A3 splicing events.

For 50 these 53 overlapping A3 events, WT vs *mog-5*(Δ17+9) showed a smaller ΔPSI compared to WT vs *sacy-1*(G533R). On average, WT vs *mog-5*(Δ17+9) showed a ΔPSI that was 21 percentage point lower than WT vs *sacy-1*(G533R). The lower number of events seen with WT vs *mog-5*(Δ17+9) compared to WT vs *sacy-1*(G533R), and the smaller

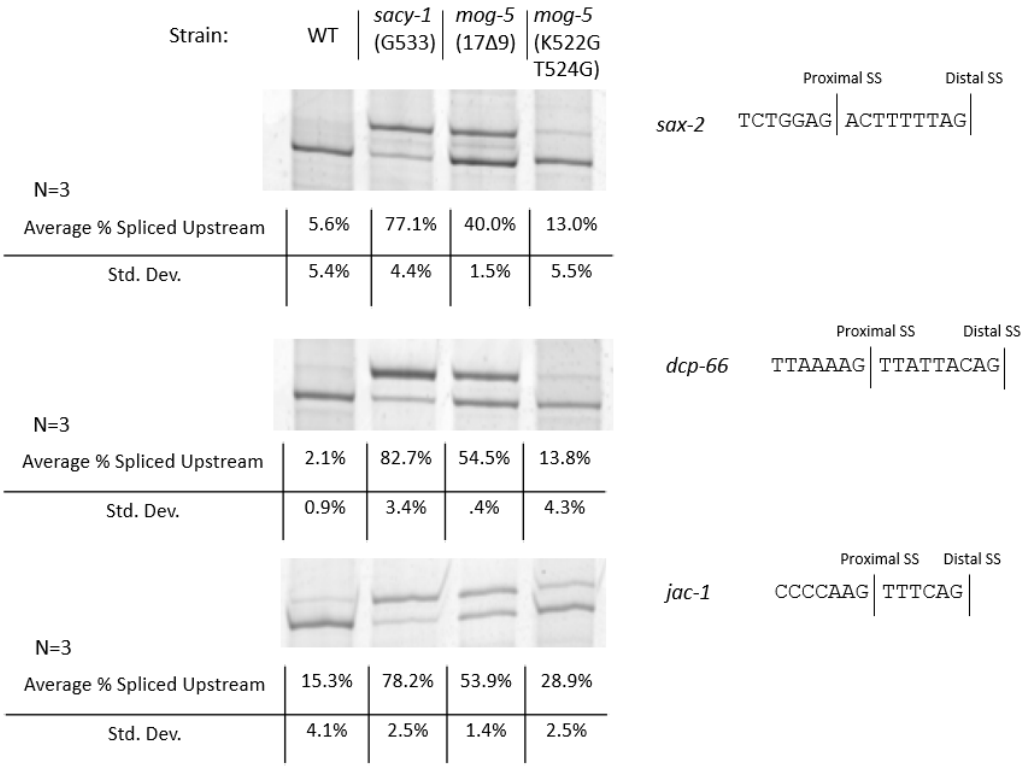

Figure 7. Disruption of the MOG-5 region predicted to interface with SACY-1 changes proximal 3' splice site usage. Splicing assay of RNA from synchronized L3 populations 34 hours after embryo harvest. RT-PCR products run on 6% polyacrylamide gels. The strains are indicated above each lane. The sequence of the alt. spliced 3'ss is shown to the right of each gel. Quantification and standard deviation from three independent experiments for each condition are shown below.

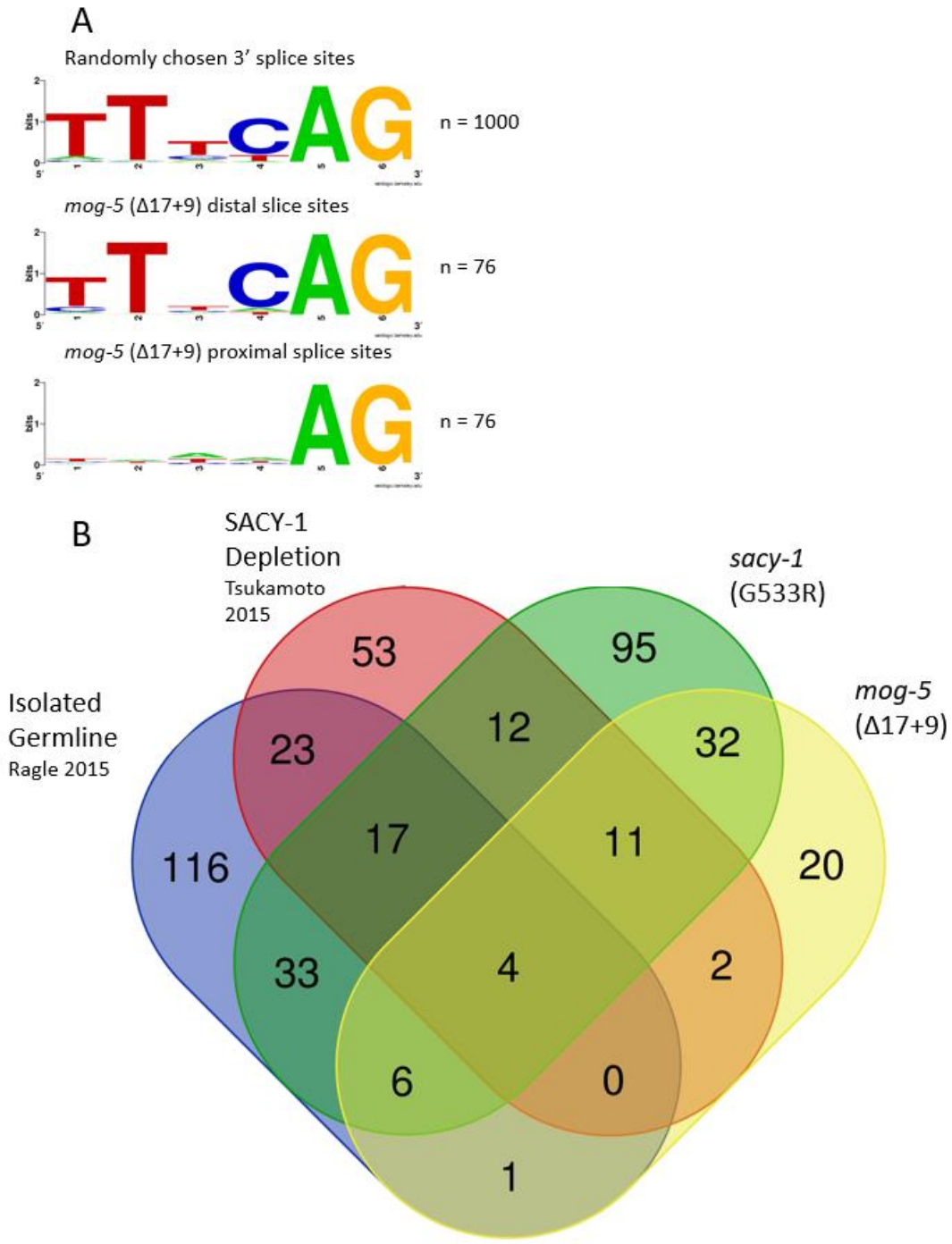

Figure 8. Global splicing analysis indicates a functional overlap between *sacy-1*(G533R) and *mog-5*(Δ17+9). (A) Sequence logo. The height of the base represents significant enrichment of nt identity at each position over random chance, created on

(https://weblogo.berkeley.edu/) (Crooks *et al.* 2004). (B) Overlap of A3 events between 4 RNA-seq datasets. Image created on (https://bioinformatics.psb.ugent.be/webtools/Venn/)

ΔPSI, show that the *mog-5*(Δ17+9) has a weaker splicing effect than *sacy-1*(G533R). *mog-5*(Δ17+9) likely does not completely prevent SACY-1 from interacting with the spliceosome.

The overlap between *sacy-1*(G533R) and *mog-5*(Δ17+9) A3 events is much stronger than the overlaps with the germline events from (Ragle *et al.* 2015) and the SACY-1 depletion events from (Tsukamoto *et al.* 2020) (Figure 7B). Since many differences exist between these experiments, we do not draw conclusions from the absence of overlap. The depletion data are from adult animals on auxin plates, and the germline-specific data are from gonads isolated by dissection. In the two new sequencing libraries prepared for this manuscript, *sacy-1*(G533R) and *mog-5*(Δ17+9), the RNA was prepared from staged L3 animals under the exact same conditions, and in this case we observe 70% overlap for A3 events between alternatively spliced by mutants.

Discussion

We show direct evidence that *mog-5* and *sacy-1* have a phenotypic overlap in 3' splice site choice. Mutant alleles for either gene lead to an inability to proofread against a large and substantially overlapping set of proximal 3' splice sites in favor of adjacent distal splice sites.

The precise location, within the human C* spliceosome, of an intron's path between the BP and a candidate 3'ss has recently been observed (Dybkov *et al.* 2023). Intriguingly, this intron loops around the C-terminus of PRP22, and the BP-to-3'ss-length could change the tension of this interaction, with a longer length potentially causing a much looser interaction. This suggests that PRP22 could sense whether the candidate 3'ss in the catalytic core is a short length from the BP. Because we find that mutant *mog-5* becomes proofreads against proximal 3' splice sites, we hypothesize that PRP22/MOG-5 can sense the short length of the bp-to-proximal-alt.-splice-site distance and proofread against that 3'ss. The protein NKAP interacts with the same portion of the intron as the PRP22 C

terminus, and NKAP also interacts with Slu7. Slu7 prevents splicing of a 3'ss that is a short length from the BP (Chua and Reed 1999), suggesting that these proteins may all be involved in a BP-to-3'ss-length proofreading mechanism. Direct biochemical evidence would be needed to demonstrate if such a mechanism exists.

Since DDX41/*sacy-1* does not have a homolog in *S. cerevisiae,* our results highlight a difference between *S. cerevisiae* and metazoan 3'ss choice mechanisms. It is possible that in metazoans, DDX41/SACY-1 assists or regulates PRP22/ MOG-5 function in splice site choice, while in *S. cerevisiae*, Prp22 performs the analogous function by itself. Metazoans likely evolved additional splice site choice mechanism to create more flexibility in alt. splicing, and to allow for 3'ss choice with less conserved BPS and 3'ss sequence. In addition to DDX41/SACY-1, metazoans have 4 more helicases involved in splicing that are not conserved in *S. cerevisiae* (De Bortoli *et al.* 2021). The evolutionary loss of a splicing helicase has been directly studied. Intriguingly, in *S. pombe*, Prp2 and Aquarius act sequentially to activate the spliceosome, while in *S. cerevisiae*, Prp2 performs activation without an Aquarius homolog (Schmitzová *et al.* 2023). The spliceosome may have divided helicase duties among an increasing number of helicases as it evolved, and PRP22 and DDX41 may likewise share a role in proofreading the second step of splicing.

DDX41/SACY-1 splicing function and C* complex location have only recently been uncovered, so its mechanism of action is still unknown. Recent advances in understanding PRP22's dynamics between pre-C*-I, pre-C*-2, and C* (Zhan *et al.* 2022) suggest models of how DDX41/SACY-1 could work in conjunction with PRP22/MOG-5. Based on these structures, it appears likely that the DDX41 to PRP22 interaction is not compatible with all these conformations, so DDX41 binding would change the kinetics of the transitions between the conformations. DDX41/SACY-1 binding to PRP22/MOG-5 may drive the spliceosome towards a proofreading state, or it may hinder the transition to a catalytic state when an unfavored potential splice site is in the catalytic core. It is possible that

DDX41/SACY-1 plays no role in splicing when the first potential 3'ss loaded into the catalytic core is favorable and proofreading is not needed. Since *sacy-1*(G533R) is potentially analogous to DDX41(R525H), which decreases ATPase activity (Kadono *et al.* 2016), it seems that ATPase activity may be involved in the proofreading effect demonstrated here. To our knowledge there currently are no data providing hints of any potential helicase substrate of DDX41 during splicing, and the only modeled contact that DDX41 makes in the C* complex is with PRP22.

Although human DDX41 proofreading function has not yet been found, we believe that this function is likely to be conserved with *C. elegans* SACY-1 since our targeted *mog-5* alleles were designed based on homology with DDX41 and the human C* structure. However, siRNA-mediated knockdown in HeLa cells of DDX41 and many other C* complex proteins was used to study adjacent 3'ss NAGNAG splicing; this resulted primarily in skipped exon events, and DDX41 knockdown resulted in only a low percentage of its alt. splicing events to be A3 (Dybkov *et al.* 2023). It may be that NAGNAGs are too close together to be recognized by a BP-to-3'SS-length proofreading mechanism. Only 1 out of the 210 A3 events seen with *sacy-1*(G533R) features a pair of splice sites 3 nt apart. It also seems likely that DDX41 knockdown in HeLa cells also disrupted additional mechanisms besides the specific proofreading mechanism, which may have overshadowed the loss of a proofreading function.

This work demonstrates that *C. elegans* genetic studies can complement human biochemical and structural studies to aid our understanding of 3'ss choice. The large number of alt. adjacent 3' splice sites that our lab has uncovered provide a useful variety of intriguing alt. splicing substrates. These substrates differ from human NAGNAG splicing substrates yet are now functionally linked to residues conserved in the human spliceosome. Given that we have demonstrated that mutations to DDX41 and *mog-5* in *C. elegans* lead to splicing changes in somatic cells that have high overlap with the changes we previously

identified as specific to germline cells, we hypothesize that the *C. elegans* germline naturally has an altered, C*-linked 3'ss proofreading mechanism compared to somatic tissue. Studying this difference may provide insight into conserved proofreading mechanisms.

DDX41 plays multiple roles in human cells, and despite recent progress in studying DDX41 oncogenic perturbation, uncovering the pathogenic mechanisms of DDX41 remains an important goal. This study draws further attention to the hypothesis that human DDX41 mutation contributes to pathogenesis by altering 3'ss choice.

Methods

C. elegans staging

Mixed staged worms were treated with bleach to isolate embryos in their egg shells for a rough synchronization of larval stages. See "Protocol 4. Egg prep" from Wormbook: Maintenance of C. elegans (http://www.wormbook.org/chapters/www_strainmaintain/strainmaintain.html)(Stiernagle 2006). Worm embryos were plated on NGM agar with *E. coli* as food and grown at 20°C. For L3 samples, we extracted RNA 34 hours post bleaching, and for adult samples, we extracted RNA 60 hours post bleaching.

RNA Extraction

Staged worms were washed three times in 0.1M NaCl to remove *E. coli*. Worms were gently pelleted, and supernatant was removed to create a pellet of about 50ul. Worm pellets were flash frozen in LN2. 500µl of Trizol was added to samples, vortexed, and incubated for 5 minutes at room temperature. 100µl of $CHCl_3$ was added to samples, vortexed, and incubated for 3 minutes at room temperature. Phases were separated by 15-minute centrifugation at 13,000rpm in a microcentrifuge. The aqueous phase was spin-

column purified and DNase treated using Zymo RNA Clean and Concentrator Kit following manufacturer's instructions.

Library Preparation and RNA sequencing

Azenta performed rRNA depletion followed by library preparation to create 150x150 paired end reads. The WT vs. *sacy-1*(G533R) libraries have strand-specific reads, while the SACY-1 depletion and WT vs mog-5(Δ17+9) libraries are not strand-specific.

RNA-Seq Analysis

Data was downloaded from GEO archive (accession number GSE144003)(Tsukamoto *et al.* 2020) or provided by Azenta. Reads were trimmed, duplicates were removed, quality control analyses were performed, and reads were two-pass aligned to *C. elegans* reference assembly (WS220/ce10) using STAR (Dobin *et al.* 2013). We examined alt. 3' (A3), alt. 3' (A5), alt. first exon (AF), alt. last exon (AL), skipped exon (SE), retained intron (RI), mutually exclusive exon (MX) and multiple skipped exon (MS) events annotated in the the Ensembl gene predictions Archive 65 of WS220/ce10 (EnsArch65) using junctionCounts "infer pairwise events" function (https://github.com/ajw2329/junctionCounts).

For the WT vs. *sacy-1*(G533R) and WT vs mog-5(Δ17+9), we also called *de novo* A5 and A3 splicing events using STAR mappings. After combining libraries, we filtered for A5 and A3 events by filtering for all exon junctions with at least 5 reads of support (total across all samples), determining which of those have a common 5' end and a divergent 3' end or vice versa, filtering for a maximum of 50 nucleotides between the alt. ends, and a minimum of 10 reads spanning the two alt. junction These de novo A3 and A5 events, along with ensArch65 events, were then examined for changes in splicing under our reaction conditions.

For each alternative splicing event, PSI was calculated in every library. A ΔPSI comparing control and test sample was then calculated. For the SACY-1 depletion data, we

did six pairwise comparisons, two replicates of CA1200(one replicate had RNA degradation prior to sequencing) vs three replicates of DG4703 with auxin treatment. For the *sacy-1*(G533R) data, we then did nine pairwise comparisons, three replicates of N2 vs three replicates of DG3430. For WT vs mog-5(Δ17+9) we did 6 pairwise replicates, 3 WT replicates vs 2 mog-5(Δ17+9) samples. Those events with a >15% Δ PSI in all of the pairwise comparisons (pairSum=6 or pairSum=9 in the different experiments) were then analyzed by eye by viewing .bam tracks on the UCSC genome browser (Nassar *et al.* 2023) to verify all alt. splicing events reported in this manuscript. The events with three adjacent splice sites were found this way, but the event calling pipeline and PSI calculations only assumed 2 splice sites.

Splicing event subclassification

The subclassifications shown in Figure 4A were done by hand. The number of used splice sites in an event was determined by viewing aligned .bam reads at the locus of the alt. events. When a pair of adjacent AG splice sites was used, if one splice site had more bases at the -5 and -3 positions (TTTCAG) that match the *C. elegans* 3' splice site consensus sequence, it was binned as stronger. If one site of the pair was not an AG dinucleotide, it was considered weaker. If neither site of the pair was binned stronger by these checks, the event was put in the "Neither splice site has a significantly closer match to consensus sequence" category. We chose to emphasize the -5 position because there is functional evidence of its importance (Itani *et al.* 2016), and the -3 position because this base interacts with the catalytic core (Dybkov *et al.* 2023).

CRISPR/Cas9 Genome editing

The *az192* and *az194* alleles of *mog-5* were created using the CRISPR/Cas9 Genome editing method described in (Suzuki *et al.* 2022).

RT-PCR splicing assay

Reverse transcription was performed using AMV RT. PCR was performed with Cy3 labeled reverse primers. Primer sequences are in Supplemental Table 3. PCR products were run on 6% polyacrylamide/urea denaturing gels for 2 hours at 42W and imaged using a Typhoon scanner. Average % Spliced Upstream is calculated as (mean intensity of upper band - background) / (mean intensity of upper band plus mean intensity of lower band - (background x 2)).

Chapter 3

A research program for the genetic study of 3’ splice site choice

The work described in chapter 2 of this thesis demonstrates that *C. elegans* is tractable for the genetic study of 3'ss choice after early spliceosomal assembly. Here I will present an outline of a research program to comprehensively study the genetics of 3'ss choice in *C. elegans.*

Our lab has used forward genetic screens to find factors that change 5'ss choice. We screened for suppressors of a 5'ss mutation in *unc-73* in which the first base of the intron is mutated from G to T. The suppressors increase usage of a mutant UU splice site in the same position as the old 5'ss. We have identified nine genes that affect 5'ss choice: *sup-6* and *sup-39*, which are both U1 snRNA genes (Roller *et al.* 2000; Zahler *et al.* 2004), *smu-1*, *smu-2*, and *snrp-27* (Dassah *et al.* 2009), *dxbp-1* and *prcc-1* (Suzuki *et al.* 2022), and *prp-8* and *snrp-200* (Cartwright-Acar *et al.* 2022). This demonstrates that forward genetic screens in *C. elegans* can produce viable mutations in a wide variety of splice site choice factors. These mutations are more subtle than full knockout, directly affecting splice site choice without causing confounding pleiotropic effects.

I have attempted to develop a screen for factors that change 3'ss choice. While this attempt was not successful, it did provide lessons to guide future attempts. In addition, the research presented in Chapter 2 provides information that can be used to improve screen design.

To screen for factors that increase usage of a proximal 3'ss that was found to be developmentally regulated in (Ragle *et al.* 2015), I used CRSPR/Cas9 gene editing to swap an alt. spliced intron into *unc-73.* To do this, we substituted a developmentally regulated alternative 3'ss intron from *lmd-1* for a constitutive intron in the *unc-73* gene (Figure 9). I replaced 15th intron in unc-73 with an intron from *lmd-1* that follows this tissue specific 3' splicing pattern. The newly engineered 3' end of the intron contains a downstream AG arranged to create a frameshift, causing Unc phenotype. It also contains an upstream AG in

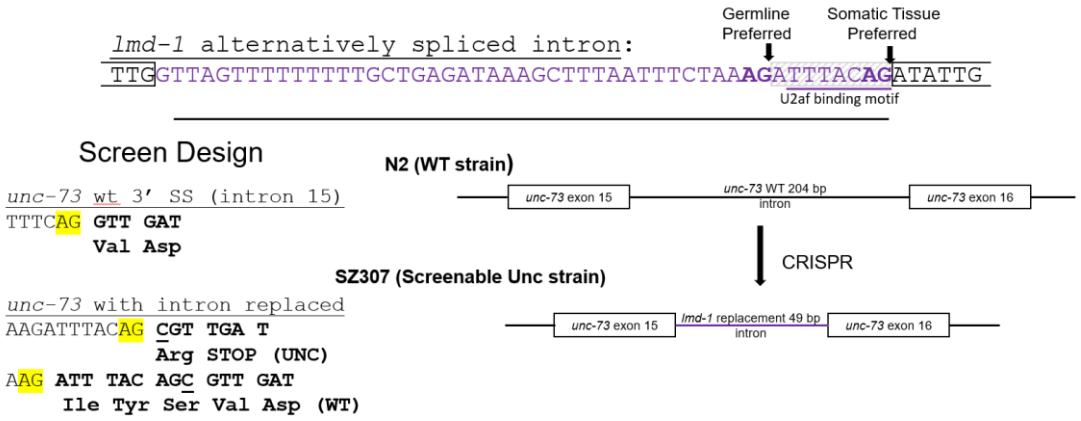

Figure 9. Diagram of splicing reporter SZ307 Adapted from Cold Spring Harbor poster "A forward genetic screen for factors controlling post-recognition choice of adjacent 3' splice sites" by author.

a different reading frame that when used will promote translation of the full protein and relieve the locomotion defect.

In this reporter strain (SZ307), neurons predominantly use the downstream AG and the worms are uncoordinated, although they are not as uncoordinated as the strain our lab normally uses for Unc suppression screens, indicating that there was some baseline of proximal splice site use in WT somatic spliceosomes. I wanted to make sure that the amino acid insertion caused by using a proximal splice site would not cause the Unc phenotype. To test this, I also created a strain with the normal 15th intron, with sequence added at the 5' end of the 16th exon to make the amino acid insertion that upstream splicing of the lmd-1 intron would make. This strain was not Unc, encouraging me to move forward.

We screened the strain by mutagenizing with ENU and picking worms with Unc suppression. We initially found 13 isolates that suppressed Unc. However, the difference in locomotion defect between these suppressors compared to the SZ307 was not very strong. When we looked for alternative 3'ss choice in these strains by RT-PCR, we saw some promising results. However, the splicing effect was inconsistent between replicates, so we decided that it was not replicable. In addition, we were unable to map the suppressor alleles to chromosomes.

One problem may have been that the *lmd-1* alt. splice site was not amenable to a large change in ΔPSI. The lmd-1 intron was chosen based on the fact that the two splice sites were 8 nt apart, so that they would create a frameshift. One lesson form this is that it may be important to consider the ΔPSI seen for a 3'ss before using that splice site as a reporter.

Now that we have found mutants that show alt. 3' splicing, we would be able to test if the reporter successfully shows suppression by crossing a *mog-5* or *sacy-1* mutant into the reporter strain, and seeing if the double mutant has the Unc phenotype clearly suppressed.

Another improvement would be to create a reporter with both Unc phenotype and fluorescence reporting on the expression of the same mRNA. Such a reporter in the gene *unc-54* has been successfully used to screen for different types of mRNA surveillance (Arribere and Fire 2018; Monem *et al.* 2023). It is possible to modify this reporter so that one spliced isoform causes NMD, and another does not. With this system, you could screen for suppression of Unc phenotype, and then get a secondary signal indicating that one isoform's product is more expressed in the suppressor by measuring increased fluorescence.

As described in chapter 2 of this thesis, we have assembled hundreds of alt. 3' splice sites that are alternatively spliced by different mechanisms. This library could be used to find different 3'ss that would match a variety of criteria that different experimental questions might call for. For example, if we wanted to study the proofreading mechanism, we may choose a 3'ss that is heavily alternatively spliced by both *sacy-1*(G533R) and *mog-5*(Δ17+9), while if we wanted to study what causes the germline specific alternative splicing pattern, we could choose an intron that is alternatively spliced germline relative to soma, but is not alternatively spliced by *sacy-1* and *mog-5* perturbations.

While the screen above would be used to functionally link splicing factors to 3'ss choice, we could also develop genetic screens to search for genetic interactions between splicing factors. For example, we could screen for suppression of Mog phenotype. We have strains in which a Mog mutation is balanced by a balancer chromosome that expresses GFP. In these strains, all of the non-GFP expressing worms are sterile due to mutation in a Mog causing gene. We could mutagenize the heterozygote balancer worms, use a robotic worm sorting system to move all the non GFP expressing F2 progeny to separate plates, and then search for fertile individuals on these plates. We would then conduct splicing assays to make sure the new alleles suppress splicing changes, rather than function in a different part of the germ-oocyte switch pathway, before mapping the mutant.

In addition to forward genetics, the research presented in this thesis shows that reverse genetics can be used to study homologs of human proteins found in the C* complex. The human C* structures can be used to find amino acids in interesting positions and generate functional hypothesis that could be tested with the reverse genetics methods used in this work. For example, in (Dybkov *et al.* 2023) they studied a set of proteins that are not conserved in *S. cerevisiae* and are recruited to the C* complex: FAM32A, SDE2, CACTIN, PRKRIP1, NKAP, TLS1, CXORF56, FAM50A, PPIL3, PPIG, ESS2, NOSIP, DDX41, DHX35, and GPATCH1. Besides the results from this study showing that siRNA-mediated knockdown of these proteins causes alternative splicing, specific functions for many of these proteins are not known. Since these proteins can affect alternative splicing, and one member of this list, DDX41, was already used to inspire our successful reverse genetic study, these proteins make good candidates for further research in our system.

Mutations in splicing factors could be studied even if they do not change splicing of the alt. 3' splice sites that we have uncovered. It is likely that the loss of function phenotypes of some of these factors would be germline perturbations. If a large number of mutants were created by reverse genetics, we may be able to place them into functional groups by their germline phenotypes. Our lab has used a method to create random missense alleles in specific amino acid positions by including repair templates with random bases during CRISPR/Cas9 mutagenesis (Cartwright-Acar *et al.* 2022). This method can be used to get multiple alleles with different phenotypes at a single genomic locus from a single round of CRISPR Cas/9 gene editing, increasing the speed at which we could collect alleles in splicing factors with different classes of germline phenotypes.

Bibliography

Arribere, J. A., and A. Z. Fire. 2018. "Nonsense mRNA suppression via nonstop decay." *Elife* 7.

Berget, S. M., C. Moore, and P. A. Sharp. 1977. "Spliced segments at the 5' terminus of adenovirus 2 late mRNA." *Proc Natl Acad Sci U S A* 74(8):3171-5.

Berget, S. M., and B. L. Robberson. 1986. "U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation." *Cell* 46(5):691-6.

Beusch, I., B. Rao, M. K. Studer, T. Luhovska, V. Šukytė, S. Lei, J. Oses-Prieto, E. SeGraves, A. Burlingame, S. Jonas, and H. D. Madhani. 2023. "Targeted high-throughput mutagenesis of the human spliceosome reveals its in vivo operating principles." *Mol Cell* 83(14):2578-2594.e9.

Black, D. L., and J. A. Steitz. 1986. "Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein." *Cell* 46(5):697-704.

Breathnach, R., and P. Chambon. 1981. "Organization and expression of eucaryotic split genes coding for proteins." *Annu Rev Biochem* 50:349-83.

Burgess, S. M., and C. Guthrie. 1993. "A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates." *Cell* 73(7):1377-91.

Cartwright-Acar, C. H., K. Osterhoudt, Jmngl Suzuki, D. R. Gomez, S. Katzman, and A. M. Zahler. 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(20):11834-11857.

Charenton, C., M. E. Wilkinson, and K. Nagai. 2019. "Mechanism of 5' splice site transfer for human spliceosome activation." *Science* 364(6438):362-367.

Chow, L. T., R. E. Gelinas, T. R. Broker, and R. J. Roberts. 1977. "An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA." *Cell* 12(1):1-8.

Chua, K., and R. Reed. 1999. "The RNA splicing factor hSlu7 is required for correct 3' splice-site choice." *Nature* 402(6758):207-10.

Company, M., J. Arenas, and J. Abelson. 1991. "Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes." *Nature* 349(6309):487-93.

Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner. 2004. "WebLogo: a sequence logo generator." *Genome Res* 14(6):1188-90.

Darman, R. B., M. Seiler, A. A. Agrawal, K. H. Lim, S. Peng, D. Aird, S. L. Bailey, E. B. Bhavsar, B. Chan, S. Colla, L. Corson, J. Feala, P. Fekkes, K. Ichikawa, G. F. Keaney, L. Lee, P. Kumar, K. Kunii, C. MacKenzie, M. Matijevic, Y. Mizui, K. Myint, E. S. Park, X. Puyang, A. Selvaraj, M. P. Thomas, J. Tsai, J. Y. Wang, M. Warmuth, H. Yang, P. Zhu, G. Garcia-Manero, R. R. Furman, L. Yu, P. G. Smith, and S. Buonamici. 2015. "Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point." *Cell Rep* 13(5):1033-45.

Dassah, M., S. Patzek, V. M. Hunt, P. E. Medina, and A. M. Zahler. 2009. "A genetic screen for suppressors of a mutated 5' splice site identifies factors associated with later steps of spliceosome assembly." *Genetics* 182(3):725-34.

De Bortoli, F., S. Espinosa, and R. Zhao. 2021. "DEAH-Box RNA Helicases in Pre-mRNA Splicing." *Trends Biochem Sci* 46(3):225-238.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics* 29(1):15-21.

Doherty, M. F., G. Adelmant, A. D. Cecchetelli, J. A. Marto, and E. J. Cram. 2014. "Proteomic analysis reveals CACN-1 is a component of the spliceosome in Caenorhabditis elegans." *G3 (Bethesda)* 4(8):1555-64.

Dybkov, O., M. Preußner, L. El Ayoubi, V. Y. Feng, C. Harnisch, K. Merz, P. Leupold, P. Yudichev, D. E. Agafonov, C. L. Will, C. Girard, C. Dienemann, H. Urlaub, B. Kastner, F. Heyd, and R. Lührmann. 2023. "Regulation of 3' splice site selection after step 1 of splicing by spliceosomal C* proteins." *Sci Adv* 9(9):eadf1785.

Fica, S. M., C. Oubridge, W. P. Galej, M. E. Wilkinson, X. C. Bai, A. J. Newman, and K. Nagai. 2017. "Structure of a spliceosome remodelled for exon ligation." *Nature* 542(7641):377-380.

Fica, S. M., N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley, and J. A. Piccirilli. 2013. "RNA catalyses nuclear pre-mRNA splicing." *Nature* 503(7475):229-34.

Graham, P. L., and J. Kimble. 1993. "The mog-1 gene is required for the switch from spermatogenesis to oogenesis in Caenorhabditis elegans." *Genetics* 133(4):919-31.

Graham, P. L., T. Schedl, and J. Kimble. 1993. "More mog genes that influence the switch from spermatogenesis to oogenesis in the hermaphrodite germ line of Caenorhabditis elegans." *Dev Genet* 14(6):471-84.

Grainger, R. J., and J. D. Beggs. 2005. "Prp8 protein: at the heart of the spliceosome." *Rna* 11(5):533-57.

Hernandez, N., and W. Keller. 1983. "Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts." *Cell* 35(1):89-99.

Hiller, M., K. Huse, K. Szafranski, N. Jahn, J. Hampe, S. Schreiber, R. Backofen, and M. Platzer. 2004. "Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity." *Nat Genet* 36(12):1255-7.

Hodnett, J. L., and H. Busch. 1968. "Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei." *J Biol Chem* 243(24):6334-42.

Hollins, C., D. A. Zorio, M. MacMorris, and T. Blumenthal. 2005. "U2AF binding selects for the high conservation of the C. elegans 3' splice site." *Rna* 11(3):248-53.

Horowitz, D. S. 2011. "The splice is right: guarantors of fidelity in pre-mRNA splicing." *Rna* 17(4):551-4.

Itani, O. A., S. Flibotte, K. J. Dumas, C. Guo, T. Blumenthal, and P. J. Hu. 2016. "N-Ethyl-N-Nitrosourea (ENU) Mutagenesis Reveals an Intronic Residue Critical for Caenorhabditis elegans 3' Splice Site Function in Vivo." *G3 (Bethesda)* 6(6):1751-6.

Jiang, W., and L. Chen. 2021. "Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing." *Comput Struct Biotechnol J* 19:183-195.

Kadono, M., A. Kanai, A. Nagamachi, S. Shinriki, J. Kawata, K. Iwato, T. Kyo, K. Oshima, A. Yokoyama, T. Kawamura, R. Nagase, D. Inoue, T. Kitamura, T. Inaba, T. Ichinohe, and H. Matsui. 2016. "Biological implications of somatic DDX41 p.R525H mutation in acute myeloid leukemia." *Exp Hematol* 44(8):745-754.e4.

Kastner, B., C. L. Will, H. Stark, and R. Lührmann. 2019. "Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes." *Cold Spring Harb Perspect Biol* 11(11).

Kent, W. J., and A. M. Zahler. 2000. "The intronerator: exploring introns and alternative splicing in Caenorhabditis elegans." *Nucleic Acids Res* 28(1):91-3.

Kerins, J. A., M. Hanazawa, M. Dorsett, and T. Schedl. 2010. "PRP-17 and the pre-mRNA splicing pathway are preferentially required for the proliferation versus meiotic development decision and germline sex determination in Caenorhabditis elegans." *Dev Dyn* 239(5):1555-72.

Kim, S., J. A. Govindan, Z. J. Tu, and D. Greenstein. 2012. "SACY-1 DEAD-Box helicase links the somatic control of oocyte meiotic maturation to the sperm-to-oocyte switch and gamete maintenance in Caenorhabditis elegans." *Genetics* 192(3):905-28.

Krainer, A. R., and T. Maniatis. 1985. "Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro." *Cell* 42(3):725-36.

Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech. 1982. "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena." *Cell* 31(1):147-57.

Lerner, M. R., J. A. Boyle, S. M. Mount, S. L. Wolin, and J. A. Steitz. 1980. "Are snRNPs involved in splicing?" *Nature* 283(5743):220-4.

Lerner, M. R., and J. A. Steitz. 1979. "Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus." *Proc Natl Acad Sci U S A* 76(11):5495-9.

Letunic, I., S. Khedkar, and P. Bork. 2021. "SMART: recent updates, new developments and status in 2020." *Nucleic Acids Res* 49(D1):D458-d460.

Losson, R., and F. Lacroute. 1979. "Interference of nonsense mutations with eukaryotic messenger RNA stability." *Proc Natl Acad Sci U S A* 76(10):5134-7.

Lustig, A. J., R. J. Lin, and J. Abelson. 1986. "The yeast RNA gene products are essential for mRNA splicing in vitro." *Cell* 47(6):953-63.

Ma, L., and H. R. Horvitz. 2009. "Mutations in the Caenorhabditis elegans U2AF large subunit UAF-1 alter the choice of a 3' splice site in vivo." *PLoS Genet* 5(11):e1000708.

Ma, L., Z. Tan, Y. Teng, S. Hoersch, and H. R. Horvitz. 2011. "In vivo effects on intron retention and exon skipping by the U2AF large subunit and SF1/BBP in the nematode Caenorhabditis elegans." *Rna* 17(12):2201-11.

Mayas, R. M., H. Maita, and J. P. Staley. 2006. "Exon ligation is proofread by the DExD/H-box ATPase Prp22p." *Nat Struct Mol Biol* 13(6):482-90.

Mazroui, R., A. Puoti, and A. Krämer. 1999. "Splicing factor SF1 from Drosophila and Caenorhabditis: presence of an N-terminal RS domain and requirement for viability." *Rna* 5(12):1615-31.

Mercer, T. R., M. B. Clark, S. B. Andersen, M. E. Brunck, W. Haerty, J. Crawford, R. J. Taft, L. K. Nielsen, M. E. Dinger, and J. S. Mattick. 2015. "Genome-wide discovery of human splicing branchpoints." *Genome Res* 25(2):290-303.

Monem, P. C., N. Vidyasagar, A. L. Piatt, E. Sehgal, and J. A. Arribere. 2023. "Ubiquitination of stalled ribosomes enables mRNA decay via HBS-1 and NONU-1 in vivo." *PLoS Genet* 19(1):e1010577.

Mount, S. M., I. Pettersson, M. Hinterberger, A. Karmas, and J. A. Steitz. 1983. "The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro." *Cell* 33(2):509-18.

Mount, S. M., and S. L. Wolin. 2015. "Recognizing the 35th anniversary of the proposal that snRNPs are involved in splicing." *Mol Biol Cell* 26(20):3557-60.

Nassar, L. R., G. P. Barber, A. Benet-Pagès, J. Casper, H. Clawson, M. Diekhans, C. Fischer, J. N. Gonzalez, A. S. Hinrichs, B. T. Lee, C. M. Lee, P. Muthuraman, B. Nguy, T. Pereira, P. Nejad, G. Perez, B. J. Raney, D. Schmelter, M. L. Speir, B. D. Wick, A. S. Zweig, D. Haussler, R. M. Kuhn, M. Haeussler, and W. J. Kent. 2023. "The UCSC Genome Browser database: 2023 update." *Nucleic Acids Res* 51(D1):D1188-d1195.

Noller, H. F., V. Hoffarth, and L. Zimniak. 1992. "Unusual resistance of peptidyl transferase to protein extraction procedures." *Science* 256(5062):1416-9.

Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp. 1986. "Splicing of messenger RNA precursors." *Annu Rev Biochem* 55:1119-50.

Padgett, R. A., M. M. Konarska, P. J. Grabowski, S. F. Hardy, and P. A. Sharp. 1984. "Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors." *Science* 225(4665):898-903.

Pettersen, E. F., T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin. 2021. "UCSF ChimeraX: Structure visualization for researchers, educators, and developers." *Protein Sci* 30(1):70-82.

Ragle, J. M., S. Katzman, T. F. Akers, S. Barberan-Soler, and A. M. Zahler. 2015. "Coordinated tissue-specific regulation of adjacent alternative 3' splice sites in C. elegans." *Genome Res* 25(7):982-94.

Roller, A. B., D. C. Hoffman, and A. M. Zahler. 2000. "The allele-specific suppressor sup-39 alters use of cryptic splice sites in Caenorhabditis elegans." *Genetics* 154(3):1169-79.

Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green. 1984. "Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro." *Cell* 38(1):317-31.

Schirman, D., Z. Yakhini, Y. Pilpel, and O. Dahan. 2021. "A broad analysis of splicing regulation in yeast using a large library of synthetic introns." *PLoS Genet* 17(9):e1009805.

Schmitzová, J., C. Cretu, C. Dienemann, H. Urlaub, and V. Pena. 2023. "Structural basis of catalytic activation in human splicing." *Nature* 617(7962):842-850.

Schwer, B., and C. Guthrie. 1991. "PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome." *Nature* 349(6309):494-9.

Scotti, M. M., and M. S. Swanson. 2016. "RNA mis-splicing in disease." *Nat Rev Genet* 17(1):19-32.

Semlow, D. R., M. R. Blanco, N. G. Walter, and J. P. Staley. 2016. "Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites." *Cell* 164(5):985-98.

Sharp, P. A. 2005. "The discovery of split genes and RNA splicing." Pp. 279-81 in *Trends Biochem Sci*. England.

Shi, Y. 2017. "Mechanistic insights into precursor messenger RNA splicing by the spliceosome." *Nat Rev Mol Cell Biol* 18(11):655-670.

Shinriki, S., M. Hirayama, A. Nagamachi, A. Yokoyama, T. Kawamura, A. Kanai, H. Kawai, J. Iwakiri, R. Liu, M. Maeshiro, S. Tungalag, M. Tasaki, M. Ueda, K. Tomizawa, N. Kataoka, T. Ideue, Y. Suzuki, K. Asai, T. Tani, T. Inaba, and H. Matsui. 2022. "DDX41 coordinates RNA splicing and transcriptional elongation to prevent DNA replication stress in hematopoietic cells." *Leukemia* 36(11):2605-2620.

Singh, S., D. Ahmed, H. Dolatshad, D. Tatwavedi, U. Schulze, A. Sanchi, S. Ryley, A. Dhir, L. Carpenter, S. M. Watt, D. J. Roberts, A. M. Abdel-Aal, S. K. Sayed, S. A. Mohamed, A. Schuh, P. Vyas, S. Killick, A. G. Kotini, E. P. Papapetrou, D. H. Wiseman, A. Pellagatti, and J. Boultwood. 2020. "SF3B1 mutations induce R-loop accumulation and DNA damage in MDS and leukemia cells with therapeutic implications." Pp. 2525-2530 in *Leukemia*. England.

Smith, C. W., E. B. Porro, J. G. Patton, and B. Nadal-Ginard. 1989. "Scanning from an independently specified branch point defines the 3' splice site of mammalian introns." *Nature* 342(6247):243-7.

Spieth, J., D. Lawson, P. Davis, G. Williams, and K. Howe. 2014. "Overview of gene structure in C. elegans." *WormBook*:1-18.

Staley, J. P., and C. Guthrie. 1999. "An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p." *Mol Cell* 3(1):55-64.

Steitz, T. A., and J. A. Steitz. 1993. "A general two-metal-ion mechanism for catalytic RNA." *Proc Natl Acad Sci U S A* 90(14):6498-502.

Stiernagle, T. 2006. "Maintenance of C. elegans." *WormBook*:1-11.

Suzuki, Jmngl, K. Osterhoudt, C. H. Cartwright-Acar, D. R. Gomez, S. Katzman, and A. M. Zahler. 2022. "A genetic screen in C. elegans reveals roles for KIN17 and PRCC in maintaining 5' splice site identity." *PLoS Genet* 18(2):e1010028.

Tsukamoto, T., M. D. Gearhart, S. Kim, G. Mekonnen, C. A. Spike, and D. Greenstein. 2020. "Insights into the Involvement of Spliceosomal Mutations in Myelodysplastic Disorders from Analysis of SACY-1/DDX41 in Caenorhabditis elegans." *Genetics* 214(4):869-893.

Valadkhan, S., and J. L. Manley. 2001. "Splicing-related catalysis by protein-free snRNAs." *Nature* 413(6857):701-7.

Vijayraghavan, U., M. Company, and J. Abelson. 1989. "Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae." *Genes Dev* 3(8):1206-16.

Weinberg, R. A., and S. Penman. 1968. "Small molecular weight monodisperse nuclear RNA." *J Mol Biol* 38(3):289-304.

Wilkinson, M. E., C. Charenton, and K. Nagai. 2020. "RNA Splicing by the Spliceosome." *Annu Rev Biochem* 89:359-388.

Wu, J., and J. L. Manley. 1989. "Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing." *Genes Dev* 3(10):1553-61.

Yang, F., X. Y. Wang, Z. M. Zhang, J. Pu, Y. J. Fan, J. Zhou, C. C. Query, and Y. Z. Xu. 2013. "Splicing proofreading at 5' splice sites by ATPase Prp28p." *Nucleic Acids Res* 41(8):4660-70.

Zahler, A. M., L. E. Rogel, M. L. Glover, S. Yitiz, J. M. Ragle, and S. Katzman. 2018. "SNRP-27, the C. elegans homolog of the tri-snRNP 27K protein, has a role in 5' splice site positioning in the spliceosome." *Rna* 24(10):1314-1325.

Zahler, A. M., J. D. Tuttle, and A. D. Chisholm. 2004. "Genetic suppression of intronic +1G mutations by compensatory U1 snRNA changes in Caenorhabditis elegans." *Genetics* 167(4):1689-96.

Zhan, X., Y. Lu, X. Zhang, C. Yan, and Y. Shi. 2022. "Mechanism of exon ligation by human spliceosome." *Mol Cell* 82(15):2769-2778.e4.

Zhang, Z., N. Rigo, O. Dybkov, J. B. Fourmann, C. L. Will, V. Kumar, H. Urlaub, H. Stark, and R. Lührmann. 2021. "Structural insights into how Prp5 proofreads the pre-mRNA branch site." *Nature* 596(7871):296-300.

Zhuang, Y., and A. M. Weiner. 1986. "A compensatory base change in U1 snRNA suppresses a 5' splice site mutation." *Cell* 46(6):827-35.

Zorio, D. A., and T. Blumenthal. 1999. "Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans." *Nature* 402(6763):835-8.