UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title

Alternative Splicing of Transcript Isoforms in the Nematode Worm C. elegans

Permalink

https://escholarship.org/uc/item/4b01z4tk

Author

Ragle, James Matthew

Publication Date

2016

Copyright Information

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**Alternative Splicing of Transcript Isoforms in the
Nematode Worm *C. elegans***

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

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELL AND DEVELOPMENTAL BIOLOGY

by

James Matthew Ragle

June 2016

The Dissertation of James Matthew
Ragle is approved:

_____
Professor Alan Zahler, chair

_____
Professor Manuel Ares, Jr.

_____
Professor Melissa Jurica

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

Chapter 2: Coordinated tissue-specific regulation of adjacent alternative 3'

splice sites in *C. elegans*

Chapter 3: Alternative Splicing Regulation of Cancer-Related Pathways in

***Caenorhabditis elegans*: An In Vivo Model System with a Powerful Reverse**

Genetics Toolbox

List of Figures

Abstract

Tissue-Specific Alternative Splicing of Transcripts in the Nematode Worm
C. elegans

by

James Matthew Ragle

Alternative splicing is a form of pre-mRNA processing that modifies gene expression through selective and differential inclusion of genetic material in the mature messenger RNA. This process is dependent upon accurate selection of splice sites in the primary transcript. Site selection has been shown to be differentially regulated between organisms as well as throughout development of tissues within a single organism. In *C. elegans*, we have uncovered over 200 examples of tissue-specific regulation of a class of alternative 3' splice sites, those found within 18 nucleotides of each other. While germline cells prefer to splice at the 3' splice site closest to the 5' end of the intron, somatic cells tend to splice at the site further from the 5' end. These splicing patterns are conserved in the related nematode species *C. briggsae*. Normal *cis*-regulatory features of alternative splice sites are either overlapping in these adjacent sites or are undetermined. Branchpoint selection does not correlate with 3' splice sites chosen in these cases, but total intron length is correlated with upstream 3' splice site usage in the germline, likely due to spatial restrictions on the spliceosome. We also provide evidence that use of the upstream site in somatic cells is correlated with enrichment for pyrimidines at the upstream site and

decreased enrichment at the downstream site, a change we do not see for 3' splice sites that are preferred in the germline. We propose several models to explain this regulation: 1) Differential expression of spliceosome-associated factors in the germline; 2) The spliceosome scans downstream sequences for the first appropriate 3' splice site environment it can detect; 3) Fidelity of splice site choice is relaxed allowing for deviation of site choice from the strictly-regulated somatic splicing pattern.

Dedication

I dedicate this thesis to my wife Ann and to the joyful little animals she blessed

me with: Avery, Alyse and Mason. Without you guys, this and life both mean

nothing to me. To my parents, Joyce and Herb Ragle, for teaching me what it

means to work hard and do whatever I must to support those I love. To my

sister, Angela, for giving me someone special to grow up with and for thinking of

me no matter where we are on the planet. And finally, to myself, for all the

nights I stayed up well past the witching hour finishing assignments, sifting

through data, writing manuscripts and overcoming personal barriers.

This is the culmination of a lifelong dream.

Acknowledgements

I would like to first acknowledge Professor Al Zahler for several things. When I was left in the lurch without a lab to work in, he accepted me. He took me in and showed me what the RNA world was all about while listening to my ideas about the germline. When my project was stuck in the mud and not going anywhere, he let me change direction and follow the data resulting in a publication in a journal in which I previously could never have imagined myself. Mostly, I would like to thank him for understanding that not every graduate student is on the same path to post doc and principlal investigator. The time developing myself as a leader, an educator, a volunteer and a father has been just as valuable as my experiences at the bench and behind the computer screen.

I would like to thank Sergio Barberan-Soler who was a co-writer and mentor and provided resources from half way around the world, Zach Neeb for sharing the same path through grad school with me and experiencing the same stumbles and successes as I did for the most part, Sol Katzman for providing all the knowledge, time and expertise it took to process a lot of data I knew very little about, and to Taylor Akers who was a real friend and provided as many smiles as she did "Grrrrr"s. Thanks to Needhi Bhalla for believing in me to set up her lab and take on her projects before anyone else was there, Bill Saxton for allowing me to cut my teeth as a senior-level undergrad and for recommending me to Needhi, the Graduate Acceptance and Graduate Advisory Committees for bringing me into the program and making sure I was successful, Manny Ares, Melissa Jurica,

Chapter 1: Introduction

1.1 Unique Cellular Traits Are Encoded By A Common Genome

Multicellular biological organisms exhibit a division of labor between cell types. For example, neurons process signals that emanate from or transit to muscles. Bones provide structure to muscles wrapped in epithelia and other connective tissues for protection and maintenance. Likewise, germ cells, in the sex-specific forms of spermatids and oocytes, are set aside for the passage of genetic information from one generation to the next, but are of little to no significance for the health and vitality of the organism as a whole. While the external phenotypes of these cells can be quite varied, the genetic components are nearly identical. Genetic information is stored in the form of DNA sequence and is tightly controlled by regulatory processes such as chromatin packing and transcription factor networks. Much of the variability between cell types is derived from differential expression of particular genes that are present in every cell in an organism.

1.2 Splicing as RNA Processing

Post-transcriptional modifications of newly synthesized RNA molecules contribute greatly to alterations of gene expression and, at times, to cellular diversity. RNA transcripts derived from activated genes can transmit instructions stored within DNA sequence to protein actors that execute them. RNA processing is a series of biochemical reactions that occur in every

1

eukaryotic cell to modify, amplify and regulate the pre-mRNA products of genetic transcription. A major component of RNA processing is pre-mRNA splicing, the selective removal of a region of genetic material, the intron, from a pre-mRNA transcript that results in the juxtaposition of the flanking genetic regions, the exons, to produce a mature mRNA (reviewed in Will and Lührmann 2011). The decision of each cell type concerning which introns to remove and what defines the boundaries of an intron in each cell are central areas of exploration in this thesis.

1.3 The Mechanism of Splicing

The spliceosome is a dynamic complex of proteins and small nuclear RNAs (snRNAs) that forms *de novo* on a pre-mRNA in a step-wise fashion to carry out this removal of genetic information (Brody and Abelson 1985; Grabowski *et al.* 1985; Frendewey and Keller 1985). U1, U2, U4, U5 and U6 are snRNAs that combine with over 100 proteins to make small nuclear ribonucleoproteins (snRNPs) that work cooperatively as the spliceosome to facilitate intron removal (Introduction-Figure 1). The 5' splice site at the leading edge of the intron is initially recognized by base pairing with the U1 snRNA (Zhuang and Weiner 1986, Séraphin and Rosbash 1989). The branchpoint within the intron pairs initially with the branchpoint-binding protein (BBP/SF1; Berglund *et al.* 1997). The polypyrimidine tract and 3' splice site at the lagging edge of the intron are defined by interaction with U2AF$_{65/35}$ subunit proteins (Frendewey and Keller 1985; Merendino *et al.* 1999; Zorio and Blumenthal 1999; Wu *et al.* 2011). This

is the commitment or E complex. U2 snRNP is recruited to the assembling machinery by interactions between BBP/SF1 with the branchpoint and U2AF$_{65/35}$ subunit with the polyY tract and 3' splice site forming the pre-spliceosomal or A complex (Valcarcel *et al.* 1996). These snRNPs bound to their nucleic acid binding partners promote the recruitment of the tri-snRNP containing U4/U6 and U5 to form the B complex. DExD/H-box RNA helicases unwind interactions between snRNAs within the assembling spliceosome (reviewed in Cordin and Beggs 2013). This allows for rearrangement of the spliceosomal snRNPs as the steps of the splicing process proceed. U6 tri-snRNP snRNA dissociates from U4 snRNA, pairing with U2 snRNA and the 5' splice site (Madhani and Guthrie 1992). This causes the displacement of U1 snRNP and expulsion of both U1 and U4 snRNP (Konarska and Sharp 1987), creating, along with the NTC complex, the B activated complex (Wahl et al.. 2009). Nucleophilic attack of the branchpoint adenosine 2'OH on the 5' phosphate of the guanosine at the 5' splice site and separation of the 5' exon creates the C complex. Again, DExD/H-box ATPases work to remodel the spliceosome following branching. The exons are repositioned for ligation and a 2nd nucleophilic attack, this time by the 3'-OH of the 5' exon covalently binds it to the phosphate of the downstream exon at the 3' splice site to make the P complex (reviewed in Liu and Cheng 2015). Splice site specificity is achieved by sequential binding of multiple components of the spliceosome to the correct cis-elements throughout the splicing process. Conversely, the dynamic nature of the spliceosome allows

for much greater flexibility when binding sites within pre-mRNA substrates are

selected.
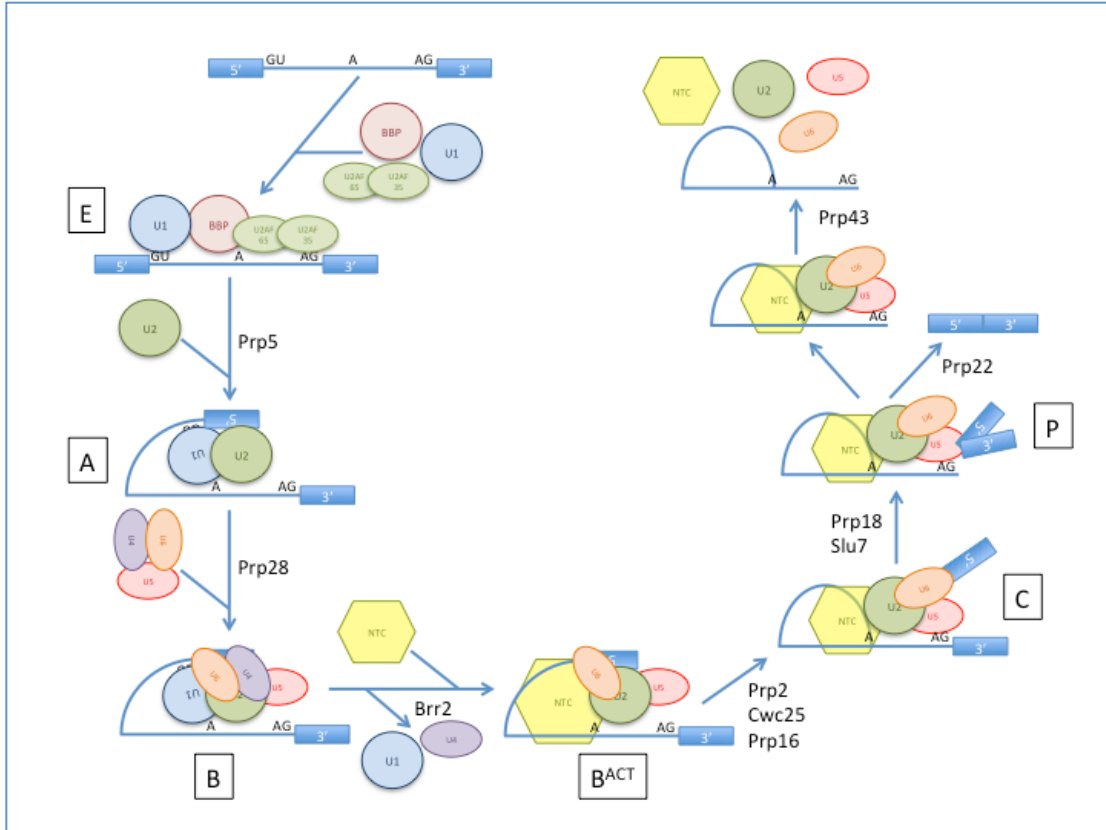
Figure 1. The splicing cycle. *cis*-elements within the intron are bound by prespliceosomal components through basepairing interactions with their associated snRNAs. Rearrangements of the splicing complex and formation of new complexes are facilitated by unwinding of basepairing interactions by helicases and the coming and going of dynamic spliceosomal components at various stages until the exons are ligated together and the lariat/protein complex is dissociated and recycled.

1.4 Alternative Splicing Expands The Proteome In Several Ways

Alternative splicing is a regulated component of pre-mRNA processing and

results in synthesis of multiple protein or ncRNA products transcribed from the

same gene. These products may contain or lack any number of domains,

regulatory sites or structural motifs, resulting in the potential for altered function, localization, reactivity or stability. Organismal complexity in higher eukaryotes, humans especially, was initially hypothesized to be a result of a comparatively large number of genes, but following sequencing of the genomes of many organisms, it became clear that much of the complexity of higher eukaryotes was derived through other means (International Human Genome Sequencing Consortium 2004). The human genome contains roughly 20,000-25,000 genes, on par with the genomes of lower eukaryotes like fruit flies and worms, but over 100,000 different proteins (Modrek and Lee 2002). The classic one gene-one protein model (Beadle and Tatum 1941) gave way to the idea of an expanded human proteome when it was discovered that the complement of diverse mRNAs in human cells, the transcriptome, greatly outnumbered the number of genes at the DNA level. In fact, nearly every gene in the human genome (>90%) produces pre-mRNA transcripts that are alternatively spliced (Wang et al.. 2009). If these alternatively spliced mRNAs are translated, the result is an expanded proteome with altered molecular and cellular behavior, cellular diversity and even organismal complexity (Kim et al.. 2007). Transcript isoforms derived from alternative splicing often differ by expansion or contraction of the 5' or 3' ends of an exon, by one of several possible terminal exons or by inclusion or skipping of an exon or grouping of exons (reviewed in Zahler 2012). In fact, cassette exons, which can be wholly included or skipped, represent the largest class (38.4%) of conserved alternative splicing events in

humans and mice with alternative 3' splice sites and alternative 5' splice sites representing a smaller subset of conserved alternative events (18.4% and 7.9% respectively; Sugnet et al. 2004).

1.5 Factors, Sequences and Mechanism of Alternative Splicing

Excision of alternative introns from transcripts of the same gene require the presence of multiple cis-regulatory elements that serve to identify the boundaries of the alternative introns as well as a complementary set of sequences and protein binding partners that modulate the alternative usage of such elements. SR proteins contain RNA-recognition motifs (RRMs) that bind to sequence-based elements within exons or introns of a pre-mRNA transcript to recruit spliceosomal snRNPs (Zahler *et al.* 1992; Zahler and Roth 1995; reviewed in Bourgeois *et al.* 2004). Conversely, hnRNPs are proteins that bind to intronic or exonic short sequence elements to inhibit formation of transcript-spliceosome interactions (reviewed in Martinez-Contreras *et al.* 2007). For example, binding of hnRNP H to an element near the 3' splice site of HIV tat exon 2 reduces the ability of U2AF to bind there (Jacquenet *et al.* 2001). Both of these classes of proteins compete and cooperate to regulate intron definition when in context with other splicing silencers or enhancers (Barberan-Soler *et al.* 2011; Zhu *et al.* 2001). They can function redundantly to regulate splicing patterns (Huelga *et al.* 2012) and can even regulate the expression of splicing factors themselves through altered splicing to unstable isoforms (Barberan-Soler and Zahler 2008a). Replacement of an hnRNP A1 binding site for an hnRNP H

6

binding site causes no change in splicing patterns of an alternative exon in hnRNP A1 transcript pre-mRNA splicing (Martinez-Contreras *et al.* 2006). In all, a combination of interactions between the transcript and protein as well as between protein interactors that associate with the transcript are required to fine tune the process of splice site selection and intron excision.

1.6 Timing and Fidelity of Splice Site Choice

Given the complex nature of the spliceosome and the balance it must maintain between specificity and flexibility throughout the splicing process, it is no surprise that splice site selection has remained somewhat of an enigma. At 3' splice sites specifically, three sequence elements contribute to site recognition within each intron, the branchpoint, polyY tract and a consensus NAG motif at the 3' splice site itself (N=any nucleotide). Interactions between BBP/SF1 and U2AF help stabilize binding of these factors to their nucleotide counterparts early in spliceosome formation (Berglund *et al.* 1998; Selenko *et al.* 2003). hSlu7 reorganizes the spliceosome following branching to align the AG dinucleotide at the 3' splice site for the 2nd catalytic reaction, nucleophilic attack of the 5' exon and subsequent exon-exon ligation (Chua and Reed 1999ab). Proofreading steps facilitated by Prp22 disruption of stabilizing spliceosome-transcript interactions have been proposed to ensure adherence to proper splice site choice (Mayas *et al.* 2006; Semlow *et al.* 2016). Given all of this, questions still remain concerning the timing and fidelity of splice site choice. It is still unknown how rearrangements of the spliceosome-transcript complex prior to exon-exon

ligation contribute to 3' splice site choice or at precisely what point in the splicing process the 3' splice site is finally determined. Furthermore, it is unclear how splice site choice changes over development and between distinct cell types.

1.7 *C. elegans* Provides a Crossroads to Understand the Correlation of Cellular Differences with Alternative Splicing Patterns

C. elegans represents an intriguing *in vivo* model organism in which to study development, cellular differentiation, genome evolution and alternative splice site choice. The somatic worm is invariably comprised of 959 cells whose developmental lineages have been extensively traced (Sulston and Horvitz 1977). Alternative splicing within tissues constructed by these fully differentiated cells has been studied. For example, roles for alternative isoforms in regulating basic physiological processes and cellular specialization have been elucidated in distinct neuron types (Norris *et al.* 2014). As well, muscle-specific alternative splicing proteins such as ASD-2 and SUP-12 have been shown to control regulators of actin network organization (Ohno *et al.* 2012). But despite extensive effort to explore alternative splicing between *C. elegans* somatic tissues, the differences in alternative splicing between the tissues of the somatic body and the germline have not been identified. This is significant given that each hermaphroditic worm produces 150-300 progeny within a three day span of its normal fourteen day life. The changes in gene expression due to alternative splicing differences between the germline and somatic tissues of *C. elegans* remained unknown until this work. Furthermore, *C. elegans* introns

have a unique set of internal regulatory sequences that complicate contemporary models for how early actors in spliceosome formation identify their pre-mRNA binding partners. The branchpoint consensus sequence motif is unknown or not used as in other organisms and the typically extensive polypyrimidine tract seen in higher eukaryotes is shortened and combined into a highly conserved 3' splice site consensus YYYNAG motif (Y=cytosine or thymine; Blumenthal and Steward 1995). Introns are much smaller in *C.* elegans than in higher eukaryotes with the median intron size in humans being 1334 nucleotides (Hong *et al.* 2006) compared to 67nt in *C. elegans* (Chapter 2) indicating that the components of the worm spliceosome are in closer physical and temporal proximity when assembling at each end of the substrate pre-mRNA intron. Rearrangements during the assembly and activation of the worm spliceosome are likely to be under increased spatial restrictions compared to human spliceosomes in order to accommodate the shorter introns. Lastly, *C. elegans* represents an evolutionary middle ground between lower eukaryotes like *saccharomyces Saccharomyces cerevisiae* that exhibit the requirement for alternative splicing in only a handful of genes (Parenteau *et al.* 2008) and higher eukaryotes like humans that produce alternative isoforms from a large proportion of genes. Roughly 25% of protein-coding genes in *C. elegans* have multiple isoform annotations, with each gene producing 2.2 isoforms on average (Ramani et al. 2011).

1.8 Initial Results and Questions of this Work

There are unique situations in which the organization of intronic regulatory elements and splicing patterns allow for insights into questions about splice site choice and the mechanisms of splicing fidelity. In this work, hundreds of genes were identified that produce transcripts that differ in splicing patterns between somatic and germline cells. Primarily, these differences are not of the most abundant class of alternative splicing events, cassette exons, but of tandem alternative 3' splice sites. These tissue-specific 3' splice sites are separated by 6-18nt and most frequently show maintenance of translational reading frame in both somatic and germline samples. In somatic cells, splicing occurs almost exclusively at the 3' splice site furthest from the 5' end of the intron (the distal site), while germline cells activate the alternative 3' splice site closer to the 5' splice site (the proximal site). We did not observe a single case in which this tissue-specific splice site choice pattern was reversed in this study. Our discovery leads to many interesting questions that we explore in Chapter 2 concerning how and why a cell chooses a 3' splice site: 1) Are there functional differences between isoforms derived from alternative 3' splice sites in such close proximity? 2) Does intron architecture play a role in determining splicing patterns at tandem 3' splice sites? 3) What sequence elements in the intron or at the 3' splice site are necessary for alternative 3' splice site choice?

1.9 Functional Differences Between Alternative Isoforms Created by

Tandem 3' Splice Sites

In 2005, Hiller et al. highlighted a class of motifs that serve as alternative splice acceptors, called NAGNAGs, that have implications for protein function in some cases (Hiller *et al.* 2005). These dual 3' splice sites consist of a tandem triplet of nucleotides, each comprised of any possible nucleotide followed by an obligate AG dinucleotide. Examining mRNAs from Genbank's RefSeq database identified these 3' splice site motifs in 30% of human genes overall with 5% of genes having them in the coding region. Given the juxtaposition of one splice site to the other, differences in the nucleotide sequence of the spliced transcripts change only one or two amino acids in the translated proteins. Nonetheless, these minor changes can be enough to alter the structure of the resulting protein, the charge of a binding pocket or the phosphorylation state of an inducible enzyme. Insulin-like Growth Factor I receptor, for example, produces two isoforms that differ by a CAG nucleotide triplet found in the intron of one isoform and the neighboring downstream exon of the second isoform (Condorelli *et al.* 1994). The receptor protein lacking the CAG exhibited higher activation of downstream effectors, but slower rates of insulin-like growth factor I uptake into cells than the isoforms containing the CAG. In another example, the cystic fibrosis trans-membrane conductance regulator (CFTR) encodes a cAMP-regulated Cl^- ion channel at the apical membrane of epithelial cells. Mutation of CFTR results in disease phenotypes of the gastro-intestinal tract, lungs,

reproductive organs and sweat glands. CFTR contains a NAGNAG motif in a linker region between functional domains. Mutation of the fourth nucleotide in the NAG**N**AG motif from the native guanine to a thymine often results in the downstream exon being skipped and a mild disease phenotype (Hinzpeter *et al.* 2010). This conversion also changes the motif from an obligate proximal splice site, in which splicing occurs constitutively at the site closest to the 5' end of the intron, to an alternative 3' splice site, using both splice acceptors to some degree. While splicing to the upstream 3' splice site causes incorporation of a premature termination codon (PTC) leading to a truncated protein and a severe disease phenotype. This phenotype can be suppressed by alternative splicing to the downstream 3' splice site (distal) which removes the single stop codon and results in a mostly functional protein. Alternative splicing at NAGNAGs, therefore, can have major ramifications for the stability of transcript isoforms as well as their function in humans.

Isoforms created by alternative splicing at tandem 3' splice sites may be similarly functional in *C. elegans*. If *cis*-regulatory elements that direct splice site selection in one species are maintained in related, but divergent species, it would highlight the necessity of those elements to produce multiple isoforms throughout evolution, especially if splicing patterns between distantly-related species at those tandem 3' splice sites are the same. Furthermore, splice sites that result in nonfunctional products would expect to be chosen without preference for reading frame. Out-of-frame splicing results in an entire class of

transcripts that contain downstream premature termination codons, making them potential targets for nonsense-mediated decay (NMD) targeting. Some genes, especially splicing factors, produce isoforms that autoregulate expression through NMD in *C. elegans*, but it is unknown if they carry out other functions (Barberan-Soler and Zahler 2008). Because NMD mutant *C. elegans* are viable, it is possible to test if isoforms from tissue-specific alternative 3' splice sites are stably expressed. Lastly, because the isoforms produced at tandem 3' splice sites differ by only a few nucleotides, potentially functional protein isoforms in the germline would contain only a single to a few added amino acids upon translation. Conservation of those amino acids in related species would provide further evidence of their function.

1.10 Intron Architecture and Splicing Patterns

The architecture of an intron can play a central role in determining splice sites. Vertebrate introns have increased in length over evolutionary time requiring stronger splice sites to mark exon-intron boundaries from within a larger pool of nucleotides (Gelfman *et al.* 2012). Even the method with which intronic borders are determined is altered with intron size. Splice sites in *Drosophila* and humans are recognized across introns until they reach 200-250 nucleotides in length. Beyond this, introns are determined across the exon (Fox-Walsh *et al.* 2005). Similarly, exons preceded by a large intron are more likely to be alternative than those with short leading introns. On the other extreme, introns do not span less than 50 nucleotides on average in most species (Deutsch and

Long 1999), but given that the spliceosome is such a dynamic complex of proteins and ncRNAs, the reason for this minimum is not understood. It is conceivable that regulatory elements that normally work to maintain tissue-specific splicing patterns succumb to limitations such as spatial constraints dictated by the spliceosome and its component parts. Likewise, branch site selection may impact other events during the splicing process. The distance between the branchpoint and 3' splice site has been shown to play a role in 3' splice site selection. Expression of a minigene splicing reporter in HEK293T cells revealed increased usage of a proximal 3' splice site when the distance from the branchpoint to the 3' splice site was increased by 4 or more nucleotides (Bradley *et al.* 2012). With these findings and with evidence that U2AF subunits interact directly with the branchpoint-binding protein (BBP/SF1; Berglund *et al.* 1998), we hypothesize that alternative 3' splice site selection would be correlated with alternative branchpoints.

1.11 Regulatory Sequence Elements that Alter 3' Splice Site Choice

cis-regulatory elements such as the 3' splice site sequence motif, the polyY tract and the branchpoint normally dictate splice site choice, but can be superseded by other factors. Intronic sequence elements that normally direct the spliceosome, such as the branchpoint consensus sequence motif and the polypyrimidine tract, are modified, absent or unknown in *C.* elegans. Short sequences within the intron or downstream exon (5 or 6 nucleotides) have been identified as binding partners for alternative splicing factors such as HRP-2,

SC35 and hnRNP A/B (Zahler *et al.* 2004; Kabat *et al.* 2009; Kabat *et al.* 2006).

These types of factors interact with core spliceosomal components or factors

that influence core splicesomal component behavior to recruit or repel

formation of the spliceosome. Higher sequence conservation has been observed

in introns preceding alternatively-spliced tandem 3' splice site sets when

compared to constitutively-spliced ones (Akerman and Mandel-Gutfreund 2006;

Bradley *et al.* 2012). Site selection at tandem 3' splice sites may be more

dependent upon alternative *cis*-elements and *trans*-factors for constitutive

splicing regulation activities such as branchpoint and/or splice site

identification, but the fact that adjacent 3' splice sites are so tightly juxtaposed

complicates this model. The splice site itself is thought to contain the primary

information necessary for identification by pre-spliceosomal components,

namely the polypyrimidine tract and the AG dinucleotide (C. elegans II, 2nd

edition 1997). Still, evidence exists that supports the idea that the 3' splice site

components are not an absolute necessity for 3' splice site choice. Mutation of

an AG dinucleotide to an AA results in splicing to a cryptic AG as well as to the

AA dinucleotide at the original site, suggesting the short polyY tract upstream

holds sufficient information to direct splice site selection, but not absolutely

(Aroian *et al.* 1993; Zhang and Blumenthal 1996). Tissue-specific differences

may exist concerning the nucleotide content requirements of each splice site and

the regulation of 3' splice site selection, especially when they are found in such

close proximity that alternative splicing factors cannot differentiate between them.

1.12 Transitioning from the 1st Step to the 2nd

This thesis addresses the larger topic of 3' splice site choice and the timing, factors and mechanism involving the rearrangements of the spliceosome as it transitions from branching to determination of exon ligation location. Following branching and 5' exon separation in the first catalytic step of splicing, Prp16 establishes the second step active site by release of Cwc25 from the branchpoint in an ATP-dependent reaction (Tseng *et al.* 2011, Ohrt *et al.* 2013). Prp18 and Slu7 bind to the 3' splice site just prior to exon ligation and dock the 3' splice site in the second step active site (Umen and Guthrie 1995, Ohrt *et al.* 2013). U5 snRNP interacts with Prp18 and Slu7 likely stabilizing their interactions with the splicing intermediate (Horowitz and Abelson 1993, Jiang *et al.* 2000, Bacikova and Horowitz 2002, Bacikova and Horowitz 2005). These and other splicing factors that transition the spliceosome from first step to second are conserved from yeast to worms to humans implying the process is conserved as well (Zhou and Reed 1998), but many splicing factors, second step actors included, have functions in biochemical processes other than splicing in *C. elegans*, such as germline development and the sperm to oocyte switch (Puoti and Kimble 1999). This major difference along with the differences in intronic regulatory motifs between worms and higher eukaryotes highlights the fact that the splicing factors and precise mechanism with which *C. elegans* spliceosomes select a 3'

splice site remain to be worked out. Tissue-specificity at adjacent 3' splice sites may be enacted through differential expression of first step/second step transition factors such as Prp18, Prp16 or Slu7, through altered interactions of these factors with each other or the substrate intermediate or through disruption of proofreading mechanisms. Prp22, for example, uses binding of sequences downstream from the 3' splice site and 3' to 5' translocation along the transcript to dissociate ligating exons when a suboptimal 3' splice site has been chosen during the second catalytic step (Semlow *et al.* 2016). In chapter 2 of this thesis, we explore these ideas propose updated models to explain coordinated regulation of tissue-specific alternative 3' splice sites in *C. elegans*.

1.13 References

Akerman M, Mandel-Gutfreund Y (2006) Alternative splicing regulation at tandem 3' splice sites. Nucleic Acids Res 34: 23–31.

Aroian RV, Levy AD, Koga M, Ohshima Y, Kramer JM, Sternberg PW. Splicing in *Caenorhabditis elegans* does not require an AG at the 3' splice acceptor site. Mol Cell Biol. 1993 Jan;13(1):626-37.

Bacíková D, Horowitz DS. Mutational analysis identifies two separable roles of the Saccharomyces cerevisiae splicing factor Prp18. RNA. 2002 Oct;8(10):1280-93.

Bacíková D, Horowitz DS. Genetic and functional interaction of evolutionarily conserved regions of the Prp18 protein and the U5 snRNA. Mol Cell Biol. 2005 Mar;25(6):2107-16.

Barberan-Soler S, Zahler AM. Alternative splicing regulation during C. elegans development: splicing factors as regulated targets. PLoS Genet. 2008 Feb 29;4(2):e1000001. doi: 10.1371/journal.pgen.1000001.

Barberan-Soler S, Medina P, Estella J, Williams J, Zahler AM. Co-regulation of alternative splicing by diverse splicing factors in Caenorhabditis elegans.

Nucleic Acids Res. 2011 Jan;39(2):666-74. doi: 10.1093/nar/gkq767. Epub 2010 Aug 30.

Beadle, George Wells, and Edward Lawrie Tatum. "Genetic Control of Biochemical Reactions in Neurospora." Proceedings of the National Academy of Sciences 27 (1941): 499–505.

Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell 89(5): 781-787.

Berglund JA, Abovich N, Rosbash M. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes Dev 12: 858–867.

Blumenthal T, Steward K. 1997. RNA processing and gene structure. In C. elegans II (ed. Riddle DL, et al..), pp. 117–145. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bourgeois CF, Lejeune F, Stevenin J. Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. Prog Nucleic Acid Res Mol Biol, 78 (2004), pp. 37–88

Bradley RK, Merkin J, Lambert NJ, Burge CB. Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. PLoS Biol. 2012 Jan;10(1):e1001229. doi: 10.1371/journal.pbio.1001229. Epub 2012 Jan 3.

Brody E, Abelson J. The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science. 1985;228:963–7.

C. elegans II, 2nd edition. Editors: Donald L Riddle, Thomas Blumenthal, Barbara J Meyer, and James R Priess. Cold Spring Harbor Mongraph Series, Vol. 33. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997.

Chua K, Reed R. 1999a. Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. Genes Dev 13: 841–850.

Chua K, Reed R. 1999b. The RNA splicing factor hSlu7 is required for correct 3′ splice-site choice. Nature 402: 207–210.

Condorelli G, Bueno R, Smith RJ. Two alternatively spliced forms of the human insulin-like growth factor I receptor have distinct biological activities and internalization kinetics. J Biol Chem. 1994 Mar 18;269(11):8510-6.

Cordin O, Beggs JD. RNA helicases in splicing. RNA Biol. 2013 Jan;10(1):83-95. doi: 10.4161/rna.22547. Epub 2012 Dec 10.

Deutsch M, Long M. Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res. 1999 Aug 1;27(15):3219-28.

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A. 2005 Nov 8;102(45):16176-81. Epub 2005 Oct 31.

Frendewey D, Keller W. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell. 1985;42:355–67.

Gelfman S, Burstein D, Penn O, Savchenko A, Amit M, Schwartz S, Pupko T, Ast G. Changes in exon-intron structure during vertebrate evolution affect the splicing pattern of exons. Genome Res. 2012 Jan;22(1):35-50. doi: 10.1101/gr.119834.110. Epub 2011 Oct 5.

Grabowski PJ, Seiler SR, Sharp PA. A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell. 1985 Aug;42(1):345-53.

Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, Schreiber S, Backofen R, Platzer M. Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. Nat Genet. 2004 Dec;36(12):1255-7. Epub 2004 Oct 31. Erratum in: Nat Genet. 2005 Jan;37(1):106.

Hinzpeter A, Aissat A, Sondo E, Costa C, Arous N, Gameiro C, Martin N, Tarze A, Weiss L, de Becdelièvre A, Costes B, Goossens M, Galietta LJ, Girodon E, Fanen P. Alternative splicing at a NAGNAG acceptor site as a novel phenotype modifier. PLoS Genet. 2010 Oct 7;6(10).

Hong X, Scofield DG, Lynch M. Intron size, abundance, and distribution within untranslated regions of genes. Mol Biol Evol. 2006 Dec;23(12):2392-404. Epub 2006 Sep 15.

Horowitz DS, Abelson J. A U5 small nuclear ribonucleoprotein particle protein involved only in the second step of pre-mRNA splicing in Saccharomyces cerevisiae. Mol Cell Biol. 1993 May;13(5):2959-70.

Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, et al.. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep. 1:167–78.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004 Oct 21;431(7011):931-45.

Jacquenet S, Mereau A, Bilodeau PS et al.. A second exon splicing silencer within human immunodefciency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. J Biol Chem 2001; 276(44):40464-40475.

Jiang J, Horowitz DS, Xu RM. Crystal structure of the functional domain of the splicing factor Prp18. Proc Natl Acad Sci U S A. 2000 Mar 28;97(7):3022-7.

Kabat JL, Barberan-Soler S, McKenna P, Clawson H, Farrer T, Zahler AM. Intronic alternative splicing regulators identified by comparative genomics in nematodes. PLoS Comput Biol. 2006 Jul 14;2(7):e86. Epub 2006 Jun 5.

Kabat JL, Barberan-Soler S, Zahler AM. HRP-2, the Caenorhabditis elegans homolog of mammalian heterogeneous nuclear ribonucleoproteins Q and R, is an alternative splicing factor that binds to UCUAUC splicing regulatory elements. J Biol Chem. 2009 Oct 16;284(42):28490-7. doi: 10.1074/jbc.M109.023101. Epub 2009 Aug 24.

Kim E, Magen A, Ast G. 2007. Different levels of alternative splicing among eukaryotes. Nucleic Acids Res. 35:125–31.

Konarska MM, Sharp PA. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. Cell, 49 (1987), pp. 763–774

Liu YC, Cheng SC. Functional roles of DExD/H-box RNA helicases in Pre-mRNA splicing. J Biomed Sci. 2015 Jul 16;22:54. doi: 10.1186/s12929-015-0161-z.

Madhani HD, Guthrie C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell, 71 (1992), pp. 803–817.

Martinez-Contreras R, Fisette JF, Nasim FU et al.. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. PLoS Biol 2006; 4(2):e21.

Martinez-Contreras R, Cloutier P, Shkreta L, Fisette JF, Revil T, Chabot B. 2007. hnRNP proteins and splicing control. Adv. Exp. Med. Biol. 623:123–47.

Mayas RM, Maita H, Staley JP. 2006. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol 13: 482–490.

Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J. 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. Nature 402(6763): 838-841.

Modrek, B. & Lee, C. A genomic view of alternative splicing. Nature Genet. 30, 13–19 (2002).

Norris AD, Gao S, Norris ML, Ray D, Ramani AK, Fraser AG, Morris Q, Hughes TR, Zhen M, Calarco JA. A pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. Mol Cell. 2014 Jun 19;54(6):946-59. Epub 2014 Jun 5.

Ohno G, Ono K, Togo M, Watanabe Y, Ono S, Hagiwara M, Kuroyanagi H. Muscle-specific splicing factors ASD-2 and SUP-12 cooperatively switch alternative pre-mRNA processing patterns of the ADF/cofilin gene in Caenorhabditis elegans. PLoS Genet. 2012;8(10):e1002991. doi: 10.1371/journal.pgen.1002991. Epub 2012 Oct 11.

Ohrt T, Odenwälder P, Dannenberg J, Prior M, Warkocki Z, Schmitzová J, Karaduman R, Gregor I, Enderlein J, Fabrizio P, Lührmann R. Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. RNA. 2013 Jul;19(7):902-15.

Parenteau J, Durand M, Véronneau S, Lacombe A, Morin G, Guérin V, Cecez B, Gervais-Bird J, Koh C, Brunelle D, Wellinger RJ, Chabot B, Elela SA. Deletion of Many Yeast Introns Reveals a Minority of Genes that Require Splicing for Function. Mol Biol Cell. 2008 May; 19(5): 1932–1941. doi: 10.1091/mbc.E07-12-1254.

Puoti A, Kimble J. The Caenorhabditis elegans sex determination gene mog-1 encodes a member of the DEAH-Box protein family. Mol Cell Biol. 1999 Mar;19(3):2189-97.

Ramani AK, Calarco JA, Pan Q, Mavandadi S, Wang Y, Nelson AC, Lee LJ, Morris Q, Blencowe BJ, Zhen M, Fraser AG. Genome-wide analysis of alternative splicing in Caenorhabditis elegans. Genome Res. 2011 Feb;21(2):342-8.

Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Krämer A, Sattler M. 2003. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol Cell 11: 965–976.

Semlow DR, Blanco MR, Walter NG, Staley JP. Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. Cell. 2016 Feb 25;164(5):985-98.

Séraphin B, Rosbash M. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell, 59 (1989), pp. 349–358.

Sugnet CW, Kent WJ, Ares M Jr, Haussler D. Transcriptome and genome conservation of alternative splicing events in humans and mice. Pac Symp Biocomput. 2004:66-77.

Sulston JE, Horvitz HR. 1977. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol* **56**(1): 110-156.

Tseng CK, Liu HL, Cheng SC. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA. 2011 Jan;17(1):145-54. doi: 10.1261/rna.2459611. Epub 2010 Nov 22.

Umen JG, Guthrie C. Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. RNA. 1995 Aug;1(6):584-97.

Valcarcel J, Gaur RK, Singh R, Green MR. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. Science, 273 (1996), pp. 1706–1709.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP & Burge CB. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470-476 (27 November 2008).

Will CL and Lührmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011;3:0003707.

Zahler AM, Lane WS, Stolk JA, Roth MB. SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev. 1992 May;6(5):837-47.

Zahler AM, Roth MB. Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. Proc Natl Acad Sci U S A. 1995 Mar 28;92(7):2642-6.

Zahler AM, Damgaard CK, Kjems J, Caputi M. J . SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. J. Biol Chem. 2004 Mar 12;279(11):10077-84. Epub 2003 Dec 31.

Zahler AM. Pre-mRNA splicing and its regulation in Caenorhabditis elegans. WormBook. 2012 Mar 21:1-21.

Zhang H, Blumenthal T. Functional analysis of an intron 3' splice site in *Caenorhabditis elegans*. RNA. 1996 Apr;2(4):380-8.

Zhou Z, Reed R. Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. EMBO J. 1998 Apr 1;17(7):2095-106.

Zhu, J., Mayeda, A. & Krainer, A. R. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol. Cell 8, 1351–1361 (2001).

Zhuang Y, Weiner AM. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell 46: 827–835.

Zorio DA, Blumenthal T. 1999. Both subunits of U2AF recognize the 3' splice site in C. elegans. Nature 402: 835-838.

Chapter 2: Coordinated tissue-specific regulation of adjacent alternative 3' splice sites in *C. elegans*

2.1 INTRODUCTION

Alternative splicing is a highly regulated process by which a cell can produce multiple messenger RNAs, potentially encoding multiple proteins, from a common precursor transcript. The *de novo* assembly of a spliceosome on each intron of a pre-mRNA transcript requires *cis*-elements within the intron that are recognized as signals marking the beginning, end and branchpoint (Ares and Weiser 1995). A cassette exon is a form of alternative splicing in which an exon is either included or skipped in the mature mRNA. Conserved enhancer or silencer elements within the exon or the surrounding introns interact with an array of constitutive and tissue-specific *trans*-factors that promote or inhibit assembly of a functional spliceosome (Wang and Burge 2008). The use of alternative 3' or 5' splice sites modifies gene expression by including or skipping coding sequences at the ends of exons. Many examples of adjacent alternative 3' splice sites, defined as being separated by 18nt or less, have been observed. In many species, a form of alternative 3' splice site usage has been identified in which the alternative splicing acceptors are only 3nt apart. Except for the rare example of AC dinucleotides observed for substrates of the minor spliceosome, introns end with AG dinucleotides. Alternative 3' splice sites separated by only 3nt are referred to as NAGNAGs, as the end of the intron consists of two AG splice acceptors separated by 3nt. A recent report provided strong evidence that

NAGNAG alternative splicing can be regulated in a tissue-specific manner in mammals (Bradley et al.. 2012). The close proximity of these sites to each other makes the influence of enhancer or silencer elements on splicing to one site or the other, as with cassette exons, unlikely. Therefore, NAGNAGs provide an interesting model in which to understand more about the regulation of alternative splicing and more specifically the mechanisms by which a 3' splice site is chosen.

3' splice sites are determined by a combinatorial code that consists of nucleotides found within the site itself as well as within the intron that precedes it. Typical 3' splice sites consist of an AG dinucleotide and a stretch of upstream, intronic pyrimidines that bind to $U2AF_{35}$ and $U2AF_{65,}$ respectively (Merendino et al.. 1999; Zorio and Blumenthal 1999; Wu et al.. 2011). This complex associates with SF1/BBP, which binds to a conserved sequence surrounding an intronic adenosine nucleotide as the branchpoint for that intron (Berglund et al.. 1997). Together, SF1 and the $U2AF_{35}$/$U2AF_{65}$ complex recruit the U2 snRNP to the branchpoint region and promote spliceosome assembly (Zhang et al.. 1992; Berglund et al.. 1998). The scanning model proposes that the 3' splice site is determined by identification of the first AG dinucleotide downstream from the branchpoint, irrespective of sequence context or distance from the branchpoint (Smith et al.. 1989). A modified scanning model allows for some variability in 3' splice site choice in a small range of nucleotides within the reach of the spliceosome (Smith et al.. 1993), but still suggests the mechanism of 3' splice site

choice, even at NAGNAGs, is stochastic and common to every tissue. Refuting this model, a global analysis of gene expression in 16 human and 8 mouse tissues revealed that splicing patterns at individual NAGNAGs are tissue-specific, regulated and conserved (Bradley et al.. 2012). This work highlighted splicing patterns at adjacent 3' splice sites that suggest a role for the polypyrimidine tract, the location of the branchpoint and the identity of the N within each NAGNAG in designating a site as the end of an intron (Bradley et al.. 2012). It is still unclear, though, if these contributions apply equally to splicing patterns at adjacent 3' splice sites in every cell type and to what extent these factors influence splice site choice in other organisms.

C. elegans is an excellent model organism in which to study developmental and tissue-specific alternative splicing. The developmental lineages of all 959 somatic cells have been traced (Sulston and Horvitz 1977) and the *C. elegans* genome was the first among animals to be fully sequenced (The *C. elegans* Sequencing Consortium 1998). Intergenic regions as well as typical introns are relatively small (Spieth and Lawson 2006). Roughly 25% of protein-coding genes in *C. elegans* have multiple isoform annotations, with each gene producing 2.2 isoforms on average (Ramani et al.. 2011). In addition, many *C. elegans trans*-acting alternative splicing factors and their targeted *cis*-elements are conserved in mammals, making these worms an excellent model organism to study human diseases related to splicing factors (Kabat et al.. 2006; Zahler 2012; Barberan-Soler and Ragle 2013). The 5' splice site consensus sequence in *C. elegans* is

similar to other eukaryotes. In nematodes, introns end in AG and are preceded by a short polypyrimidine tract yielding the consensus sequence YYYNAG/R (Blumenthal and Steward 1997). This short polypyrimidine tract differs from other animals in that it is much shorter and located closer to the AG dinucleotide. The YYYNAG/R consensus sequence is a direct binding site for the *C. elegans* U2AF subunits UAF-1 and UAF-2 (Zorio and Blumenthal 1999; Hollins et al.. 2005). The phenomenon of NAGNAG alternative 3' splice sites separated by 3nt is rare in *C. elegans*, presumably due to the strong nematode 3' splice site consensus with its adjacent pyrimidines precluding a 3nt separation of two AG dinucleotide splice acceptors. A recent example of tissue-specific use of adjacent alternative 3' splice sites separated by 9nt has been uncovered in the *C. elegans* *let-363* gene (Barberan-Soler et al.. 2014). The splice acceptor site closer to the 5' end of the intron (proximal) is favored in the germline while the site further from the 5' end (distal) is preferred in somatic tissue. Regulation of this splicing in the male germline was shown to be dependent on antisense transcription of a gene located within an intron of *let-363*, which is controlled by the piRNA pathway. The extent to which adjacent 3' splice sites in *C. elegans* may be utilized, conserved or regulated on a tissue-specific level is unknown. Further study of these adjacent 3' splice sites may contribute to our understanding of the mechanisms of 3' splice site choice by the spliceosome and spliceosome-associated factors. In this study we uncover and characterize tissue-specific use

of 203 alternative adjacent 3' splice sites. In all cases, there is tissue-specific usage of the proximal splice site in the germline relative to somatic tissue.

2.2 RESULTS

Alternative splicing regulation of *top-1* and establishing a method for detecting germline-specific splicing.

We set out to study the regulation of alternative splicing in *C. elegans* in a specific tissue, the germline, which develops fully during the transition to adulthood. Previous studies in *C. elegans* have focused on the developmental timing and regulation of alternative splicing (Kuroyanagi et al.. 2000; Barberan-Soler and Zahler 2008; Barberan-Soler et al.. 2009; Ramani et al.. 2011). For example, *top-1* alternative cassette exon splicing patterns change during development (Lee et al.. 1998) and RNAi knockdown of the developmentally-regulated splicing factor gene *hrpf-1* leads to a change in cassette exon inclusion in *top-1* isoforms (Barberan-Soler and Zahler 2008). A recent study used immunostaining with antipeptide antibodies specific to the alternative cassette exon of *top-1* to demonstrate detection of the skipping isoform of *top-1* in almost all cells, and the inclusion isoform of *top-1* in neuronal cells, excretory cells, and the germline (Cha et al.., 2012). We set out to test whether alternative splicing of *top-1* pre-mRNA changes with the onset of development of the mature gonad, which corresponds with developmental expansion of the germline. Total RNA was extracted from wildtype L3 and L4 larval stage worms prior to the establishment of a mature germline, as well as wildtype young adult worms following

28

establishment of the mature germline. Reverse transcription with random hexamers followed by polymerase chain reaction (RT-PCR) with primers that anneal to sequences in exons that flank the alternative cassette exon reveals a shift to inclusion of the *top-1* cassette exon upon development of a mature germline (Figure 1A). Because RNA was extracted from whole worms, inclusion of the *top-1* alternative exon in mature adults could be a result of overall changes in alternative splicing in the worm as it ages or, more specifically, to growth and maturity of a new organ, the gonad, containing an expanded germline.

To address whether changes in *top-1* alternative splicing in adults were due to tissue-specific splicing in the germline, young adult worms 20-24 hours past L4 were dissected to isolate gonads and heads. Total RNA was extracted from these tissues as well as from whole worms from the same developmental stage. In addition, RNA was extracted from *glp-4(bn2ts)* mutant young adult worms. These *glp-4* mutant adults contain somatic cells that form a sheath surrounding the gonad, but their germlines are developmentally restricted to just 12 cells stalled in prophase of mitosis when grown at restrictive temperature (Beanan and Strome 1992). This strain has been used extensively as a tool to identify changes in gene expression between germline and somatic tissues (Roussell and Bennett 1993; Shim 1999; Aoki et al.. 2000; Higashitani et al.. 2000; TeKippe and Aballay 2010). RT-PCR revealed a dramatic shift in *top-1* pre-mRNA splicing from complete inclusion of the alternative exon in dissected gonads to partial inclusion in whole worms to almost complete skipping in dissected heads and in

glp-4(bn2ts) worms grown at restrictive temperature (Figure 1B). These data suggest *top-1* cassette exon alternative splicing is strongly subject to tissue-specific regulation. This experiment also demonstrates a useful approach toward identifying germline-specific splicing events through comparison of alternative splicing between isolated gonad and whole *glp-4(bn2ts)* worms that lack an expanded germline.

Genome-wide identification of tissue-specific alternative splicing events.

To identify tissue-specific alternative splicing events in the germline, we used high throughput RNA sequencing and bioinformatics analysis of the resulting data. Total RNA was isolated from dissected wildtype gonads, as well as wildtype whole worms and *glp-4(bn2ts)* mutant whole worms 24 hours past L4 stage (Figure 2A). cDNA libraries were prepared and high-throughput sequencing was performed to obtain strand-specific, 50bp paired-end reads. These reads were mapped to the *C. elegans* genome (ce6) with TopHat (Trapnell et al.. 2009) and splicing changes were identified using SpliceTrap (Wu et al.. 2011), which utilizes paired-end reads to quantify inclusion levels in alternative splicing events. We detected 23 events involving cassette exons with a minimum change in inclusion ratio (IR) of 0.3 between dissected gonad and *glp-4* samples (Supplemental Table 1). The majority of alternative cassette exons we identified (19/23) are highly expressed in the *glp-4* mutant, with some cassette exons nearly undetectable in gonad samples (Figure 2B and 2C). Conversely, cassette exons from only a few genes (4/23) are included more often in RNA isolated

from gonads when compared to *glp-4* mutant RNA (Figure 2D). Additionally, one gene contains an intron inclusion event enriched in RNA isolated from *glp-4* adults that is only slightly detectable in wildtype dissected gonads.

A shift from distal to proximal alternative 3' splice site usage between the soma and the germline.

Surprisingly, the majority of changes in alternative splicing we identified between germline-enriched and somatic-enriched samples were at adjacent alternative 3' splice sites. Using SpliceTrap, we found 65 splice junctions in which adjacent 3' splice sites (≤18nt apart) are used in a tissue-specific manner (Supplemental Table 2). The threshold for tissue-specificity was determined by identifying known adjacent 3' splice sites with a minimum inclusion ratio difference of 0.3 between *glp-4* and dissected wildtype gonads and a minimum of 15 junction spanning reads in both libraries. In all of these 65 cases of adjacent alternative 3' splice sites with tissue-specific usage, we observed a striking correlation: alternative splicing shifts from the distal 3' splice site (furthest from the 5' splice site) in the somatic (*glp-4* whole worm) library toward a proximal 3' splice site (closer to the 5' splice site) in the germline library (Figure 3A). As expected, RNA derived from wildtype whole worms, which contain both germline and somatic tissues, reveals an intermediate splicing pattern that uses a combination of proximal and distal 3' splice sites. To verify the tissue-specific splicing patterns at 3' splice sites seen in RNA-seq data, we isolated RNA from *glp-4* adult whole worms, wildtype adult dissected heads, wildtype adult whole

31

worms and wildtype adult dissected gonads. We made cDNA via reverse transcription with random hexamers and then PCR-amplified with ^{32}P-radiolabelled gene-specific primers that bind to sequences in exons flanking the alternative 3' splice sites (Figure 3B). The products were run on denaturing polyacrylamide gels and visualized with a PhosphorImager to verify the tissue-specific splicing at adjacent 3' splice sites. Consistent with the RNA-seq data, *glp-4* adult whole worm and wildtype adult dissected head samples showed a strong preference for the distal 3' splice site. Wildtype adult whole worms produced isoforms that use a combination of proximal and distal 3' splice sites and wildtype adult germlines overwhelmingly produce isoforms that utilize the proximal 3' splice site. Figure 3C shows the sequences of 15 representative adjacent alternative 3' splice sites that show a shift to usage of the proximal site in the germline. Unlike NAGNAGs seen in mammals, in which the two adjacent 3' splice sites are separated by 3nt, the adjacent 3' splice sites in C. *elegans* are separated by a short stretch of nucleotides enriched for pyrimidines (NAGYYYNAGs; Figure 3C). This is not surprising, given that nematodes differ from other eukaryotes in that they have a short intronic polypyrimidine tract immediately adjacent to the 3' splice site with the consensus sequence YYYNAG (Aroian et al.. 1993; Zhang and Blumenthal 1996). This requirement for immediately adjacent pyrimidines may preclude the ability of NAGNAG adjacent splice sites that are common in mammals to also function in worms.

Identification of additional introns with tissue-specific 3' splice sites in the germline.

Because current genome annotations of alternative splicing are constantly evolving towards completion, we assumed that our identification of tissue-specific alternative 3' splice sites using the SpliceTrap program, which depends on pre-annotation of alternative splicing, would be incomplete. Analysis that does not rely upon pre-annotation would allow us to identify novel tissue-specific splicing events from our sequencing data. When we measure alternative splicing by high-throughput sequencing or RT-PCR, we are measuring the steady-state levels of the alternative isoforms, which are influenced both by the splicing machinery and the relative stability of the different isoforms. Alternative isoforms with differing stabilities are seen in the phenomenon of alternative splicing coupled to nonsense-mediated decay (AS-NMD) (Hamid and Makeyev 2014). NMD is triggered when a message contains a premature termination codon (PTC), and in AS-NMD a change in reading frame at the alternative splice junction generates a PTC. In *C. elegans*, unlike in other animals, worms bearing loss-of-function mutations in genes essential for the NMD pathway are viable (Hodgkin et al.. 1989; Pulak and Anderson 1993), and this system offers us an opportunity to look for alternative adjacent 3' splice sites that are not a multiple of 3nt apart and would otherwise be destabilized by NMD because they disrupt reading frame. To understand the extent of adjacent tissue-specific 3' splice sites genome-wide, and to understand if the NMD pathway is specifically involved in

regulation of out-of-frame isoforms produced by adjacent 3' splice sites, we dissected gonads from adult *smg-2(e2008)* worms defective in NMD (*smg-2* is the *C. elegans* homolog of the essential NMD factor *UPF1*). We extracted total RNA and prepared cDNA libraries for high-throughput sequencing. We reasoned that *smg-2* germline RNA would be enriched in NMD-target RNAs produced by alternative splicing (these would be preferentially degraded in wildtype germ cells), and by analyzing this sample we would maximize our chances of discovery of the different mRNA products produced by the spliceosome. RNA-seq reads were mapped to the *C. elegans* genome with TopHat (Trapnell et al.. 2009) and introns with a common 5' splice site but divergent 3' splice sites ≤18nt apart were identified (Figure 4A).

In analyzing *smg-2* mutant RNA-seq data, we uncovered 487 unique introns with evidence of alternative 3' splice sites ≤18nt from each other (Supplemental Table 3). Of these alternative introns, 315 (64.7%) produced multiple isoforms in the same translational frame, while 172 (35.3%) resulted in alternative isoforms with a PTC-causing frameshift (Figure 4B). This is consistent with previous studies that estimated roughly 1/3 of alternative splicing events in *C. elegans* produce a PTC-containing isoform (Barberan-Soler et al.. 2009). Cross-referencing these adjacent 3' splice site introns detected in this *smg-2* mutant germline analysis with splicing patterns in wildtype germline and *glp-4* mutant whole worms, we identified 58 out of the 65 tissue-specific adjacent 3' splice sites previously identified in the SpliceTrap analysis. In addition, we found 118

more introns that exhibit tissue-specificity at adjacent 3' splice sites using the same expression and splicing change thresholds as the previous analysis (Supplemental Table 2). 53/118 of these had no previously annotated alternative splicing and these represent novel isoforms. We expect this to be an underestimation of alternative adjacent 3' splice sites due to the requirement in our analysis that at least one read be detectable for each isoform in the *smg-2* germline RNA-seq library; this requirement may have excluded tissue-specific alternative 3' splice sites in which all reads that map to an intron in this library cross only the proximal splice site.

To ensure that it is, in fact, the presence of a mature germline that is leading to the production of proximal isoforms and not the mere lack of NMD, RNA-seq data from NMD-defective L1 (somatic tissue-enriched) larval worms from Kuroyanagi et al. (2013) were analyzed and compared to our NMD-deficient gonad sample data (Kuroyanagi et al.. 2013). L1 *smg-2* mutant *s*plicing patterns for all 118 introns identified in our *smg*-2 gonad analysis mirrored that of *glp-4* whole worms with the distal AG dinucleotide as the dominant 3' splice site. Furthermore, a reciprocal analysis of common 5' splice sites with alternative 3' splice sites ≤18nt apart in the *smg-2* mutant L1 worm library failed to identify a single intron with adjacent 3' splice sites that differed significantly in splicing pattern from *glp-4*. This suggests that it is an innate characteristic of the tissue, and not simply the lack of NMD, that contributes to the production and/or stability of the proximal isoforms in the *smg-2* mutant animals.

Nonsense-mediated decay regulation of out-of-frame isoforms produced by adjacent 3' splice sites.

Many of the tissue-specific out-of-frame proximal 3' splice site isoforms expressed in wildtype germline are found in the terminal exon (Supplemental Table 2), a situation not predicted to elicit an NMD response (Nagy and Maquat 1998). To determine if NMD regulation plays an extensive role in the degradation of germline-enriched proximal isoforms, we compared the relative expression and proximal 3' splice site usage of all 487 *smg*-2 adjacent 3' splice site introns to wildtype germline sequencing data (Figure 4B). We note that a higher percentage of introns with out-of-frame proximal 3' splice sites either drop below overall expression thresholds (minimum 10 junction-spanning reads) or decrease usage of the proximal 3' splice site (minimum 0.3 PSI change) when compared to introns with in-frame proximal 3' splice sites, suggesting they are regulated by NMD. Figure 4D shows an example of this phenomenon with alternative adjacent 3' splice sites separated by 7nt in an intron of the *trxr-1* gene. We conclude that even in the absence of NMD, the majority of alternative adjacent 3' splice sites (54.0%) are found to be 6nt or 9nt from each other (Figure 4C). Interestingly, out-of-frame adjacent 3' splice sites are most commonly 7nt or 8nt apart, indicating that alternative adjacent 3' splice sites are separated within a preferred range of 6 to 9nt.

Proximal alternative 3' splice site usage is favored in the germline.

To better understand the characteristics of introns capable of tissue-specific alternative adjacent 3' splice site usage in C. *elegans,* we further explored introns with an AG dinucleotide 6nt upstream of a known 3' splice site (NAGYYYNAGs). We used the UCSC Genome Browser Table Browser tool (Karolchik et al.. 2004) to identify 106,891 annotated introns that terminate in an AG dinucleotide. From that set we then identified 1880 introns containing an AG dinucleotide 6nt upstream from the terminal AG. We looked by hand at 1245 (66.2%) of these introns (Supplemental Table 4) to determine if these represent a rich source of adjacent alternative 3' splice sites. By requiring a minimum expression threshold of 10 junction spanning reads for an intron in both the wildtype germline and *glp-4* whole worm libraries, we identified 192 introns for further study (Figure 5A). Tissue-specific variation of proximal AG usage is evident when we arranged the 192 introns with sufficient expression in this study (>10 junction-spanning reads) according to increasing percentage of proximal 3' splice site usage in wildtype germline from 0% to 100% (Figure 5B, black line) and then identified the corresponding percentage of proximal 3' splice site usage for each intron in the somatic *glp-4* library (Figure 5B, grey line). Strikingly, 120 (62.5%) of these introns showed >10% usage of the upstream AG dinucleotide in either the soma or the germline. This indicates that the presence of an AG dinucleotide 6nt upstream of the distal 3' splice site, in a gene that is well expressed in germline, is a strong indicator of potential adjacent alternative 3' splice site usage. Conversely, of the 192 introns with sufficient expression, 72 used the proximal 3'

37

splice site minimally or not at all (<10% of junction-spanning reads) in *glp-4* as expected, but also in wildtype germline. This is despite the presence of an AG dinucleotide 6nt upstream from the distal 3' splice site.

Of the 120/192 with >10% proximal 3' splice site usage, 80 showed tissue-specificity (minimum 0.3 PSI change), with proximal isoform expression in wildtype germline and distal isoform expression in *glp-4*. 60 of these were previously identified in this study through SpliceTrap or *smg-2* germline analysis, but 20 were novel discoveries of tissue-specific alternative 3' splice sites. Using three methods (SpliceTrap, *smg-2* gonad analysis and this 6nt Shift analysis), we have identified 203 alternative adjacent 3' splice sites in total that show tissue-specific alternative splicing (Supplemental Table S2), all with a shift to the proximal splice site in the germline. It is important to note that for only a few cases did we detect proximal 3' splice site usage in *glp-4* roughly equivalent to the level of usage we observed in wildtype germline. In fact, we have yet to identify a single intron in which there is a significant splicing change in the opposite direction (30% higher proximal splice site usage in *glp-4* whole worm samples than in wildtype germline).

Distance between the 5' and 3' splice sites correlates with the ability to detect splicing to the proximal 3' splice site in the germline.

In order to better understand why some introns with an AG dinucleotide 6nt upstream of the annotated 3' splice site do not splice to the proximal site in the germline while others do, we analyzed the lengths of introns in these two

classes. *C. elegans* introns tend to be smaller on average than mammalian introns, though the factors that compromise the spliceosome tend to be well conserved. We compared the median length of introns that contain an AG dinucleotide 6nt upstream that are either used or not used as a 3' splice site in the germline (as previously defined in figure 5A). The class of introns that do not allow for proximal 3' splice site usage in the germline have a median length of 49nt compared to a median length of 95nt for introns with proximal 3' splice site usage and a median length of 67nt for 108,604 *C. elegans* introns identified in the UCSC Genome Browser (Figure 6). This suggests that intron length may have influence over 3' splice site choice, particularly in the context of adjacent 3' splice sites. This is consistent with a model in which introns below a threshold length may not allow for splicing to a 6nt proximal AG dinucleotide, even when expressed in the germline.

Nucleotide content requirements for 3' splice site selection differ between germline and somatic tissues.

A large number of introns with an AG dinucleotide 6nt upstream of the splice acceptor show adjacent alternative 3' splice site usage in a tissue-specific manner. To understand the requirement for specific nucleotides in identifying a 3' splice site in the context of these tissue-specific alternative adjacent 3' splicing events, multiple sequence alignments of various intron classes (Supplemental Table 5) from the 3' splice site to 40 nucleotides upstream were analyzed through the on-line WebLogo program (Crooks et al.. 2004). A random set of

typical introns revealed the previously identified *C. elegans* 3' splice site

consensus motif TTTCAG (Zhang and Blumenthal 1996), while the consensus

motif at tissue-specific alternative adjacent 3' splice sites separated by 6nt added

only an AG dinucleotide consensus immediately upstream from the typical splice

site motif (Figures 7A and 7B, respectively). Surprisingly, the short stretch of

pyrimidines that accompany the AG dinucleotide in the typical 3' splice site motif

are absent in the region upstream from the proximal AG dinucleotide preferred

in germline. Similarly, in consensus sequences from introns with 9nt or 12nt

between tissue-specific alternative adjacent 3' splice sites, the proximal AG

dinucleotide is pushed 3nt and 6nt further upstream, respectively, with no

further accompanying sequence consensus (Figure 7C and 7D). Consensus

sequence alignment motifs of introns in which the AG dinucleotide at the

proximal 3' splice site is minimally or not at all used in wildtype germline

(Figure 7E) do not reveal any major differences in nucleotide content when

compared to introns in which usage of the proximal AG is enriched in wildtype

germline (Figure 7B). This suggests that the proximal AG dinucleotide is not

sufficient for germline 3' splice site selection at adjacent 3' splice sites, and

supports the influence of other factors such as intron length (Figure 6). On the

other hand, a consensus sequence alignment motif derived from introns with

significant usage of the proximal AG in *glp-4* whole worms (non tissue-specific

adjacent 3' splice sites) shows the decreased presence of pyrimidines preceding

the distal 3' splice site AG dinucleotides with a concurrent increase preceding

the proximal AG dinucleotide (Figure 7F). Somatic cell-derived tissues may generally depend on nucleotide composition leading to U2AF binding upstream for alternative adjacent 3' splice site decisions, while germline-specific use of upstream alternative 3' splice sites correlates with a poor consensus sequence for the proximal site and a strong consensus for the distal site.

An intron with a single branchpoint used in both soma and germline exhibits tissue-specific alternative adjacent 3' splice site usage.

One hypothesis to explain tissue-specific alternative 3' splice site choice is that distinct intronic branch points are used in the tissues. Expansion or contraction of the distance between the branchpoint and NAGNAG 3' splice sites in transgenes expressed in HEK293T cells was previously reported to enrich splice site selection at the proximal or distal 3' splice site, respectively (Bradley et al.. 2012). Conserved sequence motifs that typically mark intronic branchpoint locations in other eukaryotes have not been identified in *C. elegans*, making their identification more elusive. To assess whether a single branchpoint may be sufficient to allow for splicing at multiple adjacent 3' splice sites, we set out to map the branchpoint(s) from an intron of *cdk-12*, which contains strongly tissue-regulated adjacent 3' splice sites (Figure 7G). RNA was extracted from *glp-4* adults and wildtype gonads as previously described. A single, gene-specific primer complementary to the intron was used to reverse transcribe through the branchpoint. cDNA from this reaction was amplified using divergent, nested primers within the intron (Figure 7H). The PCR products were then ligated into

plasmids, and 10-20 plasmid inserts were sequenced for each tissue type. A single branchpoint, 14nt from the proximal 3' splice site and 20nt from the distal 3' splice site, was observed for this intron of *cdk-12* for both tissues examined (Figure 7I). While an exhaustive, global mapping of *C. elegans* branchpoints has yet to be performed, these data provide evidence that the determination of the tissue-specific 3' splice site for this intron occurs independently of the step in splicing at which the branchpoint is determined.

Adjacent alternative 3' splice sites are conserved in related *Caenorhabditis* species.

To test the evolutionary importance of alternative proximal 3' splice site usage in the germline, we tested whether this tissue-specific phenomenon is conserved in related nematodes. The proximal AG dinucleotide is the only identifiable common sequence element in proximal 3' splice sites in the *C. elegans* germline. Retention of this dinucleotide at these adjacent 3' splice sites over evolutionary time in related nematode species *C. briggsae, C. remanei* and *C. brenneri* would be indicative of conserved alternative splicing in these species. If alternative splicing is not conserved between the species, we would expect the proximal AG dinucleotide intronic sequence to change over evolutionary time, as intron sequences have been observed to change rapidly between *Caenorhabditis* species, with the exception of splicing regulatory regions (Kabat et al.. 2006). Because evolutionary conservation typically suggests functional relevance, we compared the conservation of the proximal AG dinucleotide in species related to

42

C. elegans to the degree of proximal AG dinucleotide usage as a 3' splice site in the germline (Figure 8A). For this, we evaluated the 192 introns with AG dinucleotides spaced 6nt apart (Figure 5A) that are sufficiently expression in both the wildtype germline and *glp-4* whole worm libraries. The vast majority of introns with low (0-19%) proximal 3' splice site usage in germline do not exhibit proximal AG dinucleotide conservation in species related to C. *elegans*; 64% of the proximal AG dinucleotides from this class are only found in *C. elegans* while only 15% are conserved across all four *Caenorhabditis* species. In contrast, introns with intermediate (20-59%) and high (60-100%) usage of the proximal 3' splice site in germline contain proximal AG dinucleotides more frequently in multiple nematodes species (34% and 43% are conserved in all four *Caenorhabditis* species, respectively). This shows that tissue-specific use of the AG dinucleotide in the germline in *C. elegans* is correlated with its conservation. Tissue-specificity of splicing at adjacent 3' splice sites may exist in these related nematode species.

To understand if splicing patterns at tissue-specific alternative 3' splice sites in *C. elegans* are conserved among related species, we extracted total RNA from *C. briggsae* dissected heads, whole worms and dissected gonads as in previous experiments. *C. briggsae* is a species of nematode that diverged from *C. elegans* roughly 50-100 million years ago (Coghlan and Wolfe 2002). The two species have maintained similar genome sizes and structure, but evolutionary changes in the number of genes and the amount of repetitive sequence exist (Stein et al..

2003). Comparative genomics approaches have more recently identified intronic sequence elements associated with alternative cassette exon splicing events conserved between *C. elegans* and *C. briggsae* (Kabat et al.. 2006). RT-PCR with radiolabelled primers that anneal to flanking exons in *atx-2* and *daf-15* revealed conservation of tissue-specificity at adjacent 3' splice sites (Figure 8B). As in *C. elegans*, splice site selection at these *C. briggsae* adjacent 3' splice sites is directed to the proximal 3' splice site in germline and to the distal 3' splice site in heads, suggesting these tissue-specific splicing patterns are evolutionarily conserved.

Conservation across species of the amino acids added by alternative splicing is an important test of protein functionality. However, this is difficult to determine in the case of many adjacent alternative 3' splice sites because the 6 nucleotides upstream of the distal splice site match the *C. elegans* 3' splice site consensus motif TTTCAG. While these nucleotides will likely be translated into amino acids if included by splicing at the proximal splice site in the germline, they already have evolutionary constraints on them to promote splicing at the distal site in somatic tissues. It is, therefore, difficult to determine if the amino acids added to the protein through alternative splicing to a proximal 3' splice site 6nt upstream in the gonad is important for protein function. However, for alternative adjacent 3' splice sites that are separated by 9nt, only the 6nt immediately upstream from the proximal site would have these sequence constraints. This lack of positional nucleotide conservation can be observed in the WebLogo alignment of 53 tissue-

specific alternative 3' splice sites separated by 9nt detailed in Table S2 (Figure 8C). The three nucleotides upstream from the consensus motif (-7 through -9) do not have this evolutionary constraint for splicing. Of the 53 introns with adjacent 3' splice sites separated by 9nt, 24 show conservation of the proximal AG in all 4 nematode species analyzed, suggesting splicing pattern to these 3' splice sites may also be conserved. Furthermore, the majority of amino acids derived from the -7 to -9 nucleotides within these 24 are either fully conserved between all 4 nematode species, conserved between two or more species or the charge/polarity of the residues are conserved (Figure 8C). The conservation of these amino acids suggests that the use of tissue-specific alternative adjacent 3' splice sites corresponds with conservation of changes in the primary structure of the alternative protein isoforms produced.

Non-canonical alternative adjacent 3' splice acceptor dinucleotides are used in wildtype germline

Previously, the TTTC polypyrimidine tract was found to be sufficient to induce splicing following mutation of a TTTCAG 3' splice site to TTTCAA (Zhang and Blumenthal 1996), indicating that cryptic splicing could occur at non-AG dinucleotide acceptors. In our analyses of tissue-specific adjacent 3' splice sites we identified three introns in *C. elegans* in which the germline-preferred proximal 3' splice site does not end in the canonical AG (Figure 9A-C and Supplemental Table 2). Previous EST evidence demonstrates splicing to these sites, but this work provides the first evidence that this splicing is done tissue-

specifically in germline cells and is not a result of cryptic splicing from mutation of a native AG acceptor. In these examples, splicing at TG, AT and GG dinucleotide acceptors (*par-4*, *ubxn-6* and *icd-2*, respectively) is observed and strongly preferred in the germline, despite the apposition of 3' splice sites used heavily in somatic cells containing strong consensus sequence motifs. This small number of cases further delineates the deviation of 3' splice site choice in germline cells from canonical 3' splice site dinucleotide composition requirements found in somatic tissues.

2.3 DISCUSSION

Adjacent alternative 3' splice sites have been identified in several organisms and tissues. Studies in humans and mice, as well as plants have inspired debate concerning how these tandem splice sites are chosen. Statistical and predictive computational models trained on human datasets propose splicing between adjacent 3' splice sites is randomly distributed or relies strictly on the content of nucleotides within the sites (Sinha et al.. 2009). These models struggled with the unique structure of the 3' splice site-determining *cis*-elements in *C. elegans* introns. It was also perceived that the lack of a clear branchpoint consensus sequence and long polypyrimidine tract in *C. elegans* would not allow for alternative splicing at tandem acceptor sites (Hiller et al.. 2004).

Our work shows that adjacent alternative 3' splice sites are present and functional in *C. elegans* as well as its nematode relative *C. briggsae*. Furthermore, we show that splicing patterns to these sites are regulated and tissue-specific. In

this study, we identify alternative splicing differences between *C. elegans* somatic and germline tissues. Through RNA sequencing of isolated gonads as well as *glp-4(bn2ts)* mutant worms that fail to produce a functional, expanded germline, we used pre-annotated and *de novo* transcriptomes to identify 23 cassette exons, 1 intron retention and 203 adjacent 3' splice sites that are differentially spliced between the two tissue types. A large portion of these tissue-regulated alternative 3' splice sites differ by 6 or 9 nucleotides, even in an NMD abrogated background. There is evidence that this preservation of frame is not by chance, as mutations that favor frame preservation at *Drosophila* NAGNAGs allows splicing regulation divergence to be more easily tolerated (McManus et al.. 2014).

The distal splice site used in somatic tissues mirrors a consensus 3' splice site in *C. elegans* with an AG dinucleotide preceded by a short stretch of pyrimidines. The frequency of splicing to the proximal site, which is normally preferred in germline samples, increases in somatic samples when a proximal AG dinucleotide is preceded by a stretch of pyrimidines (ie. when the proximal site is composed of nucleotides that make up a traditional 3' splice site). This indicates that 3' splice site selection in somatic tissues relies on the TTTCAG 3' splice site consensus. In an effort to identify if there are sequence clues that drive splicing to the proximal site in germline cells, we compared the frequency of 4nt, 5nt and 6nt DNA words in introns containing tissue-specific alternative adjacent 3' splicing against total introns (data not shown). We did this with

either the 50nt closest to the 3' end of the introns (the region upstream of the proximal AG) or with total introns. We could find no evidence of consensus sequence motifs that were enriched in introns undergoing tissue-specific alternative adjacent 3' splicing. This differed markedly from our previous analysis of intron sequences flanking alternative cassette exons, where words enhanced in these introns relative to total introns matched known components of the splicing code and allowed for identification of new ones (Kabat et al.., 2006).

Introns in *C. elegans* tend to be relatively small with a median length of 67nt (Figure 6). At some minimal distance, spatial constraints of the spliceosome will likely dictate limits on the branchpoints and splice sites used in the excision of the intron. The contribution to 3' splice site choice that the physical distance from the alternative 3' splice site to the branchpoint makes has been explored in other organisms (Akerman and Mandel-Gutfreund 2006; Tsai et al.. 2007; Tsai et al.. 2010; Bradley et al.. 2012). Expression analysis of a PTB minigene with 4nt or 7nt insertions between the branchpoint and tandem acceptor set caused an increase in splicing to the proximal site, suggesting that the choice of branchpoint may influence the choice of 3' splice site at NAGNAGs (Tsai et al., 2010). Conversely, tissue-specific branchpoint mapping in this study revealed no alternative branchpoint usage in an intron with adjacent tissue-specific 3' splice sites (Figure 7). Instead, introns with activated proximal AG dinucleotides in the germline tend to be significantly longer overall (Figure 6) and possibly

48

more permissive of spliceosomal spatial modulations that lead to 3' splice site choice fluctuations. If alternative branchpoints do not lead to alternative 3' splice sites, perhaps the scanning mechanism from branchpoint to splice acceptor differs between the germline and somatic splicing machineries.

The discovery of so many tissue-specific alternative adjacent 3' splice sites, in which the proximity to the 5' splice site is the main determinant of alternative splicing in the germline, suggests that there is an overall difference in 3' splice site selection in the germline relative to other tissues. Differential expression of splicing *trans*-factors between germline and somatic tissues that are part of the U2 snRNP or are interactors with the U2AF subunit homologs UAF-1 and UAF-2 could explain this phenomenon. To begin to address this, we used a database of 494 spliceosome-associated components (Cvitkovic and Jurica, 2012) to identify spliceosomal proteins and RNA binding proteins in C. *elegans*. Cross-referencing our tissue-specific RNA-seq data using the DESeq analysis software (Anders and Huber 2010), we identified candidate factors enriched or depleted in germline relative to somatic cells (Supplemental Table 6 and Supplemental Figure 1). Two candidates to look at, UAF-1 and UAF-2, did not yield expression changes of significance in this analysis and no detectable change in splicing pattern at adjacent 3' splice sites was observed in a mutant allele (*uaf-1)* or following RNAi knockdown (*uaf-2)* (Supplemental Figure 2). Similarly, we tested other candidate splicing regulators, including *C. elegans mog-2* (mammalian U2A'

homolog), *ptb-1* (SAP49; polypyrimidine tract-binding protein), *hrpf-1* and *sym-2* (hnRNP H/F homologs) and *sfa-1* (Branchpoint-binding protein homolog) and detected no change in splicing patterns at tissue-specific 3' splice sites. One possibility is that splicing fidelity is lower in the germline and proofreading mechanisms may not be as robust. This could explain the use of proximal sites that lack the YYYNAG consensus and even proximal sites that lack the AG dinucleotide.

Our data indicate that the presence of an AG dinucleotide 6 to 9nt upstream of a strong consensus 3' splice site, will likely lead to proximal alternative splice site usage if the pre-mRNA is expressed in the germline and the intron is above a minimal length. Importantly, our data indicate that these tissue-specific alternative splicing events, their component nucleotides and the resulting amino acids within the exonic extensions, are conserved in related nematode species. No matter the mechanism by which germline-specific splicing is established at adjacent 3' splice sites, this mechanism has been adapted to drive conserved alternative splicing events.

2.4 METHODS

Strain Maintenance and Dissections

Strains: *glp-4(bn2ts)*, *smg-2(e2008)* and wildtype (Bristol N2 strain) worms were obtained from the *Caenorhabditis* Genetics Center. Synchronized worm populations were obtained by axenization of a mixed population of worms

cultured on liquid media at 20°C to isolate embryos (Lewis and Fleming 1995). After synchronization, all worms were cultured at 25°C, the restrictive temperature for *glp-4(bn2ts),* and the developmental stage was confirmed by microscopy of vulva development (white-crescent stage).

RNA Extraction For RT-PCRs

Gonads were dissected on slides in 50µl Dissection Buffer (110µl 10X Egg Buffer (1M HEPES, 5M NaCl, 1M $MgCl_2$, 1M $CaCl_2$, 1M KCl), 25µl 20% TWEEN, 100µl 10mM levamisole, 865µl ddH_2O] with 30 gauge needles. 100-150 heads were removed by cutting below the pharynx. 75-100 extruded gonads were isolated by removing the heads and cutting at the spermatheca. 100-200 adult worms 20-24 hours past L4 were used for whole worm samples. Tissues were placed into 300µl TRIzol (Invitrogen) on ice. 60µl chloroform was added to the TRIzol/tissue samples and phases were separated using prespun Phase Lock Gel-Heavy tubes (5 Prime). The aqueous phase was transferred to a new tube and mixed with 125-175µl isopropanol and 1µl GlycoBlue (Ambion). RNA was precipitated overnight at -20°C and pelleted at 13,000xg at 4°C for 30 minutes. The pellet was washed with 75% ethanol, dried, and resuspended in 10-20µl ddH_2O.

Reverse Transcription

1-10µl of RNA was mixed with 1µl 10mM dNTPs, 1µl 50µM Random Hexamers and ddH_2O to 13µl. The solution was incubated at 65°C for 5 minutes and then for 1 minute on ice. 4µl 1st-strand Buffer, 2µl DTT and 1µl SuperScript III

(Invitrogen) were added and the following incubation protocol was followed: 4°C 10 minutes, 15°C 10minutes, 42°C 20 minutes, 70°C 15 minutes.

Polymerase Chain Reaction

For cassette exons, gene specific primers (10µM) that anneal to sequences in exons that flank an alternative exon were used in a PCR reaction with 1-2µl cDNA from RT reactions. The annealing temperature was 50°C and the elongation time was 30 seconds with 25-28 cycles of amplification. Primer sequences are available in Supplemental Table 8. 1µl of product was run on an Agilent 2100 Bioanalyzer with a DNA 1000 kit and/or separated on a 2% agarose gel. Dividing the molarity of the inclusion isoform by the sum of molarities of inclusion and skipping isoforms determined the Percent Spliced In (PSI). For alternative 3' splice sites, PCR reactions were run similar to cassette exons, but the elongation time was 15 seconds and the reactions were run in the presence of ^{32}P-labelled reverse primer. Radiolabelled PCR products from alternative 3' splice sites were phenol, then chloroform extracted and ethanol precipitated at -20°C overnight. After microcentrifugation, the pellets were resuspended in 10µl formamide dye, heated to 95°C and run on a long 6% polyacrylamide urea denaturing gel at 1700V/45W for 3.5 to 4 hrs. This gel was exposed to a PhosphorImager screen overnight and bands were imaged on a Typhoon scanner.

RNA-seq

RNA extractions were performed with TRIzol (Invitrogen) and further purified using the RNeasy Plus Micro kit (Qiagen). Gonad samples were depleted of rRNA with a RiboZero kit (Epicentre) and whole worm samples were poly-A selected. mRNA sequencing libraries were sequenced according to manufacturer recommendations with TruSeq Stranded mRNA Sample kit (Illumina) at the Centre for Genomic Regulation sequencing facility in Barcelona, Spain. 50 bp paired-end reads were mapped to the *C. elegans* reference genome (ce6) with TopHat (Trapnell et al.. 2009) and PCR duplicates were removed. SpliceTrap (Wu et al.. 2011) was used to detect pre-annotated splicing changes between tissues that met the threshold of a 0.3 minimum inclusion ratio change and a minimum count of 15 junction spanning reads (either skipping or inclusion on each side). Identification of *smg-2(e2008)* alternative 3' splice sites was performed on dissected gonad sample libraries. Splice junctions were extracted from TopHat mappings and code was written to identify introns that were filtered according to: 1) introns with more than one 3' splice site for the same 5' splice site and 2) among the 3' splice sites for a given 5' splice site, the distance between 3' splice sites is ≤18nt. The resulting introns were hand curated to detect tissue-specific splicing changes at adjacent 3' splice sites.

Consensus Motifs

Typical non-alternatively spliced introns were selected by random gene and intron selection in the UCSC Genome Browser. Intron sequences contain the

previously annotated 3' splice site through 40nt upstream. A consensus motif was created on-line with WebLogo (http://weblogo.threeplusone.com/create.cgi). 25 introns from the following classes, 1) 6nt, 9nt and 12nt shifted alternative 3' splice sites, 2) 3' splice sites in which the proximal AG is unused in gonad samples and 3) often used 3' splice sites in *glp-4(bn2ts)*, were chosen and consensus motifs were created in the same manner as typical 3' splice sites.

Branchpoint Identification

RNA was extracted from *glp-4* whole worms or dissected wildtype gonads. Divergent nested primer sets (A1/B1 and A2/B2) were designed to anneal to an intron within *cdk-12* that contains a tissue-specific alternative 3' splice site. Reverse transcription was performed as previously stated substituting the A1 primer for random hexamers. cDNA from this reaction was amplified in a primary PCR reaction with primers A1 and B1 (28 cycles), purified in a PCR Clean-up Kit (Qiagen) and reamplified in a secondary PCR reaction using primers A2 and B2 (25 cycles). Primer sequences are available in Supplemental Table 8. The resulting amplicon was gel purified, cloned into pCR2.1 TOPO (Invitrogen) and transformed into DH5alpha chemically competent cells. A minimum of 12 colonies were selected for sequencing each from the samples representing *glp-4* adult RNA and wildtype gonad RNA.

Intron Length Analysis

Nucleotide lengths were measured for introns with >10 junction-spanning reads and <10% (n=72) or >10% (n=120) proximal 3'SS usage in the germline (Figure 5). In addition, introns were identified in the UCSC Genome Table Browser (ce6 release) and nucleotide lengths were measured. The minimum, 1st quartile, median, 3rd quartile, and maximum lengths were determined from each set and compared.

2.5 DATA ACCESS

FASTQ files of raw reads from five high throughput RNA-seq libraries and a table of genes with tissue-specific alternative 3' splice sites can be accessed from the Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/gds) under accession number GSE64672.

2.6 ACKNOWLEDGEMENTS

Figure 1. Developmental and tissue-specific changes in alternative splicing of the *top-1* gene. (A) Exon inclusion of *top-1* alternative cassette exon occurs with the onset of germline development. A 1% agarose gel of reverse transcription-polymerase chain reaction (RT-PCR) products of RNA extracted from hand-picked *C. elegans* whole worms at indicated developmental stages. Exon inclusion and skipping products are indicated at right. (B) A 2100 Bioanalyzer Image of *top-1* RT-PCR products of RNA extracted from L4+1day *C. elegans* dissected wildtype gonads, adult whole worms, and dissected heads, along with *glp-4(bn2ts)* mutant whole worms that fail to develop a germline.
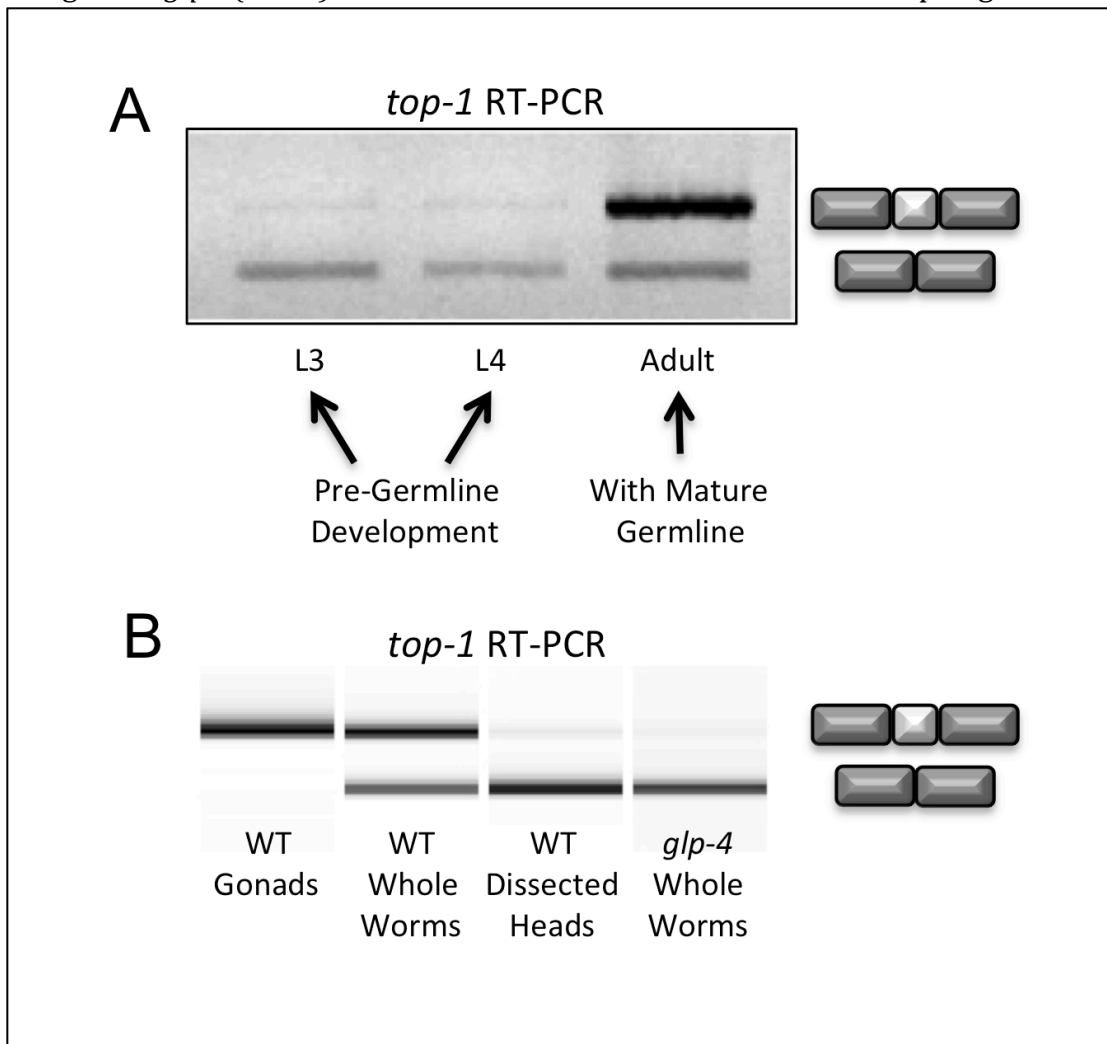
Figure 2. Cassette exon alternative splicing changes between *C. elegans* **germline and somatic tissues.** (A) Flowchart depicting the process of RNA isolation, high-throughput sequencing and computational analysis. (B) RT-PCR products from primers that flank alternative cassette exons highlighting tissue-specific inclusion/skipping. Products were separated on an Agilent 2100 Bioanalyzer. PSI was calculated by dividing the molarity of the inclusion product by the sum of the molarities of the inclusion and skipping products. (C) Representation of the Wormbase gene annotation for *rgr-1* alternative isoforms and normalized RNA-sequencing coverage tracks for the indicated libraries. This demonstrates skipping of the alternative cassette exon in gonad. (D) Representation of the Wormbase gene annotation for C06A6.4 showing alternative isoforms and RNA-sequencing coverage. This demonstrates an example of a cassette exon that is specifically included in the germline.
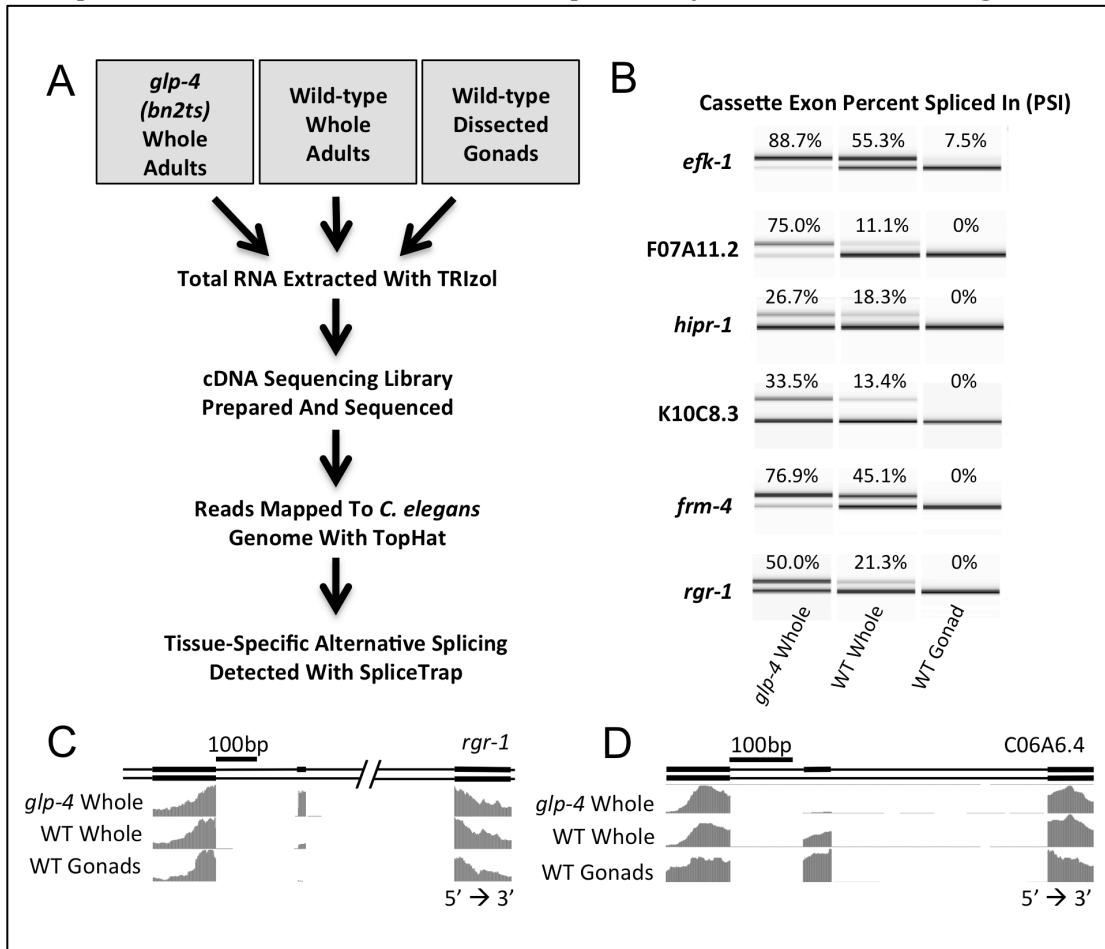
Figure 3. The major class of splicing changes between germline and somatic cells consists of adjacent alternative 3' splice sites. (A) *taf-10* alternative isoforms and RNA-sequencing coverage tracks showing a tissue-specific alternative 3' splicing event. Germline reads primarily cross the 3' splice site closer to the 5' splice site (proximal) while reads in somatic cells primarily use the 3' splice site further from the 5' splice site (distal). (B) Cartoon depicting the proximal and distal 3' splice sites, the tissue-type in which each 3' splice site is primarily used (in parenthesis) and the location of the ^{32}P-labelled oligos used to validate the tissue-specific enrichment of each isoform by RT-PCR. ^{32}P RT-PCR products from 3 sample genes were separated on a 6% polyacrylamide denaturing gel, and visualized with a PhosphorImager. (C) Nucleotides preceding and within tissue-specific 3' splice sites in a representative set of genes. The proximal and distal 3' splice sites are in bold for each sequence (left and right, respectively). Nucleotides are spaced to show the 3nt periodicity and maintenance of frame between the 3' splice sites.

Figure 3

C

	Proximal			Distal			
taf-10	GAT TTA **AAG**	TAT TTT **CAG**				
Y69A2AR.1	TTT CTA **AAG**	CTT TTG **AAG**				
ufd-3	TTA AAT **TAG**	TTC **CAG**				
lst-4	GCT TGC **TAG**	CGT **CAG**				
F56C9.10	TTA ATT **GAG**	AAT GTG **AAG**				
M03F8.3	CTT CGC **TAG**	TAC TTT **CAG**				
dli-1	TTG CAA **AAG**	TTT **TAG**				
sel-10	ACG TAC **CAG**	TTC **CAG**				
ced-7	TGT CTA **AAG**	TTT **CAG**				
F27C1.2	CCG ACC **TAG**	TTT GCA CTT **CAG**				
gfi-2	CTG TGT **AAG**	TAT TAT TTC **CAG**				
B0336.5	AGT TCC **CAG**	AAG TTG **CAG**				
daf-15	TAT TAA **AAG**	TTG TCG AAA TGA TTC **CAG**				
R06F6.8	TTT AAA **AAG**	TTT **CAG**				
C36B1.8	GAA ATA **AAG**	CTA TTG **CAG**				

Figure 4.   The majority of adjacent alternative 3' splice sites are maintained in frame in the absence of nonsense-mediated decay (NMD). (A) Approach for identification of alternative 3' splice site isoforms in the absence of NMD. Total RNA from *smg*-2 mutant gonads was extracted, 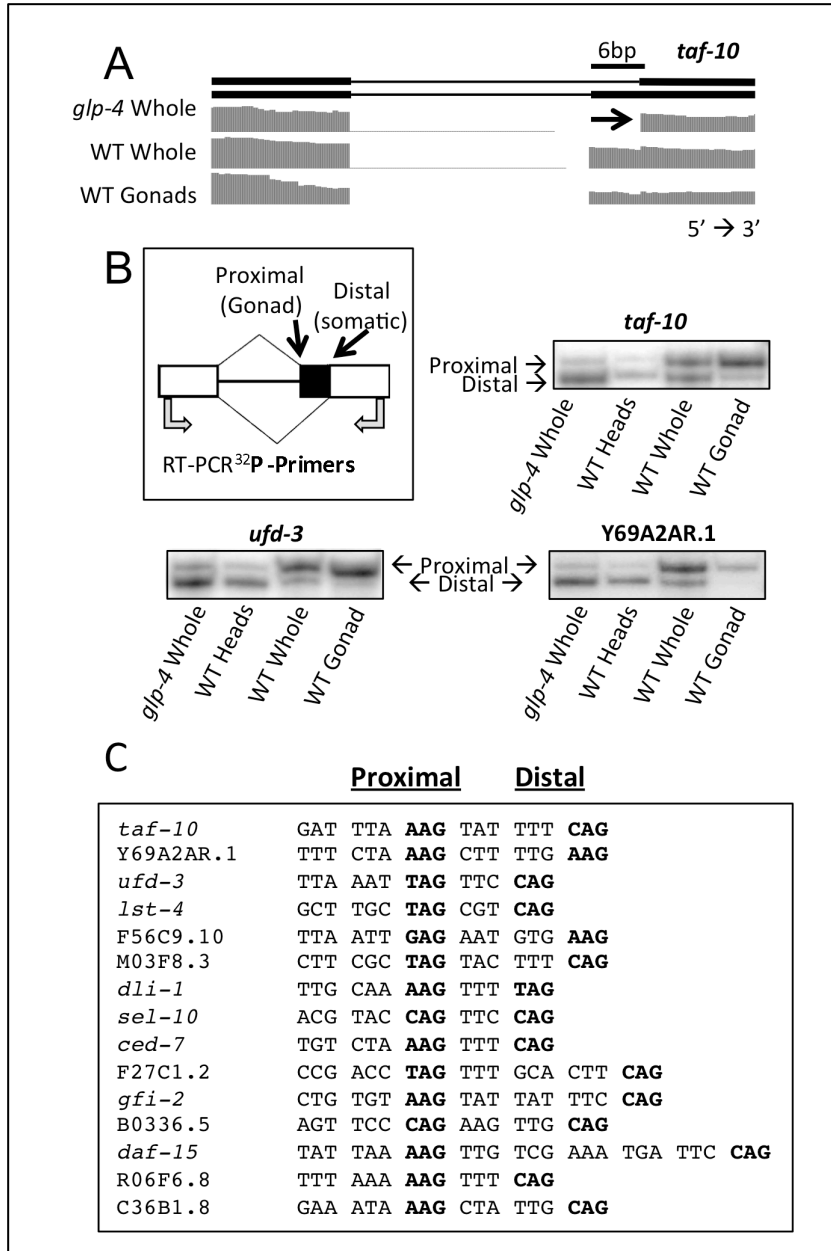made into a cDNA library and subjected to high-throughput sequencing.   Introns with common 5' splice sites and adjacent alternative 3' splice sites ≤18nt apart were identified. (B) Table showing the number of alternative 3' splice sites ≤18nt apart in the *smg*-2 germline that maintain reading frame and those that cause a frameshift. This is further subdivided to show the corresponding change in expression or proximal splice site usage in the wildtype relative to *smg-2* germline. (C) Graph depicting the percentage of introns (y-axis) identified in the *smg*-2 germline that have specified numbers of nucleotides separating the adjacent 3' splice sites (x-axis). (D) Representation of the Wormbase gene annotation for *trxr-1*. Note that alternative splicing is not annotated for this intron. The sequencing tracks show a frameshift-causing alternative 3' ss 7nt upstream with germline-specific isoform enrichment that is enhanced in the *smg-2* mutant germline.

A	B	3' Splice Sites In Frame (in smg-2 gonad)	315

3' Splice Sites In Frame (in smg-2 gonad)	315
<10 Junction-Spanning Reads (in WT Gonad)	27 (8.6%)
No Change in Proximal 3'SS Usage (in WT Gonad)	180 (57.1%)
Increased Proximal 3'SS Usage (in WT Gonad)	75 (23.8%)
Decreased Proximal 3'SS Usage (in WT Gonad)	33 (10.5%)
3' Splice Sites Out of Frame (in smg-2 gonad)	172
<10 Junction-Spanning Reads (in WT Gonad)	31 (18.0%)
No Change in Proximal 3'SS Usage (in WT Gonad)	93 (54.1%)
Increased Proximal 3'SS Usage (in WT Gonad)	12 (7.0%)
Decreased Proximal 3'SS Usage (in WT Gonad)	36 (20.9%)
TOTAL	**487**

Figure 5. Proximal alternative 3' splice site usage is favored in the germline. (A) Breakdown of annotated introns with a terminal AG dinucleotide and an AG dinucleotide 6nt upstream. Introns separated according to the total number of reads crossing the splice junction, the presence of splicing at the proximal 3' splice site, the change in percentage of isoforms using the proximal 3'ss between *glp-4* whole worm and wildtype germline, and the method used to identify the tissue-specific 3'ss. (B) Graph of alternative adjacent proximal 3' ss usage for the 192 introns with an AG dinucleotide 6nt upstream of an annotated 3' splice site and >10 junction spanning reads in both germline and somatic libraries. Introns are arranged in order of increasing proximal 3' splice site use percentage in wildtype germline (black line, left to right) and the corresponding proximal 3' splice site usage of each intron in the somatic *glp-4* whole worm library (grey line).

Figure 5
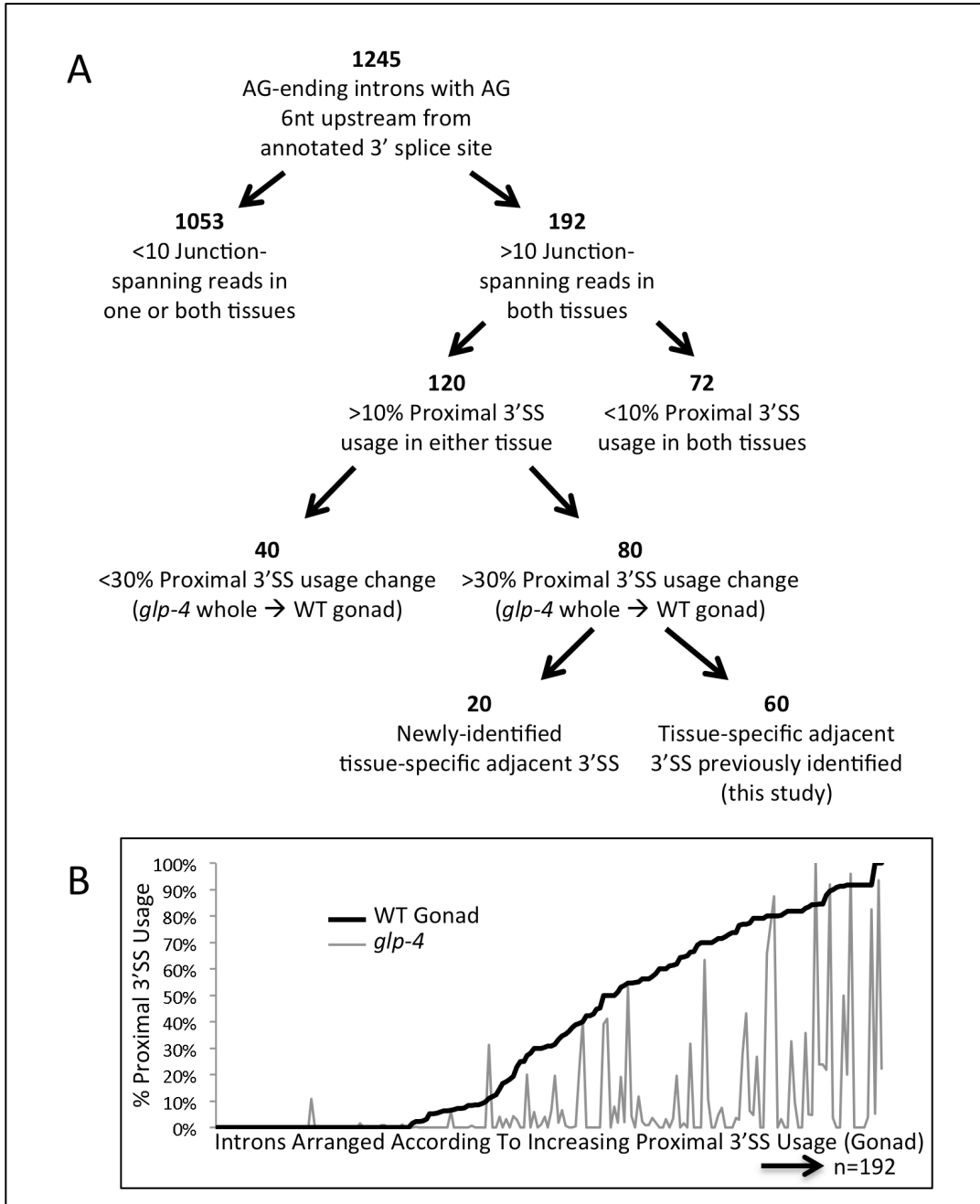
Figure 6. Introns that utilize the proximal AG dinucleotide as a 3' splice site in the germline are longer than those that do not.

Length of introns that utilize the proximal AG in <10% and >10% of junction-spanning reads as well as total *C. elegans* introns identified in the UCSC Genome Browser. The bottom and top of the boxes are the beginning of the 2nd and 4th quartiles respectively, with the median represented by the line in the center of the boxes.

	<10% Prox 3'SS Usage	>10% Prox 3'SS Usage	C. elegans Total Introns
Minimum	43	31	1
Q1	47	52	48
Median	49	95	67
Q3	75	467	332
Maximum	1150	2611	100913
Count	72	120	108604
Mean	125	361	323

Figure 7. Sequence comparison and branchpoint analysis of introns with tissue-specific adjacent alternative 3' splice sites. (A) WebLogo sequence comparison of regions upstream of a randomly selected set of 3' splice sites. Upstream frequencies were measured from the distal splice site to 40nt upstream. A logo displays the frequencies of bases at each position as the relative heights of letters, along with the degree of sequence conserva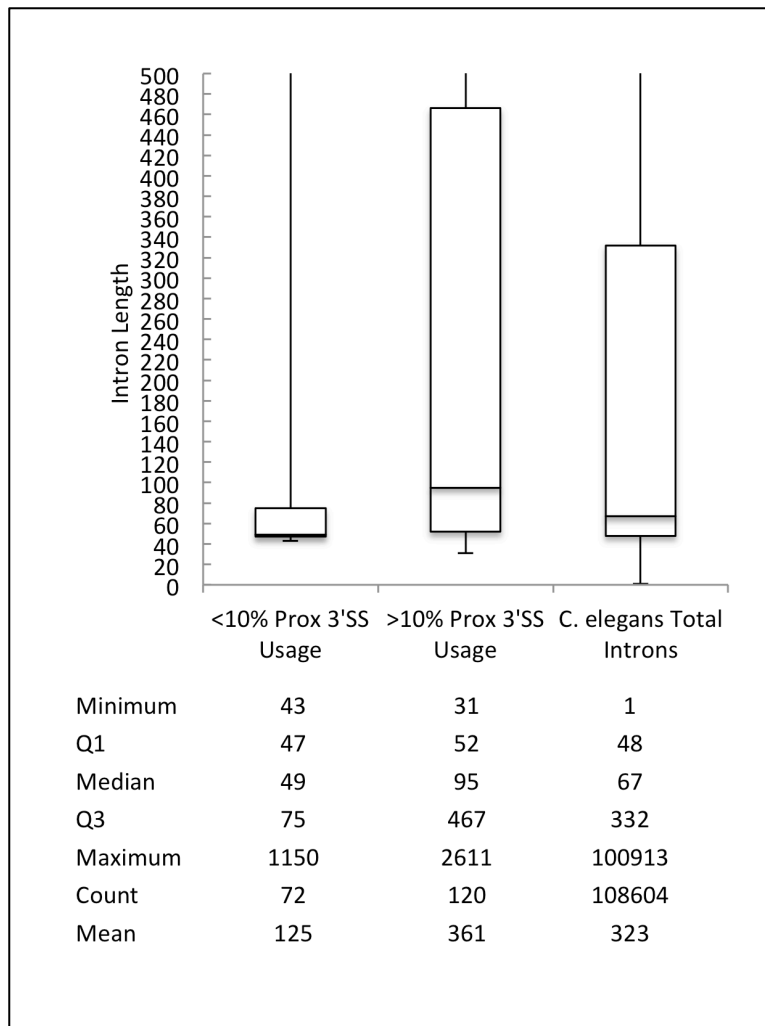tion as the total height of a stack of letters, measured in bits of information. (B-D) WebLogo sequence comparisons of regions upstream of tissue-specific alternative 3' splice sites that are 6nt, 9nt and 12nt apart. (E) WebLogo sequence comparison of tissue-specific alternative 3' splice sites 6nt apart in which the proximal AG dinucleotide is used <10% in germline. Note the similarity of this sequence logo to that in which the dominant 3' splice site in germline is the proximal AG dinucleotide (Figure 7B). (F) WebLogo sequence comparison of the 25 most used AG dinucleotide proximal 3' splice sites in *glp-4(bn2ts)* when found 6nt upstream from an intron-terminating AG dinucleotide. Note the signature of TTTCAG at the proximal 3' splice site that is lacking in the tissue-specific proximal 3' splice site (Figure 7B). (G) Tissue-specific alternative 3' splice site usage in the *cdk-12* gene. (H) Orientation of divergent primers used to map branchpoints within an intron. (I) Lariat structure with alternative 3' splice sites, the convergent orientation of the primers within the structure and the branchpoint through which reverse transcriptase passes. Reverse transcription was first performed using the A1 primer, amplified with A1/B1 primers, purified using a Qiagen PCR clean-up kit, amplified a second time using the A2/B2 primers and gel purified. (J) Terminal 27nt of the *cdk-12* intron containing the tissue-specific alternative 3' splice site with the proximal and distal 3' splice sites and the location of the only branchpoint identified in our analysis. Also shown are the distances from this branchpoint to the proximal and distal 3' splice sites.
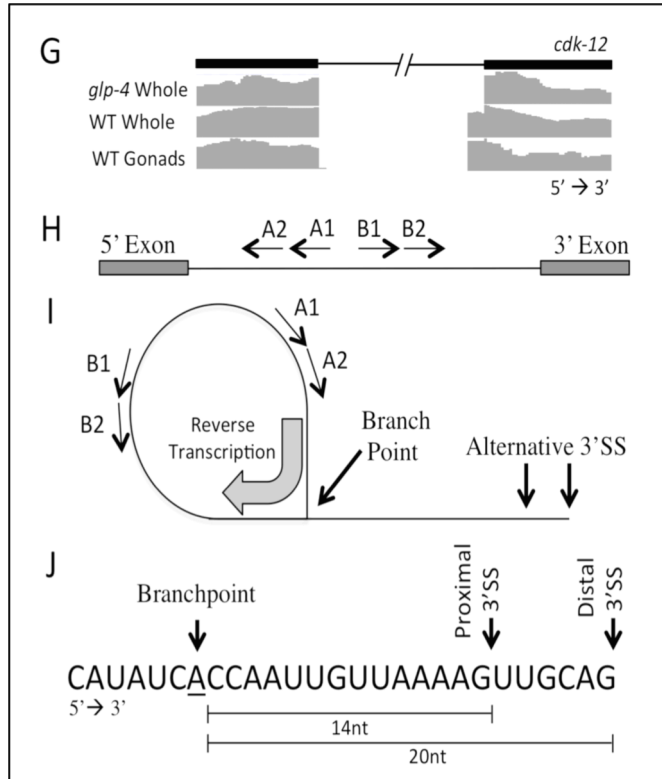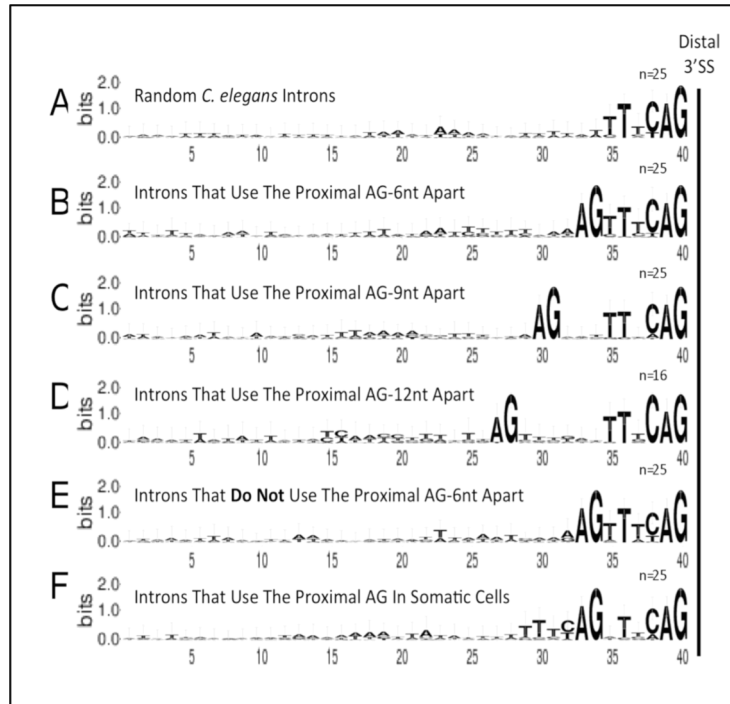
Figure 7

65

Figure 8. Adjacent alternative 3' splice sites are conserved in related *Caenorhabditis* **species.** (A) Bar graph depicting conservation of the proximal AG dinucleotide in tissue-specific alternative 3' splice sites. Introns were separated according to percent of proximal 3' splice site usage in wildtype germline (0-19% in black, 20-59% in dark grey and 60-100% in light grey). The comparison is made with three species related to *C. elegans*: *C. remanei, C. briggsae and C. brenneri*. Displayed is a bar graph depicting the percentage of events in each category in which the upstream AG dinucleotide is conserved in only *C. elegans* (1) or in *C. elegans* plus 1, 2 or 3 other related species (2, 3 and 4). (B) ^{32}P-labelled RT-PCR products made from primers that flank introns in *daf-15* and *atx-2* with tissue-specific alternative adjacent 3' splice sites. Enrichment in *C. elegans* or *C. briggsae* for somatic cells (g*lp-4* or dissected heads, respectively) and germline cells (dissected gonads) shows a shift in abundance from the distal site isoform to the proximal site isoform in the germline of both species. (C) WebLogo consensus sequence motif of the 9nt preceding the distal 3' splice site of the 53 introns with tissue-specific 3' splice sites separated by 9nt (from Supplemental Table S2). Note the lack of conserved nucleotides at positions -7 to -9 from the distal 3' splice site. On the right, a table depicting the 24 introns with the AG dinucleotide at the proximal site 9nt upstream of the distal splice site conserved in all four nematode species analyzed. Shown is the gene name, common name, the frame of translation and the identity of the amino acid(s) derived from the -7 to -9 nucleotides.
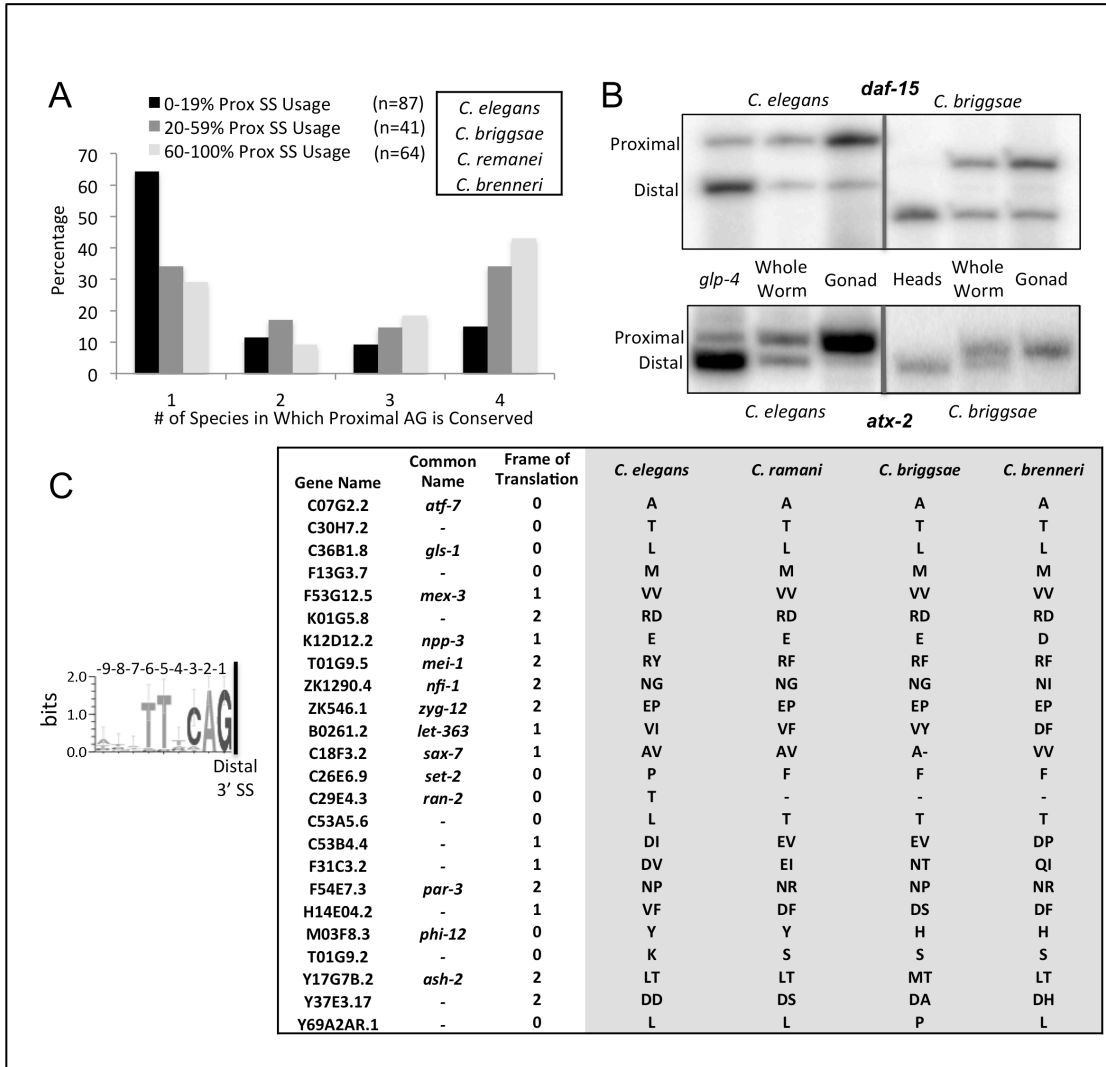
Figure 8

A

- 0-19% Prox SS Usage (n=87)
- 20-59% Prox SS Usage (n=41)
- 60-100% Prox SS Usage (n=64)

C. elegans
C. briggsae
C. remanei
C. brenneri

Percentage (y-axis: 0, 10, 20, 30, 40, 50, 60, 70)

# of Species in Which Proximal AG is Conserved (x-axis: 1, 2, 3, 4)

B

daf-15

C. elegans — *C. briggsae*

Proximal
Distal

glp-4 | Whole Worm | Gonad | Heads | Whole Worm | Gonad

Proximal
Distal

C. elegans — *atx-2* — *C. briggsae*

C

bits (y-axis: 0.0, 1.0, 2.0)
-9-8-7-6-5-4-3-2-1
TT CAG
Distal 3' SS

| Gene Name	Common Name	Frame of Translation	*C. elegans*	*C. ramani*	*C. briggsae*	*C. brenneri*
C07G2.2	atf-7	0	A	A	A	A
C30H7.2	-	0	T	T	T	T
C36B1.8	gls-1	0	L	L	L	L
F13G3.7	-	0	M	M	M	M
F53G12.5	mex-3	1	VV	VV	VV	VV
K01G5.8	-	2	RD	RD	RD	RD
K12D12.2	npp-3	1	E	E	E	D
T01G9.5	mei-1	2	RY	RF	RF	RF
ZK1290.4	nfi-1	2	NG	NG	NG	NI
ZK546.1	zyg-12	2	EP	EP	EP	EP
B0261.2	let-363	1	VI	VF	VY	DF
C18F3.2	sax-7	1	AV	AV	A-	VV
C26E6.9	set-2	0	P	F	F	F
C29E4.3	ran-2	0	T	-	-	-
C53A5.6	-	0	L	T	T	T
C53B4.4	-	1	DI	EV	EV	DP
F31C3.2	-	1	DV	EI	NT	QI
F54E7.3	par-3	2	NP	NR	NP	NR
H14E04.2	-	1	VF	DF	DS	DF
M03F8.3	phi-12	0	Y	Y	H	H
T01G9.2	-	0	K	S	S	S
Y17G7B.2	ash-2	2	LT	LT	MT	LT
Y37E3.17	-	2	DD	DS	DA	DH
Y69A2AR.1	-	0	L	L	P	L

Figure 9. 3' splice sites in dissected gonad samples that deviate from the canonical AG dinucleotide. [32]P RT-PCR and RNA-sequencing coverage tracks of handpicked *glp-4* whole worm, wildtype whole worm and wildtype germline RNA showing tissue-specific expression of isoforms with alternative 3' splice sites. Analysis of *par-4* (A), *ubxn-6* (B) and *icd-2* (C) RNA-sequencing reveals significant usage of non-canonical 3' splice sites primarily in germline-containing samples. Sequences above each set of coverage tracks shows the dinucleotide set found at the somatic site preceded by the dinucleotide set found at the germline site (underlined) and the accompanying sequence immediately upstream. Genomic and mRNA sequence in these regions were validated.
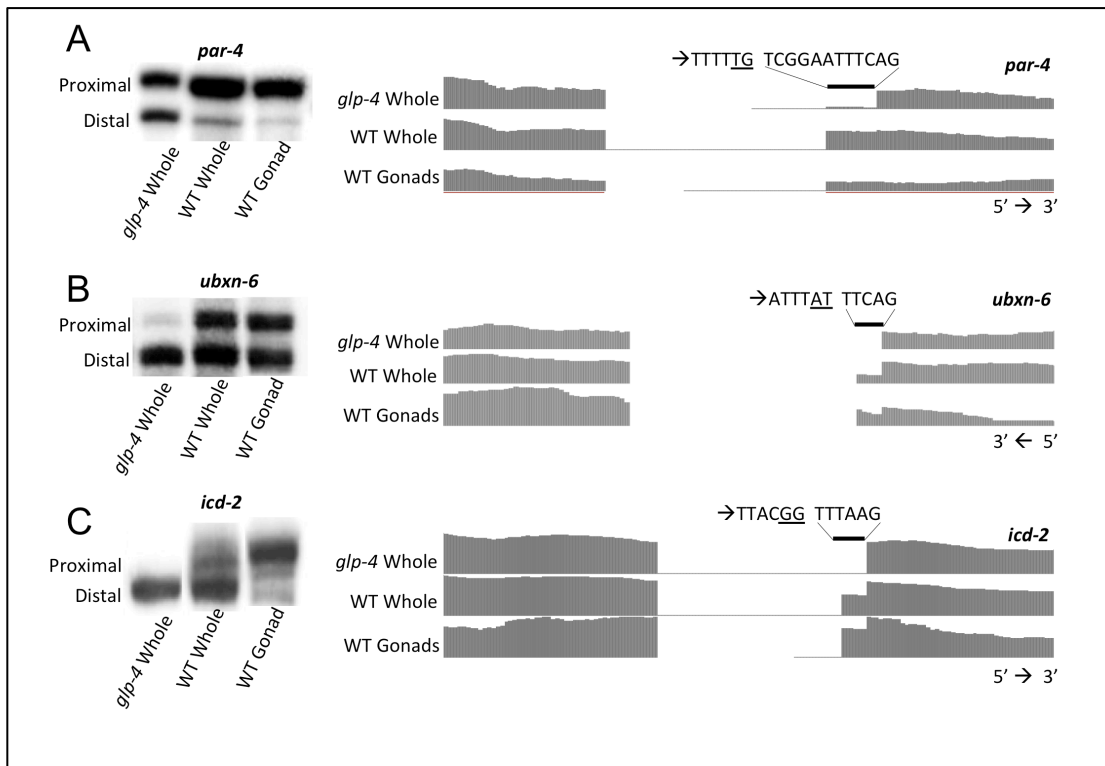
2.7 REFERENCES

Akerman M, Mandel-Gutfreund Y. 2006. Alternative splicing regulation at tandem 3' splice sites. *Nucleic Acids Res* **34**(1): 23-31.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**(10): R106.

Aoki H, Sato S, Takanami T, Ishihara T, Katsura I, Takahashi H, Higashitani A. 2000. Characterization of Ce-atl-1, an ATM-like gene from *Caenorhabditis elegans*. *Molecular & general genetics : MGG* **264**(1-2): 119-126.

Ares M, Jr., Weiser B. 1995. Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog Nucleic Acid Res Mol Biol* **50**: 131-159.

Aroian RV, Levy AD, Koga M, Ohshima Y, Kramer JM, Sternberg PW. 1993. Splicing in *Caenorhabditis elegans* does not require an AG at the 3' splice acceptor site. *Mol Cell Biol* **13**(1): 626-637.

Barberan-Soler S, Fontrodona L, Ribo A, Lamm AT, Iannone C, Ceron J, Lehner B, Valcarcel J. 2014. Co-option of the piRNA Pathway for Germline-Specific Alternative Splicing of C. elegans TOR. *Cell reports* **8**(6): 1609-1616.

Barberan-Soler S, Lambert NJ, Zahler AM. 2009. Global analysis of alternative splicing uncovers developmental regulation of nonsense-mediated decay in C. elegans. *RNA* **15**(9): 1652-1660.

Barberan-Soler S, Ragle JM. 2013. Alternative Splicing Regulation of Cancer-Related Pathways in *Caenorhabditis elegans*: An In Vivo Model System with a Powerful Reverse Genetics Toolbox. *International journal of cell biology* **2013**: 636050.

Barberan-Soler S, Zahler AM. 2008. Alternative splicing regulation during C. elegans development: splicing factors as regulated targets. *PLoS Genet* **4**(2): e1000001.

Beanan MJ, Strome S. 1992. Characterization of a germ-line proliferation mutation in C. elegans. *Development* **116**(3): 755-766.

Berglund JA, Abovich N, Rosbash M. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev* **12**(6): 858-867.

Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**(5): 781-787.

Blumenthal T, Steward K. 1997. RNA processing and gene structure. In *C elegans II*, (ed. DL Riddle, T Blumenthal, BJ Meyer, JR Priess), pp. 117-145. Cold Spring Harbor Laboratory Press, Plainview, NY.

Bradley RK, Merkin J, Lambert NJ, Burge CB. 2012. Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. *PLoS Biol* **10**(1): e1001229.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**(5396): 2012-2018.

Coghlan A, Wolfe KH. 2002. Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Res* **12**(6): 857-867.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**(6): 1188-1190.

Hamid FM, Makeyev EV. 2014. Emerging functions of alternative splicing coupled with nonsense-mediated decay. *Biochem Soc Trans* **42**(4): 1168-1173.

Higashitani A, Aoki H, Mori A, Sasagawa Y, Takanami T, Takahashi H. 2000. *Caenorhabditis elegans* Chk2-like gene is essential for meiosis but dispensable for DNA repair. *FEBS Lett* **485**(1): 35-39.

Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, Schreiber S, Backofen R, Platzer M. 2004. Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat Genet* **36**(12): 1255-1257.

Hodgkin J, Papp A, Pulak R, Ambros V, Anderson P. 1989. A new kind of informational suppression in the nematode *Caenorhabditis elegans*. *Genetics* **123**(2): 301-313.

Hollins C, Zorio DA, MacMorris M, Blumenthal T. 2005. U2AF binding selects for the high conservation of the C. elegans 3' splice site. *Rna* **11**(3): 248-253.

Kabat JL, Barberan-Soler S, McKenna P, Clawson H, Farrer T, Zahler AM. 2006. Intronic alternative splicing regulators identified by comparative genomics in nematodes. *PLoS Comput Biol* **2**(7): e86.

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**(Database issue): D493-496.

Kuroyanagi H, Kimura T, Wada K, Hisamoto N, Matsumoto K, Hagiwara M. 2000. SPK-1, a C. elegans SR protein kinase homologue, is essential for embryogenesis and required for germline development. *Mech Dev* **99**(1-2): 51-64.

Kuroyanagi H, Watanabe Y, Suzuki Y, Hagiwara M. 2013. Position-dependent and neuron-specific splicing regulation by the CELF family RNA-binding protein UNC-75 in *Caenorhabditis elegans*. *Nucleic Acids Res* **41**(7): 4015-4025.

Lee MH, Jang YJ, Koo HS. 1998. Alternative splicing in the *Caenorhabditis elegans* DNA topoisomerase I gene. *Biochim Biophys Acta* **1396**(2): 207-214.

Lewis JA, Fleming JT. 1995. Basic culture methods. *Methods in cell biology* **48**: 3-29.

McManus CJ, Coolon JD, Eipper-Mains J, Wittkopp PJ, Graveley BR. 2014. Evolution of splicing regulatory networks in *Drosophila*. *Genome Res* **24**(5): 786-796.

Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J. 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**(6763): 838-841.

Nagy E, Maquat LE. 1998. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* **23**(6): 198-199.

Pulak R, Anderson P. 1993. mRNA surveillance by the *Caenorhabditis elegans* smg genes. *Genes Dev* **7**(10): 1885-1897.

Ramani AK, Calarco JA, Pan Q, Mavandadi S, Wang Y, Nelson AC, Lee LJ, Morris Q, Blencowe BJ, Zhen M et al.. 2011. Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res* **21**(2): 342-348.

Roussell DL, Bennett KL. 1993. glh-1, a germ-line putative RNA helicase from *Caenorhabditis*, has four zinc fingers. *Proc Natl Acad Sci U S A* **90**(20): 9300-9304.

Shim YH. 1999. elt-1, a gene encoding a *Caenorhabditis elegans* GATA transcription factor, is highly expressed in the germ lines with msp genes as the potential targets. *Mol Cells* **9**(5): 535-541.

Sinha R, Nikolajewa S, Szafranski K, Hiller M, Jahn N, Huse K, Platzer M, Backofen R. 2009. Accurate prediction of NAGNAG alternative splicing. *Nucleic Acids Res* **37**(11): 3569-3579.

Smith CW, Chu TT, Nadal-Ginard B. 1993. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol Cell Biol* **13**(8): 4939-4952.

Smith CW, Porro EB, Patton JG, Nadal-Ginard B. 1989. Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature* **342**(6247): 243-247.

Spieth J, Lawson D. 2006. Overview of gene structure. *WormBook*: 1-10.

Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A et al.. 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol* **1**(2): E45.

Sulston JE, Horvitz HR. 1977. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol* **56**(1): 110-156.

TeKippe M, Aballay A. 2010. C. elegans germline-deficient mutants respond to pathogen infection using shared and distinct mechanisms. *PLoS One* **5**(7): e11777.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**(9): 1105-1111.

Tsai KW, Chan WC, Hsu CN, Lin WC. 2010. Sequence features involved in the mechanism of 3' splice junction wobbling. *BMC molecular biology* **11**: 34.

Tsai KW, Tarn WY, Lin WC. 2007. Wobble splicing reveals the role of the branch point sequence-to-NAGNAG region in 3' tandem splice site selection. *Mol Cell Biol* **27**(16): 5835-5848.

Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *Rna* **14**(5): 802-813.

Wu J, Akerman M, Sun S, McCombie WR, Krainer AR, Zhang MQ. 2011. SpliceTrap: a method to quantify alternative splicing under single cellular conditions. *Bioinformatics* **27**(21): 3010-3016.

Zahler AM. 2012. Pre-mRNA splicing and its regulation in *Caenorhabditis elegans*. *WormBook*: 1-21.

Zhang H, Blumenthal T. 1996. Functional analysis of an intron 3' splice site in *Caenorhabditis elegans*. *Rna* **2**(4): 380-388.

Zhang M, Zamore PD, Carmo-Fonseca M, Lamond AI, Green MR. 1992. Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc Natl Acad Sci U S A* **89**(18): 8769-8773.

Zorio DA, Blumenthal T. 1999. Both subunits of U2AF recognize the 3' splice site in *C. elegans*. *Nature* **402**: 835-838.

Chapter 3: Alternative Splicing Regulation of Cancer-Related Pathways in *Caenorhabditis elegans*: An In Vivo Model System with a Powerful Reverse Genetics Toolbox

3.1 INTRODUCTION

Since the early 1960s with the efforts of Sydney Brenner, the nematode Caenorhabditis elegans has been established as a popular model organism in developmental biology and neurobiology. There are many biological advantages that make it an attractive system for several fields of research. The adult worm contains 959 somatic cells, making its anatomy relatively simple. Experiments in the late 1970s showed that it has an invariant cell lineage during establishment of the somatic tissues (Sulston and Horvitz 1977). It has two sexes, a self-fertilizing hermaphrodite and males, allowing genetic crosses to be performed. C. elegans has a short life cycle of less than three days and each hermaphrodite produces about 300 progeny by self-fertilization or up to 1000 progeny from cross progeny with males. The gonad is a relatively large organ in this animal, allowing for studies of organogenesis, cell proliferation, meiosis, and embryogenesis. During its development, the hermaphrodite worms produce sperm at one stage of their life cycle before switching to produce oocytes. The molecular pathways of this sperm to oocyte transition have been studied extensively (Zanetti et al. 2012). Its complete genome, sequenced in 1998, was the first sequenced genome from a multicellular organism. It has a genome size of 97 megabases containing close to 19,000 protein coding genes (*C. elegans*

Sequencing Consortium 1998). Genome-wide alignments with other five related nematodes are now available for any comparative genomics approach (Yook et al. 2012). The modENCODE project systematically generated genome-wide data from transcriptome profiling, transcription factor-binging sites, and maps of chromatin organization to improve genome annotation (Gerstein et al. 2011). When comparing the C. elegans genome to higher eukaryotes, it was found that close to 40% of the genes that have been associated with diseases in humans have worm orthologues, and cancer is not the exception (Culetto and Satelle 2000).

3.2 ALTERNATIVE SPLICING PREVALENCE

In 1990, the first report of an alternative splicing event in the C. elegans genome was published. Interestingly the event corresponds to the PKA mRNA, a kinase implicated in the onset and progression of several cancers (Gross et al. 1990). Since then, several groups have used different approaches to predict the percentage of genes in the C. elegans genome that undergo alternative splicing. Initial estimates using a limited number of Expressed Sequence Tags (ESTs) predicted fewer than 1000 genes to be alternatively spliced Kent and Zahler 2000). By normalizing the occurrence of alternative splicing, taking into account the coverage of ESTs, it was estimated that close to 10% of the C. elegans genome undergoes alternative splicing (Brett et al. 2002). More recent analysis using next-generation sequencing identified 8,651 putative novel splicing events, suggesting that up to 25% of genes have an alternative splicing event (Figure

75

1(a)) (Ramani et al. 2011). While this percentage is far from the >90% reported for the human genome, it does show that alternative splicing in C. elegans is not an uncommon mechanism to generate protein diversity.

3.3 EFFECTS OF GENOME COMPACTION ON AS

The worm genome appears to be under selective pressure to promote a reduction in genome size (Denver et al. 2004). This natural selection towards a small genome can be seen in features like short intergenic regions, short UTRs, and small introns (Kent and Zahler 2000, Jan et al. 2012, Hiller et al. 2005). For example, in humans, the median size of introns in the coding sequences is 1,334 bases, while in worms the average intron is just 65 bases (Figure 1(b)) (Yook et al. 2012, Hong et al. 2006). More than 20 years ago, the small size of C. elegans introns was already the subject of study. It was demonstrated that a short 53 nucleotide worm intron could not be efficiently spliced in human extracts, while an expansion of this intron with 31 extra nucleotides allowed for efficient splicing (Ogg et al. 1990). In spite of this reduction in intron size, the worm spliceosome is still capable of removing big introns (144 introns in the C. elegans genome are bigger than 10 kb (Figure 1(c)) (Yook et al. 2012). Larger intron size has been correlated with alternative splicing in several species (Kandul and Noor 2009). Worms also have complex patterns of alternative splicing where multiple exons in the same gene are alternatively spliced to generate multiple isoforms (Figure 1(d)). This means that the information content of a worm intron is on average greater than in higher eukaryotes (higher density of

functional elements in introns). This makes the molecular dissection of worm introns easier to achieve. By using genomic alignments between two Caenorhabditis species, the identification of novel intronic elements important for alternative splicing regulation has been described (Kabat et al. 2006). Other groups have also used comparative genomics together with UV cross-linking and Electrophoretic Mobility Shift Assays (EMSA) to identify cis-elements important for alternative splicing regulation (Kuroyanagi et al. 2007).

3.4 EVIDENCE THAT AS IN NEMATODES IS UNDER STABILIZING SELECTION

Several studies comparing alternative splicing events in mammals and insects concluded that a high percentage of the events are not conserved and are species specific (Modrek and Lee 2003, Malko et al. 2006). This high variability in alternative splicing makes it necessary to validate the functionality of the events studied. To test whether the smaller percentage of genes with alternative splicing in worms also follows these patterns of high variability during evolution, several groups have measured the levels of conservation between different C. elegans populations or between related species (Barberan-Soler and Zahler 2008a, Irima et al. 2008). In comparison to the findings in higher eukaryotes, the regulation of alternative splicing in natural populations of nematodes appears to be under strong stabilizing selection with low intra- an interspecies variability. These results point to an essential intrinsic characteristic of alternative splicing in worms: its functionality. The detection of alternative splicing in a C. elegans

transcript has a higher probability of being a functional and regulated event than in other systems.

3.5 ALTERNATIVE SPLICING REGULATION

A proxy for the functionality of an alternative splicing event that has been used for other systems is its regulation. If a particular event is detected as regulated across different conditions or during development then the possibility that the isoforms have specific functions is higher. In the last five years C. elegans joined other species in terms of the detection of changes in alternative splicing at genome-wide levels. Initially with the use of splicing-sensitive microarrays and later with the use of next-generation sequencing the regulation of alternative splicing in worms has been studied in detail across different conditions and mutations (Ramani et al. 2011, Barberan-Soler and Zahler 2008b, Ramani et al. 2009). Initial measurement of changes in splicing during development demonstrated that up to 40% of the events detected are regulated (>2 fold) between different stages of worms (Figure 1(a)) (Barberan-Soler and Zahler 2008b). This result was further validated with tiling arrays and next-generation sequencing (Ramani et al. 2009). Several examples of tissue-specific splicing are known. The alternative splicing regulation of a neuron-specific exon of unc-32, the worm a subunit of V0 complex of vacuolar-type H+-ATPases, was recently characterized (Kuroyanagi et al. 2013). A male-specific isoform of unc-55, a transcription factor, has also been reported (Shan and Walthall 2008). While individual examples of tissue-specific splicing are known, a genome-wide

78

analysis of tissue-specific splicing is still missing. With the increased sensitivity of next-generation sequencing techniques, together with the availability of manual or molecular dissections that allow the isolation of tissue-specific mRNA (mRNA-tagging; Pauli et al. 2006) a complete catalog of tissue-specific splicing should be possible in the near future.

While a big set of splicing events produce two protein isoforms, another important group introduces premature termination codons (PTC) to one of the isoforms. It is known that the introduction of a PTC to one isoform targets it to nonsense-mediated decay (NMD) (reviewed in Kervestin and Jacobson 2012). NMD mutants were first described in C. elegans almost 25 years ago (Hodgkin et al. 1989). Contrary to other systems where mutations in NMD factors are lethal, in worms, null mutations for all the seven core factors of NMD are viable (smg-1 to smg-7). This has facilitated the study of AS coupled to NMD in worms. Some of the first events of wild-type transcripts that were shown to be regulated by NMD are splicing factors in the C. elegans genome (Morrison et al. 1997). An important discovery concerning the regulation of alternative splicing in worms is that between 20–35% of the events in the genome appear to be targets of NMD (Figure 1(a)) (Ramani et al. 2009, Barberan-Soler et al. 2009). The conclusion from these studies is, then, that alternative splicing in C. elegans is not just a generator of protein diversity but also an important regulator that fine-tunes gene expression levels by targeting specific isoforms for degradation. Furthermore, it has been proposed that NMD in worms can also be regulated

with the potential to stabilize particular NMD targets allowing them to be translated into truncated proteins with putative dominant negative functions (Barberan-Soler et al. 2009).

3.6 POWERFUL REVERSE GENETICS TO DISSECT ALTERNATIVE SPLICING

One of the advantages of C. elegans as a model system is the availability of powerful tools for reverse genetics. Any laboratory can obtain stable mutants for many genes from the Caenorhabditis Genetics Center (CGC) at The University of Minnesota. The National Bioresource Project in Japan runs a program where mutants for a gene of interest can be requested and they are obtained at the facility by a protocol involving random mutagenesis with TMP/UV. These two centers allow for any group to obtain mutant strains for the gene of interest in an inexpensive and expedited manner. Recently, a more ambitious project to create a million different mutants has been performed by the Moerman and Waterston labs (Thompson et al. 2013). The aim of this project was to use next-generation sequencing and a collection of 2000 mutagenized strains to identify multiple mutations in virtually all the genes in the worm genome. This project will likely allow the study of mutant worms that have mutations in isoform-specific regions. All these resources allow for the use of stable mutants to characterize the roles of either putative splicing factors or more specifically of particular isoforms. This approach to screen the effects that mutations in different splicing factors have on a particular splicing event has been used before (Barberan-Soler

et al. 2011). The conclusion from this work was that the coregulation of alternative splicing by diverse factors is a common phenomenon in worms.

Another great resource for the worm community is the availability of genome-wide RNAi libraries (Kamath et al. 2003). Researchers aiming to characterize genetic pathways have used these libraries for genome-wide screens. Some of them have found that several splicing-related components have interesting phenotypes when targeted by RNAi (Ceron et al. 2007). More focused screens with a subset of clones from these libraries are used to characterize a specific group of genes. For example, Kerins et al.. used an RNAi screen of all the predicted C. elegans splicing factors and found that many of them showed an overproliferation phenotype in a sensitized germline background (Kerins et al. 2010).

3.7 IN VIVO ALTERNATIVE SPLICING REPORTERS

One of the most advantageous tools for the study of alternative splicing in C. elegans is the use of bichromatic alternative splicing reporters (Kuroyanagi et al. 2006). This transgenic reporter system allows the visualization of splicing events by tagging each one of the isoforms with a different fluorescent protein. A worm population with differences in splicing regulation can then be sorted using FACS cytometry and worms with a particular splicing ratio obtained. This together with the use of mutagenesis or RNAi allows performing an in vivo screen for regulators of any splicing event of interest. The advantage of these reporters in worms is that the characterization of an alternative splicing event

81

can be performed in vivo. Different events that have been studied with this technology are egl-15, let-2, unc-60, and unc-32 (Kuroyanagi et al. 2013, Kuroyanagi et al. 2006, Ohno et al. 2008, Ohno et al. 2012, Kuroyanagi et al. 2013). This technology has allowed the identification of splicing factors that regulate these events, as well as the cis-elements that are necessary for its regulation.

3.8 *C. ELEGANS* AS REGULATION OF CANCER-RELATED PATHWAYS

Given the diversity of tools available to perform studies in C. elegans, researchers have set out to understand details about cancer that were long a mystery. Questions surrounding cell proliferation and cell-cell communication as well as the cellular and molecular components that make up a stem cell niche often must be answered in a multicellular context. Furthermore, the fact that biological processes and factors active in splicing and human cancers are almost wholly homologous with those in C. elegans makes the move to worm studies rewarding. Disrupted cancer pathways in C. elegans, for the most part, have clear, simple, and observable phenotypes that extend beyond just cell death or proliferation. Using these pathways as tools and read-outs, splicing factors and splicing events have been identified, in C. elegans, to interact with, regulate, and cooperate with homologous pathways that lead to cancer in humans (Figure 2).

Extensive Splicing of Ras/let-60 Pathway Components Contributes to Cell Fate Determination and Proliferation

Overwhelming evidence links cell signaling and gene expression changes in the promotion of human cancers. Alternative splicing plays a major role in defining the activity of signal transduction pathways such as Ras/let-60. Strikingly, of the 50–60 genes that are known components of, regulate or, are directly regulated by the Ras/let-60 pathway in C. elegans (Sundaram 2006), approximately half of them produce multiple isoforms differing by at least one alternative exon (Yook et al. 2012). This highlights the importance of proper gene expression through post-transcriptional regulation (Table 1).

The ligand-dependent Ras pathway in C. elegans stimulates the induction of vulval precursor cells (VPCs) in the hypodermis to divide and differentiate into a functional vulva during larval development (Kimble 1981, Sternberg and Horvitz 1986). Increased signaling from the Ras pathway leads to a multivulval (Muv) phenotype, while decreased signaling causes a vulvaless (Vul) phenotype Han and Sternberg 1990). These simple and observable phenotypes provide an excellent system to ascertain the contributions of alternative splicing to Ras signaling activity. The C. elegans homolog of the Ras-MAPK pathway stimulant Epidermal Growth Factor, lin-3, produces several isoforms with unique characteristics. LIN-3 isoform specialization of function has previously been observed in pathways aside from Ras-mediated cell signaling and has been found to differentially mediate growth rates, feeding behavior, and cellular quiescence (van Buskirk and Sternberg 2007). In VPCs, the activity of the LIN-3L isoform is dependent on interaction with the C. elegans Rhomboid protease

homolog, ROM-1, while the activity of LIN-3S is not (Dutt et al. 2004). LIN-3S is expressed in a specialized Anchor Cell (AC) responsible for inducing a primary set of VPCs during larval development, and LIN-3L expression and possible secretion by primary VPCs themselves act as a ligand stimulant of the Ras pathway in secondary VPCs (Kimble 1981, Sternberg and Horvitz 1986). This stimulation may cause the secondary VPCs to switch their cell fate from hypodermis to vulva (Sternberg and Horvitz 1986). It is unknown if coordinated expression of these isoforms from distinct cell types establishes a gradient through Ras signaling that induces nearby VPCs to proceed through necessary cell divisions to develop into a proper vulva.

RasGAPs are GTPase-activating proteins that specifically promote the hydrolysis of Ras-bound GTP molecules, thereby inactivating the Ras molecule and its signaling (Trahey and McCormick 1987). gap-2, in C. elegans, is similar to the p120 Ras-GAP family (Trahey et al. 1988). The gene contains 25 exons that use alternative splicing and transcription start sites to produce 9 mRNA products (Hayashizaki et al. 1998). These isoforms were identified by a rapid amplification of cDNA end (5′RACE) technique using primers specific to a sequence common to many mature mRNAs in C. elegans, the SL1 trans-splice leader. This 22nt sequence is naturally spliced onto 5′ ends of transcripts from ~70% of genes (Allen et al. 2011). In this case, they served as 5′ primer binding sites in reverse transcription reactions followed by PCR to identify isoforms with alternative 5′ ends and promoters (Dutt et al. 2004). Differential expression of

gap-2 isoforms in several tissue types was revealed by expression of transgenes containing these alternative gap-2 promoters fused to GFP in C. elegans. The function of RasGAP proteins from various genes in other organisms (Iwashita and Song 2008) may be relegated to multiple isoforms from a small number of RasGAP genes producing many isoforms in C. elegans. While the effects of this dynamic expression pattern of isoforms on molecular signaling pathways such as the Ras pathway are difficult to tease apart, target sites for PKC, PKA, and protein tyrosine kinases present or absent in each isoform (Hayashizaki et al. 1998) may provide clues to individual isoform activities.

Redundant Retinoblastoma/lin-35 Pathways Are Populated by Splicing Factors

Homozygous mutations of the retinoblastoma (Rb) gene have been shown to promote retinal cancer, small-cell lung carcinomas, and osteosarcomas (Knudson 1978). In worms, a single homolog of Rb, lin-35, represses the multivulval phenotype observed in Ras pathway hyper-signaling mutants (Lu and Horvitz 1998). Two redundant pathways, termed synthetic multivulval (synmuv) class A and class B, are active in the VPCs and induce changes in expression of genes that lead to a multivulval fate through histone modification and chromatin remodeling (Lu and Horvitz 1998, Ceol and Horvitz 2001). To induce a phenotype, disruptive mutations in single genes within each class must be present. Interestingly, mutations in the lin-15 locus created alleles with Muv, class A synmuv, and class B synmuv phenotypes, leading to speculation that two

genes existed at the locus with distinct functions in each pathway. In RNAi screens for interactors of the class B synmuv pathway in C. elegans, 57 genes were found to enhance the Muv phenotype observed in lin-35 or other synmuv class B mutant lines (Ceron et al. 2007, Ferguson and Horvitz 1989). Ten of these potential interactors (rsr-2, lsm-2/gut-2, lsm-4, snr-1–7) have roles in splicing. Their synmuv B mutant phenotypes are dependent on Ras activity and lin-3 ligand binding. While the activities, interactions, and mutant phenotypes of most of these splicing-related Rb enhancers are not well understood, RNAi of some of the snr genes (snr-1, snr-2, snr-4, snr-5, snr-6, and snr-7) led to embryonic lethality and nuclear pore organization disruption [56]. The question of whether these phenotypes arise from disrupted functions that have no role in splicing regulation is still unanswered, but evidence is mounting supporting the idea that constitutive splicing and alternative splicing are major contributors in C. elegans to pathways that, when altered, lead to cancer in human tissues. Undoubtedly, future studies in C. elegans will enrich current knowledge concerning the interwoven activities of splicing and the Rb/lin-35 pathway.

Protein Kinase A Isoform Diversity Replaces the Need for Multiple Genetic Loci

Protein kinase A (PK-A) is involved in many cellular processes and has been implicated in several cancers (Merkle and Hoffmann 2011). The mammalian PK-A is composed of two distinct subunits, the regulatory and catalytic, that are each interchangeable with protein subunits from several genetic loci. The two

subunits of PK-A are conserved in C. elegans (Gross et al. 1990) but derived from possibly only three genes, kin-1, kin-2, and F47F2.1. Interestingly, the modularity of the subunits that make up PK-A in C. elegans has been conserved through alternative usage of exons. The catalytic subunit (C-subunit) alone is thought to undergo alternative splicing at both N- and C-termini to create at least 12 different isoforms (Tabish et al. 1999). These isoforms are conserved in C. briggsae, a relative of C. elegans, and show differential expression during development, suggesting that the diversity of potential subunits itself has functional importance (Bowen et al. 2006). An isoform of the C-subunit containing the N′1 exon harbors a myristoylation site and is highly expressed in eggs, where an alternative exon N′4 lacking the myristoylation site is highly expressed in adult worms (Bowen et al. 2006). The presence of N′1 protein isoform expression and myristoylation was found to affect substrate targeting of PK-A but not the catalytic activity of the enzyme itself (Clegg et al. 2012). Similarly, isoforms of the regulatory (R-subunit) have recently been identified and found to contain or lack domains necessary for docking to A-kinase anchor proteins and the C-subunit of the PK-A holoenzyme (Pastok et al. 2013). A second gene on the X chromosome of C. elegans, F47F2.1, has homology to murine Cα subunits (Tabish et al. 1999). Like the kin-1 N′4 isoform, expression of the longer F47F2.1b isoform is low in eggs but high in mixed populations of worms (Bowen et al. 2006). A truncated isoform, F47F2.1a, lacks amino acids important for ATP-binding, and knockdown of N′4 kin-1 isoform has no obvious

phenotype leading to speculation that some of these isoforms may be redundant or unnecessary (Bowen et al. 2006, Murray ey al 2008).

Topoisomerase-1 Isoforms Have Differential Expression and Potential SR Kinase Activity

Isoforms from the same genetic locus often have distinct expression patterns and functions that are seemingly unrelated. C.elegans DNA topoisomerase I (top-1) produces two isoforms that vary in temporal and spatial expression patterns. Cellular and subcellular localization differences between the isoforms lead to speculation they may function in different processes. The TOP-1β isoform skips exon 2 and is more ubiquitously expressed throughout C. elegans development, being found in multiple cell types and all stages from embryo through adulthood. Conversely, the inclusion isoform, TOP-1α, is detectably present in embryos and, interestingly, in neurons at the comma stage, then decreases in abundance as worms enter larval stages (Lee et al. 1998). Isoform-specific immuno-histochemical localization assays in germline cells identified TOP-1β concentration in nucleoli and TOP-1α concentration at centrosomes and on chromosomes (Lee et al. 2001, Cha et al. 2012). Expression of a GFP transgene driven by the topoisomerase promoter was also detected strongly in the distal tip cell of L3/L4 worms indicating that somatic gonad sheath development and/or germline stem cell proliferation may be regulated by TOP-1 activity. RNAi of C. elegans top-1 reduced the number of germ cells in the mature germline by anywhere from 50–100% (Cha et al. 2012). It is currently unknown

how top-1 isoforms contribute to the maintenance of stem cell proliferation or exactly which isoform would be performing such functions. Localization of the inclusion isoform, TOP-1α, to centrosomes and chromosomes suggests it could be involved in chromosome segregation, in regulation of transcription, or possibly in posttranscriptional regulation. DNA topoisomerase I was identified as an SR protein kinase in HeLa cell extracts (Rossi et al. 1996). It was found to phosphorylate splicing factors involved in cell cycle regulation, such as SF2/ASF (Labourier et al. 1998, Dubatolova et al. 2013). It may be possible that these isoforms are working more closely than originally thought if one or both of the top-1 isoforms affect the cell cycle and carry the same SR kinase activity as shown in HeLa cells. The activity of one or both of the TOP-1 isoforms as SR protein kinases suggests that top-1 alternative splicing may be autoregulated. Kinase activity of TOP-1 and its autoregulation have yet to be determined in worms.

Splicing Factors as Regulators of Cell Proliferation and Differentiation

Several genes involved in splicing regulation have been implicated in the proliferation/differentiation decision in the C. elegans germline. glp-1(ar202gf) worms were used to perform an RNAi screen searching for genes that led to increased cell proliferation in the germline. The resulting genes included factors involved in every step of the splicing process (spliceosome construction initiation, reorganization of the snRNP complex and removal of the lariat; Kerins et al. 2010). prp-17, for example, encodes an ortholog of yeast CDC40 and human

prp17 and is involved in the 2nd catalytic step of intron removal in the spliceosome (Kerins et al. 2010, Kassir and Simchen 1978). RNAi of prp-17 in the rrf-1(pk1417);glp-1(oz264) background enhanced germline tumors, suggesting that it is involved in the proliferation versus differentiation decision. Similarly, the mog genes represent C. elegans homologs of core yeast splicing factors that regulate the pathway downstream of glp-1 to promote differentiation and the oocyte fate of maturing germ cells (Graham et al. 1993, Puoti and Kimble 1999, Puoti and Kimble 2000). The U2 snRNP-associated splicing complex SF3b is essential for splicing (Caspary et al. 1999, Dziembowski et al. 2004). TEG-4 is the worm homolog of human SF3b subunit 3 (aka SAP130) that increases excessive cell proliferation in the glp-1(ar202gf) background (Mantina et al. 2009). Interestingly, epistasis experiments attempting to determine the genetic relationship between teg-4 and the glp-1/Notch signaling pathway in various cell types were inconclusive. Human CD2BP2 was suggested to regulate splicing via U4_U5_U6 tri-snRNP formation (Laggerbauer et al. 2005). TEG-1, the C. elegans homolog of CD2BP2 has been found to bind UAF-1, the worm U2AF large subunit homolog, suggesting that it could work in two important steps in the mechanism of pre-mRNA splicing (Wang et al. 2012). teg-1(oz230), in combination with a glp-1 mild gain-of-function mutant background, produces a tumorous germline phenotype as well. To what extent splicing factors, themselves, have individual genetic interactions within the proliferation versus

differentiation pathways or if global splicing regulation is more indirectly influential is still unknown.

CED-4 Isoforms Promote and Inhibit Systematic Apoptosis during Development

Disruption of alternative splicing patterns may affect the onset of programmed cell death so intricately regulated in C. elegans. The worm homolog of APAF1, ced-4, was first identified as a core factor in the developmental induction of neuronal programmed cell death (Ellis and Horvitz 1986). ced-4 physically links ced-3, a member of a caspase family of proteases to ced-9, a cell death suppressor (Wu et al. 1997, Chinnaiyan et al. 1997, Seshagiri and Miller 1997). As a regulatory switch, ced-4 is a core determinant of the decision of a cell to be systematically culled or not. ced-4 encodes two protein isoforms, CED-4L and CED-4S, differing by 24 amino acids at the 5′ end of exon 4. While CED-4S normally promotes programmed cell death, overexpression of the CED-4L isoform led to ectopic cell growth and rescue of lethal phenotypes seen in ced-9 loss-of-function mutants (Shaham and Horvitz 1996, Hengartner et al. 1992). It is CED-4L association with CED-3 that surprisingly inhibits cell death in a dominant negative manner. The presence of the longer P-loop in the long isoform still allows for association of CED-4L with the protease domain of CED-3 but disrupts the association of CED-4L with the prodomain of CED-3, which normally contributes to its activation (Chaudhary et al. 1998). Studies in mammalian cells have similarly identified ced-3 and ced-4 homologous genes

that produce isoforms with opposing cell death functions (Wang et al. 1994, Jiang et al. 1998). An imbalance in the concentrations of these isoforms or in the abundance or activities of specific SR or hnRNP proteins often leads to one developmental cell fate over the other (Moore et al. 2010). Complex regulation of alternative splicing factors, therefore, may direct developmental apoptosis programs. Loss-of-function of an SR protein kinase, spk-1, in a ced-4 partial loss-of-function background, as well as mutant alleles of several SR proteins themselves, leads to an increase in apoptosis (Galvin et al. 2011), suggesting that control of alternative splicing factor activity may play a direct or indirect role in the regulation of developmentally programmed apoptosis. In vitro studies have shown that SPK-1 can bind and phosphorylate alternative splicing factors like SF2 and RNAi-mediated knockdown of spk-1 in C. elegans causes embryonic lethality and germline development defects (Kuroyanagi et al. 2000, Longman et al. 2000). Experiments aimed at understanding the complex code that dictates what genetic material is included or skipped in mature mRNA and subsequent protein products will undoubtedly clarify further the impact that alternative splicing has in the mediation of programmed cell death through factors such as ced-4.

Research Focusing on the Links between Splicing and Cancer in C. elegans Is Bright

Great strides have been taken in research to understand the involvement of splicing factors in biological processes such as cell proliferation, differentiation,

migration, communication, and death (as discussed previoussly). RNAi screens have identified splicing factors that promote or inhibit cell proliferation. Transgenic assays have revealed cellular and subcellular localization of specific isoforms that may be involved in apoptotic or cell cycle regulation. Transparent cuticles and eggs make worms ideal specimens for the use of fluorescent reporters. Lineage tracing experiments have not only identified every cell in an adult worm but also precursors and cell behavior throughout development. Furthermore, only in a multicellular organism, such as C. elegans, can the effects of cell-to-cell signaling on processes such as differentiation or cell proliferation become apparent. High-throughput sequencing and bioinformatic analyses are opening a new chapter on comparisons in isoform usage between genetic backgrounds. The modEncode project seeks to identify functional elements within the Drosophila and Caenorhabditis genomes by providing access to high-quality gene expression datasets. Furthermore, the million-mutation project in C. elegans will allow researchers to obtain gene and even isoform-specific mutant strains for further study. Given the speed and depth of discoveries using methods such as RNAi in worms, this project should make C. elegans a core model to study alternative splicing and its link to genes homologous to those tied to cancer in humans.

3.9 ACKNOWLEDGEMENTS

Figure 1. From: Alternative Splicing Regulation of Cancer-Related Pathways in Caenorhabditis elegans: An In Vivo Model System with a Powerful Reverse Genetics Toolbox.

Caenorhabditis elegans alternative splicing events. (a) Genome-wide analysis of alternative splicing in C. elegans; (b) comparison of the human APAF1 and the C. elegans homolog ced-4 gene models shows significant differences in intron size between species for genes with important alternative splicing events; (c) Y48C3A.5 intron 4 (19,927 bp) is one example of the 144 introns in the C. elegans genome that are more than 10 kb in length; (d) C. elegans unc-52 gene undergoes complex alternative splicing that generates at least 12 different isoforms by the use of nine different cassette exons.

Figure 1.

(a)

Human APAF1, median intron size 2746 bp, median exon size 132 bp

(b)

C. elegans ced-4, median intron size 50 bp, median exon size 181 bp

(c)

C. elegans Y48C3A.5

Intron 4 size 19,927 bp

(d)

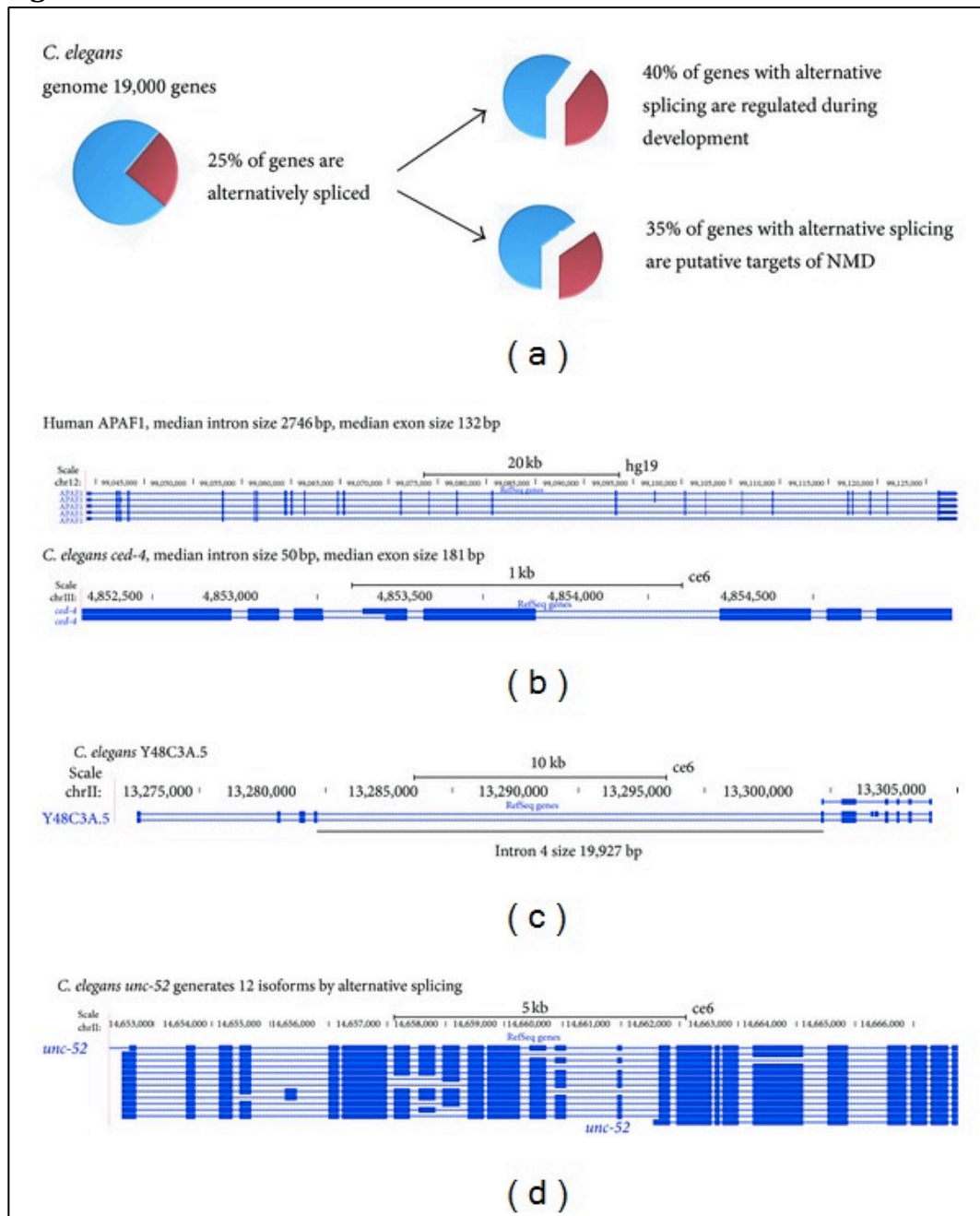*C. elegans* unc-52 generates 12 isoforms by alternative splicing

Figure 2. From: Alternative Splicing Regulation of Cancer-Related Pathways in Caenorhabditis elegans: An In Vivo Model System with a Powerful Reverse Genetics Toolbox.

Connections between alternative splicing in C. elegans and pathways homologous to those that cause excessive cell proliferation or apoptosis in humans, as reviewed in this paper.
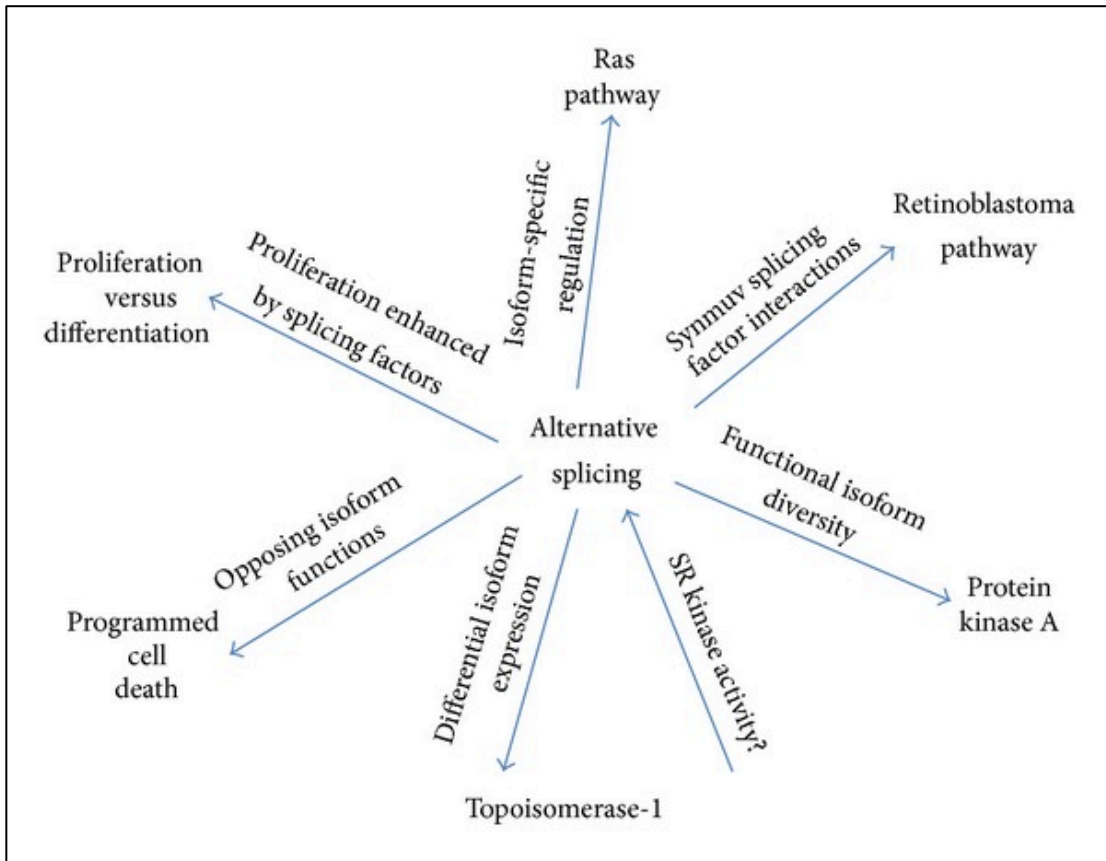
3.10 REFERENCES

Allen MA, Hillier LW, Waterston RH, Blumenthal T. A global analysis of C. elegans trans-splicing. Genome Research. 2011;21(2):255–264.

Barberan-Soler S, Lambert NJ, Zahler AM. Global analysis of alternative splicing uncovers developmental regulation of nonsense-mediated decay in C. elegans. RNA. 2009;15(9):1652–1660.

Barberan-Soler S, Medina P, Estella J, Williams J, Zahler AM. Co-regulation of alternative splicing by diverse splicing factors in Caenorhabditis elegans. Nucleic Acids Research. 2011;39(2):666–674.

Barberan-Soler S, Zahler AM. Alternative splicing and the steady-state ratios of mRNA isoforms generated by it are under strong stabilizing selection in Caenorhabditis elegans. Molecular Biology and Evolution. 2008;25(11):2431–2437.

Barberan-Soler S, Zahler AM. Alternative splicing regulation during C. elegans development: splicing factors as regulated targets. PLoS Genetics. 2008;4(2)e1000001.

Bowen LC, Bicknell AV, Tabish M, Clegg RA, Rees HH, Fisher MJ. Expression of multiple isoforms of the cAMP-dependent protein kinase (PK-A) catalytic subunit in the nematode, Caenorhabditis elegans. Cellular Signalling. 2006;18(12):2230–2237.

Brett D, Pospisil H, Valcárcel J, Reich J, Bork P. Alternative splicing and genome complexity. Nature Genetics. 2002;30(1):29–30.

C elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science. 1998;282(5396):2012–2018.

Caspary F, Shevchenko A, Wilm M, Séraphin B. Partial purification of the yeast U2 snRNP reveals a novel yeast pre-mRNA splicing factor required for pre-spliceosome assembly. EMBO Journal. 1999;18(12):3463–3474.

Ceol CJ, Horvitz HR. dpl-1 DP and efl-1 E2F act with lin-35 Rb to antagonize Ras signaling in C. elegans vulval development. Molecular Cell. 2001;7(3):461–473.

Ceron J, Rual J-F, Chandra A, Dupuy D, Vidal M, Van Den Heuvel S. Large-scale retinoblastoma pathway as well as splicing-related components with synMuv B activity. BMC Developmental Biology. 2007;7, article 30.

Cha DS, Hollis SE, Datla US, et al.. Differential subcellular localization of DNA topoisomerase-1 isoforms and their roles during Caenorhabditis elegans development. Gene Expression Patterns. 2012;12(5):189–195.

Chaudhary D, O'Rourke K, Chinnaiyan AM, Dixit VM. The death inhibitory molecules CED-9 and CED-4L use a common mechanism to inhibit the CED-3 death protease. Journal of Biological Chemistry. 1998;273(28):17708–17712.

Chinnaiyan AM, Chaudhary D, O'Rourke K, Koonin EV, Dixit VM. Role of CED-4 in the activation of CED-3. Nature. 1997;388(6644):728–729.

Clegg RA, Bowen LC, Bicknell AV, et al.. Characterisation of the N'1 isoform of the cyclic AMP-dependent protein kinase (PK-A) catalytic subunit in the nematode, Caenorhabditis elegans. Archives of Biochemistry and Biophysics. 2012;519(1):38–45.

Culetto E, Sattelle DB. A role for Caenorhabditis elegans in understanding the function and interactions of human disease genes. Human Molecular Genetics. 2000;9(6):869–877.

Denver DR, Morris K, Lynch M, Thomas WK. High mutation rate and predominance of insertions in the Caenohabditis elegans nuclear genome. Nature. 2004;430(7000):679–682.

Dubatolova TD, Volkova EI, Omelyanchuk LV. Drosophila splicing factor SF2 knock-down mutant shows altered cell-cycle in vivo. Cell Biology International. 2013;37(2):187–190.

Dutt A, Canevascini S, Froehli-Hoier E, Hajnal A. EGF signal propagation during C. elegans vulval development mediated by ROM-1 Rhomboid. PLoS Biology. 2004;2(11, article e334).

Dziembowski A, Ventura A-P, Rutz B, et al.. Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. EMBO Journal. 2004;23(24):4847–4856.

Ellis HM, Horvitz HR. Genetic control of programmed cell death in the nematode C. elegans. Cell. 1986;44(6):817–829.

Ferguson EL, Horvitz HR. The multivulva phenotype of certain Caenorhabditis elegans mutants results from defects in two functionally redundant pathways. Genetics. 1989;123(1):109–121.

Galvin BD, Denning DP, Horvitz HR. SPK-1, an SR protein kinase, inhibits programmed cell death in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(5):1998–2003.

Gerstein MB, Lu ZJ, Van Nostrand EL, et al.. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE Project. Science. 2011;330(6012):1775–1787.

Graham PL, Schedl T, Kimble J. More mog genes that influence the switch from spermatogenesis to oogenesis in the hermaphrodite germ line of Caenorhabditis elegans. Developmental Genetics. 1993;14(6):471–484.

Gross RE, Bagchi S, Lu X, Rubin CS. Cloning, characterization, and expression of the gene for the catalytic subunit of cAMP-dependent protein kinase in Caenorhabditis elegans. Identification of highly conserved and unique 1990;265(12):6896–6907.

Han M, Sternberg PW. let-60, a gene that specifies cell fates during C. elegans vulval induction, encodes a ras protein. Cell. 1990;63(5):921–931.

Hayashizaki S, Iino Y, Yamamoto M. Characterization of the C. elegans gap-2 gene encoding a novel Ras-GTpase activating protein and its possible role in larval development. Genes to Cells. 1998;3(3):189–202.

Hengartner MO, Ellis RE, Horvitz R. Caenorhabditis elegans gene ced-9 protects cells from programmed cell death. Nature. 1992;356(6369):494–499.

Hillier LW, Coulson A, Murray JI, Bao Z, Sulston JE, Waterston RH. Genomics in C. elegans: so many genes, such a little worm. Genome Research. 2005;15(12):1651–1660.

Hodgkin J, Papp A, Pulak R, Ambros V, Anderson P. A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics. 1989;123(2):301–313.

Hong X, Scofield DG, Lynch M. Intron size, abundance, and distribution within untranslated regions of genes. Molecular Biology and Evolution. 2006;23(12):2392–2404.

Irimia M, Rukov JL, Penny D, Garcia-Fernandez J, Vinther J, Roy SW. Widespread evolutionary conservation of alternatively spliced exons in Caenorhabditis. Molecular Biology and Evolution. 2008;25(2):375–382.

Iwashita S, Song S-Y. RasGAPs: a crucial regulator of extracellular stimuli for homeostasis of cellular functions. Molecular BioSystems. 2008;4(3):213–222.

Jan CH, Friedman RC, Ruby JG, Bartel DP. Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. Nature. 2011;469(7328):97–103.

Jiang Z-H, Zhang W-J, Yi R, Wu JY. Regulation of Ich-1 pre-mRNA alternative splicing and apoptosis by mammalian splicing factors. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(16):9155–9160.

Joseph-Strauss D, Gorjánácz M, Santarella-Mellwig R, Voronina E, Audhya A, Cohen-Fix O. Sm protein down-regulation leads to defects in nuclear pore complex disassembly and distribution in C. elegans embryos. Developmental Biology. 2012;365(2):445–457.

Kabat JL, Barberan-Soler S, McKenna P, Clawson H, Farrer T, Zahler AM. Intronic alternative splicing regulators identified by comparative genomics in nematodes. PLoS Computational Biology. 2006;2(7, article e86).

Kamath RS, Fraser AG, Dong Y, et al.. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature. 2003;421(6920):231–237.

Kandul NP, Noor MAF. Large introns in relation to alternative splicing and gene evolution: a case study of Drosophila bruno-3. BMC Genetics. 2009;10, article 1471:p. 67.

Kassir Y, Simchen G. Meiotic recombination and DNA synthesis in a new cell cycle mutant of Saccharomyces cerevisiae. Genetics. 1978;90(1):49–68.

Kent WJ, Zahler AM. The intronerator: exploring introns and alternative splicing in Caenorhabditis elegans. Nucleic Acids Research. 2000;28(1):91–93.

Kerins JA, Hanazawa M, Dorsett M, Schedl T. 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. Developmental Dynamics. 2010;239(5):1555-1572.

Kervestin S, Jacobson A. NMD: a multifaceted response to premature translational termination. Nature Reviews Molecular Cell Biology. 2012;13(11):700–712.

Kimble J. Alterations in cell lineage following laser ablation of cells in the somatic gonad of Caenorhabditis elegans. Developmental Biology. 1981;87(2):286–300.

Knudson AG., Jr. Retinoblastoma: a prototypic hereditary neoplasm. Seminars in Oncology. 1978;5(1):57–60.

Kuroyanagi H, Kimura T, Wada K, Hisamoto N, Matsumoto K, Hagiwara M. SPK-1, a C. elegans SR protein kinase homologue, is essential for embryogenesis and required for germline development. Mechanisms of Development. 2000;99(1-2):51–64.

Kuroyanagi H, Kobayashi T, Mitani S, Hagiwara M. Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms in vivo. Nature Methods. 2006;3(11):909–915.

Kuroyanagi H, Kuroyanagi H, Watanabe Y, Suzuki Y, Hagiwara M. Position-dependent and neuron-specific splicing regulation by the CELF family RNA-binding protein UNC-75 in Caenorhabditis elegans. Nucleic Acids Research. 2013;41(7):4015–4025.

Kuroyanagi H, Ohno G, Mitani S, Hagiwara M. The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo. Molecular and Cellular Biology. 2007;27(24):8612–8621.

Kuroyanagi H, Watanabe Y, Hagiwara M. CELF family RNA-binding protein UNC-75 regulates two sets of mutually exclusive exons of the unc-32 gene in neuron-specific manners in Caenorhabditis elegans. PLoS Genetics. 2013;9(2)e1003337.

Labourier E, Rossi F, Gallouzi I-E, Allemand E, Divita G, Tazi J. Interaction between the N-terminal domain of human DNA topoisomerase I and the arginine-serine domain of its substrate determines phosphorylation of

SF2/ASF splicing factor. Nucleic Acids Research. 1998;26(12):2955–
        2962.

Laggerbauer B, Liu S, Makarov E, et al.. The human U5 snRNP 52K protein
        (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP
        bridging protein, but dissociates upon tri-snRNP formation. RNA.
        2005;11(5):598–608.

Lee MH, Jang YJ, Koo H-S. Alternative splicing in the Caenorhabditis elegans DNA
        topoisomerase I gene. Biochimica et Biophysica Acta. 1998;1396(2):207–
        214.

Lee MH, Park H, Shim G, Lee J, Koo H-S. Regulation of gene expression, cellular
        localization, and in vivo function of Caenorhabditis elegans DNA
        topoisomerase I. Genes to Cells. 2001;6(4):303–312.

Longman D, Johnstone IL, Cáceres JF. Functional characterization of SR and SR-
        related genes in Caenorhabditis elegans. EMBO Journal.
        2000;19(7):1625-1637.

Lu X, Horvitz HR. Lin-35 and lin-53, two genes that antagonize a C. elegans Ras
        pathway, encode proteins similar to Rb and its binding protein RbAp48.
        Cell. 1998;95(7):981–991.

Malko DB, Makeev VJ, Mironov AA, Gelfand MS. Evolution of exon-intron
        structure and alternative splicing in fruit flies and malarial mosquito
        genomes. Genome Research. 2006;16(4):505–509.

Mantina P, MacDonald L, Kulaga A, Zhao L, Hansen D. A mutation in teg-4, which
        encodes a protein homologous to the SAP130 pre-mRNA splicing factor,
        disrupts the balance between proliferation and differentiation in the C.
        elegans germ line. Mechanisms of Development. 2009;126(5-6):417–429.

Merkle D, Hoffmann R. Roles of cAMP and cAMP-dependent protein kinase in the
        progression of prostate cancer: cross-talk with the androgen receptor.
        Cellular Signalling. 2011;23(3):507–515.

Modrek B, Lee CJ. Alternative splicing in the human, mouse and rat genomesis
        associated with an increased frequency of exon creation and/or loss.
        Nature Genetics. 2003;34(2):177–180.

Moore MJ, Wang Q, Kennedy CJ, Silver PA. An alternative splicing network links
        cell-cycle control to apoptosis. Cell. 2010;142(4):625–636.

Morrison M, Harris KS, Roth MB. smg mutants affect the expression of alternatively spliced SR protein mRNAs in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(18):9782–9785.

Murray P, Clegg RA, Rees HH, Fisher MJ. siRNA-mediated knockdown of a splice variant of the PK-A catalytic subunit gene causes adult-onset paralysis in C. elegans. Gene. 2008;408(1-2):157–163.

Ogg SC, Anderson P, Wickens MP. Splicing of a C. elegans myosin pre-mRNA in a human nuclear extract. Nucleic Acids Research. 1990;18(1):143–149.

Ohno G, Hagiwara M, Kuroyanagi H. STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo. Genes and Development. 2008;22(3):360–374.

Ohno G, Ono K, Togo M, et al.. Muscle-specific splicing factors ASD-2 and SUP-12 cooperatively switch alternative pre-mRNA processing patterns of the ADF/cofilin gene in Caenorhabditis elegans. PLoS Genetics. 2012;8(10)e1002991.

Pastok MW, Prescott MC, Dart C, et al.. Structural diversity of the cAMP-dependent protein kinase regulatory subunit in Caenorhabditis elegans. Cell Signaling. 2013;25(1):168–177.

Pauli F, Liu Y, Kim YA, Chen P-J, Kim SK. Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in C. elegans. Development. 2006;133(2):287–295.

Puoti A, Kimble J. The Caenorhabditis elegans sex determination gene mog-1 encodes a member of the DEAH-box protein family. Molecular and Cellular Biology. 1999;19(3):2189–2197.

Puoti A, Kimble J. The hermaphrodite sperm/oocyte switch requires the Caenorhabditis elegans homologs of PRP2 and PRP22. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(7):3276–3281.

Ramani AK, Calarco JA, Pan Q, et al.. Genome-wide analysis of alternative splicing in Caenorhabditis elegans. Genome Research. 2011;21(2):342–348

Ramani AK, Nelson AC, Kapranov P, Bell I, Gingeras TR, Fraser AG. High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective Caenorhabditis elegans. Genome Biology. 2009;10(9, article R101).

Rossi F, Labourier E, Forné T, et al.. Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. Nature. 1996;381(6577):80–82.

Seshagiri S, Miller LK. Caenorhabditis elegans CED-4 stimulates CED-3 processing and CED-3-induced apoptosis. Current Biology. 1997;7(7):455–460.

Shaham S, Horvitz HR. An alternatively spliced C. elegans ced-4 RNA encodes a novel cell death inhibitor. Cell. 1996;86(2):201–208.

Shan G, Walthall WW. Copulation in C. elegans males requires a nuclear hormone receptor. Developmental Biology. 2008;322(1):11–20.

Sternberg PW, Horvitz HR. Pattern formation during vulval development in C. elegans. Cell. 1986;44(5):761–772.

Sulston JE, Horvitz HR. Post embryonic cell lineages of the nematode. Caenorhabditis elegans. Developmental Biology. 1977;56(1):110–156.

Sundaram MV. RTK/Ras/MAPK signaling. WormBook. 2006:1–19.

Tabish M, Clegg RA, Rees HH, Fisher MJ. Organization and alternative splicing of the Caenorhabditis elegans cAMP-dependent protein kinase catalytic-subunit gene (kin-1) Biochemical Journal. 1999;339(1):209–216.

Thompson O, Edgley M, Strasbourger P, et al.. The Million Mutation Project: a new approach to genetics in Caenorhabditis elegans. Genome Research. 2013.

Trahey M, McCormick F. A cytoplasmic protein stimulates normal N-ras p21 GTPase, but does not affect oncogenic mutants. Science. 1987;238(4826):542–545.

Trahey M, Wong G, Halenbeck R, et al.. Molecular cloning of two types of GAP complementary DNA from human placenta. Science. 1988;242(4886):1697–1700.

van Buskirk C, Sternberg PW. Epidermal growth factor signaling induces behavioral quiescence in Caenorhabditis elegans. Nature Neuroscience. 2007;10(10):1300–1307.

Wang C, Wilson-Berry L, Schedl T, Hansen D. TEG-1 CD2BP2 regulates stem cell proliferation and sex determination in the C. elegans germ line and physically interacts with the UAF-1 U2AF65 splicing factor. Developmental Dynamics. 2012;241(3):505–521.

Wang L, Miura M, Bergeron L, Zhu H, Yuan J. Ich-1, an ice/ced-3-related gene, encodes both positive and negative regulators of programmed cell death. Cell. 1994;78(5):739–750.

Wu D, Wallen HD, Inohara N, Nuñez G. Interaction and regulation of the Caenorhabditis elegans death protease CED-3 by CED-4 and CED-9. Journal of Biological Chemistry. 1997;272(34):21449–21454.

Yook K, Harris TW, Bieri T, et al.. WormBase 2012: more genomes, more data, new website. Nucleic Acids Research. 2012;40:D735–D741.

Zanetti S, Grinschgl S, Meola M, et al.. The sperm-oocyte switch in the C. elegans hermaphrodite is controlled through steady-state levels of the fem-3 mRNA. RNA. 2012;18(7):1385–1394.