CHARACTERIZING THE EFFECT OF THE *rbm-39* GENE IN *C. ELEGANS* GERMLINE DEVELOPMENT

A THESIS

Presented to

The Faculty of the Department of Molecular Biology

The Colorado College

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Arts

By

Duffy Doyle

May/2022

Darrell J. Killian, Ph.D.
Associate Professor, Department Chair, Molecular Biology
Supervisor, Primary Thesis Advisor

Sara Hanson, Ph.D.
Associate Professor of Molecular Biology
Secondary Thesis Advisor

Acknowledgements

I would first like to express my gratitude to my lab supervisor and advisor Darrell Killian for giving me the opportunity to work in his lab and guiding me through my research. I learned more than I could have ever hoped and discovered a love for the process of setting out to uncover a biological mystery and figuring out how to do it along the way. Without his support or advice, I never could have completed this thesis or known what I would do after graduating. I would also like to thank Dr. Sara Hanson for her advice and mentorship. She taught me everything I know about genomics and without her this paper would be sorely incomplete.

I want to thank the Colorado College Department of Molecular Biology for their support in my research and academic work. I thank Carrie Moon for her help in my technical endeavors and Kelley Mathers for all her administrative assistance, and for all the coffee. I want to thank the Colorado College SCoRE program for funding my research in the summer of 2021, as otherwise I would never have had this amazing opportunity.

I also want to thank the other members of the Killian lab, Cade Thumann and Ana Musto, who helped make my first time doing serious research a fun and unforgettable experience. I especially want to thank Cade, who shared many of my lab struggles with me and helped me overcome them.

Lastly, I want to thank my parents for making it possible for me to go to college and learn all I have. Their support allowed me to keep working towards my goal during the Covid-19 pandemic and helped me bounce back and go beyond what I thought I possibly could.

ABSTRACT

The *rbm-39* family of genes encode RNA-binding proteins that are predicted to regulate alternative RNA splicing. However, the exact biological and molecular roles of *rbm-39* remain largely unexplored. To learn more, we investigated the role of *rbm-39* in *C. elegans* (a worm used for genetic studies). Loss of *rbm-39* leads to sterility, but the cause has not yet been explored. Using staining and microscopy techniques we found that *rbm-39* knockout worms have reduced germline development and a smaller pool of germline stem cells. This suggests that *rbm-39* is required for normal germline stem cell regulation and the regulation germline development. However, the somatic gonad cells that regulate germ cell development appears morphologically normal. In addition, we also tested the hypothesis that RBM-39 regulates alternative RNA splicing. By isolating and sequencing mRNA from controls and *rbm-39* mutants, we found that *rbm-39* is implicated in the alternative splicing of a large group of genes. Alternative splicing analysis revealed that *rbm*-39 mutants have a significant difference in the usage of alternative 3' splice sites for the RNAs for many genes. Gene ontology (GO) enrichment revealed that some of these RNAs are highly associated with the Notch signaling pathway, a key part of *C. elegans* transcriptional regulation for germline development, some misspliced RNAs are implicated in the cellular response to amino acid stimuli, aminoacylase activity, and the NuRD and CHD-type complexes. *rbm*-39 mutants also had differences in 3' alternative splice site usage for many other RNAs implicated as regulators of gene expression in germ cells, such as *nos-3, pal-1, cye-1,* and *mex-3*. Overall, misregulation of these genes may explain why *rbm-39* mutants are sterile.

BACKGROUND

RNA-Binding Proteins

RNA-Binding Proteins (RBPs) are a part of large group of proteins understood to regulate the expression of other genes. They perform this by influencing mRNA degradation, localization, translation, or via alternative splicing. There are currently thousands of recognized RBPs, and many diseases originate from mutations that impact their functionality. RBPs attach to RNA to form ribonucleoprotein particles (RNPs) which imposes regulation to said RNA. The specific molecular mechanisms of RBPs are not completely understood, and there is even less

understanding of how they affect development. RNPs can change how the RNA interacts with the environment, decreasing or increasing its longevity. RBPs are effective regulators of genes, often forming RNPs but not altering expression until/unless certain cellular conditions are met (pH, presence of amino acids, large quantities of that RNP being present, etc.). RNPs can affect RNA localization, and the rate and location at which it is translated. RNPs can also change the frequency of occurrence for different splice forms of the same gene, among many other things. RBPs bind to RNA via RNA-binding domains (RBDs), which can be predicted computationally given the residue sequence of the RBP (Gebauer et al., 2021).

Generally more highly conserved than transcription factors, RBPs are widely distributed, and phenotypes resulting from mutations are often tissue-specific, with the RBP responsible for the maintenance of that tissue leading to a cascade of aberrant gene expression (Gebauer et al., 2021). This is because RBPs are responsible for much of the tissue differentiation that results from the same genome, so when mutations occur in an RBP, phenotypes are localized to the RBP-governed tissue. Novel proteome-wide RNA interactome capture analysis has shown in a human cell there are 4,257 RBPs that bind to polyadenylated RNA, and oftentimes are not strictly correlated to an RBD. Instead, they can have protein-protein interaction interfaces, act as enzymes, or have DNA-binding domains. These areas have high rates of post-transcriptional modification and are prone to mutations (Hentze et al., 2018). 1,054 of the RBPs were found to have some sort of disease-causing mutation associated with them, either inherited or somatic - roughly one third of the 4,257. Compare this to transcription factor disease mutations, which account for 10% of annotated mutations that impact protein coding. These diseases can be loss of function (LOF), but often alter expression levels of the RBPs, which in turn alter the levels of their affected RNA which can cause cascading signaling effects. For example, the most common inherited intellectual disability, fragile X syndrome (FXS), is caused by a CGG triplet expansion in the 5' UTR of *FMR1*, leading to the hypermethylation and subsequent silencing of the gene. A function FRMP protein, encoded for by *FMR1*, is an RBP that transports and translates neuronal mRNA (Gebauer et al., 2021).

RBPs are understood to have key roles in embryonic and germline development. For example, RBP Igf2bp3 in zebrafish is required for adequate RNA expression for the germline, and a loss-of-function allele for it leads to the misregulation of key stages in the development of the germline (Vong et al., 2021). In *C. elegans*, the RBP MEX-3 is required for repression of gene expression in the germline, and *mex-3* is largely regulated by its 3’ UTR (Albarqi et al., 2021).

RBM39 family of proteins

Originally identified and studied from a fetal pig hind limb muscle cDNA library, RBM39 was then identified as a U2AF^{65} homolog (Stratil et al., 2008; Prigge et al., 2009). RBM39 forms a complex with U2AF^{65} that primarily works as a splicing factor. RBM39 expression was later found to be increased in tumors, though not directly causing their growth. RBM39 was identified as a potential cancer target. The goal of the drug would be for RBM39 to bind to the DCAF15-associated E3 ubiquitin ligase complex via aryl sulphonamides. This leads to the selective degradation of RBM39, which in turn causes increased splicing events, differential expression, and inhibition of cell cycle progression. This has led to cancer regression in preclinical models. Aryl sulphonamides are generally well tolerated as a means of treatment but this tactic fails as a clinical method due to the poor current understand of RBM39 and its molecular mechanisms, as well as a lack of predictive biomarkers (Xu et al., 2020; Xu et al., 2021). RBM39 presents an excellent opportunity for research towards cancer drugs.

RBPs are often associated with diseases affecting the nervous system. In *Drosophilia,* the RBM39 homolog Caper is found to be required for sensory neuron development. The disruption of this gene impacts dendrite morphogenesis in larvae and damages the neuronal positioning of embryonic proprioceptors, as well as impacting the maintenance of mechanosensory bristles in adults (Olesnicky et al., 2017).

The *C. elegans* ortholog RBM39 was identified in a genetic screen for RBPs associated with dendrite morphogenesis (Antonacci et al., 2015). The *rbm-39* gene, then unnamed and referred to as *Y55F3AM.3*, was found during a genetic screen for dendrite defects to the PVD multidendritic sensory neuron. It also colocalizes with spliceosomal proteins and was implicated in some alternative splicing events (Wang et al., 2019). These conclusions were made based on the human homolog RBM39, but despite the importance of *rbm-39* for dendrite development, it is poorly understood.

To learn more about this gene, the Killian lab created a CRISPR/Cas9 knockout of *rbm-39* in *C. elegans*, a null allele called *rbm-39(cnj4)*. This allele was found to cause sterility in worms homozygous for the mutation. Sterility was not expected and was understood to be an issue with oogenesis, as male *rbm-39(cnj4)* homozygotes could still fertilize other worms and sterile *rbm-39(cnj4)* homozygote hermaphrodites could not be fertilized by healthy wild-type males (Judy Cheng senior thesis *Significance of RNA Binding Motif Protein (RBM-39) in Developmental Processes in C. elegans*, 2021).

C. elegans somatic gonad and germline development

The *C. elegans* gonadal arm is a long tube where germline stem cells differentiate, develop, and undergo oogenesis as they progress from the start at the distal tip cell (DTC) to the uterus (Fig. 1). The development of the germline of *C. elegans* is highly dependent on the somatic gonad, tissues that are not a part of the germline but aid in reproduction, at various stages of development. Somatic gonadal organs require strict spatial and temporal regulation for the proliferation and differentiation of reproductive stem cells. There are five tissues to the somatic gonad: DTC, gonadal sheath, spermatheca, uterus, and spermathecal-uterine valve. Germline differentiation first begins in the L3 larval development stage, where meiotic germ cells rely on the proximal gonad to begin development. The transition through the gonad from distal to proximal, managed by the somatic gonad, is crucial to successful fertility (Killian and Hubbard, 2004). The DTC utilizes the GLP-1/Notch signaling pathway to keep local stem cells in an undifferentiated state. LAG-2 and APX-1 are expressed by the DTC and the germline stem cells express GLP-1. This maintains a delicate balance which, if interrupted, could lead to early stem cell differentiation and subsequent sterility (Kimble et al., 2013). In *C. elegans* gonadal sheath cells, the expression of *ceh-18* is required for the maturation of germline stem cells as they go through meiosis. Sheath cells have been found to either directly or indirectly promote and spatially restrict meiotic maturation (Rose et al., 1997; Hubbard and Schedl, 2019).

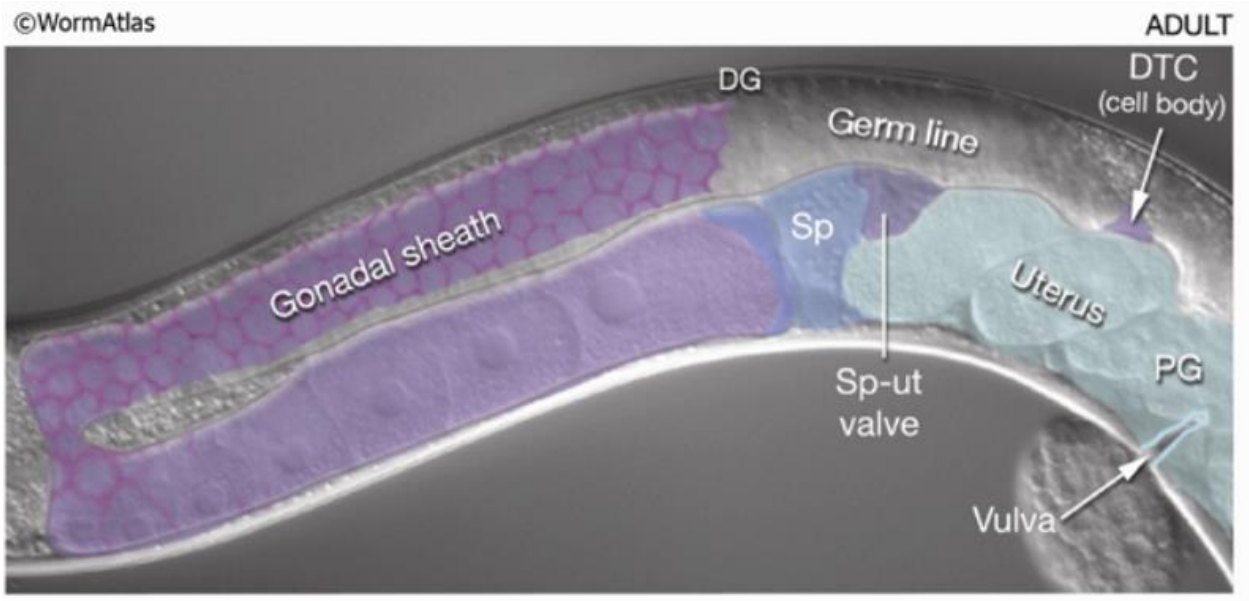


Figure 1: Somatic gonad annotated microscopic image of the *C. elegans* germline
From Worm Atlas 'Reproductive System: The Somatic Gonad', diagram of a gonadal arm of an adult hermaphrodite germline with the somatic gonad labeled.

Because we know there is no issue with sperm production, we can rule out problems with the spermatheca. Preliminary examination of sterile worms indicates a problem with the development and/or differentiation of stem cells and oogenesis, so the problem is likely not with the uterus, where developed eggs are stored prior to fertilization, nor with the spermathecal-uterine valve, where fertilization takes place. Because of this, our study of the somatic gonad was focused on the DTC and gonadal sheath cells, which many past studies have shown to be important for the regulation of germline stem cells and meiotic progression/oogenesis (review ref here).

Alternative Splicing

RBM39 has been implicated in alternative splicing (AS), often as a complex formed with spliceosomal protein U2AF(65) (Stratil et al., 2008). Alternative splicing is the source of much of the diverse genetic functionality seen in eukaryotes, where the number of different spliceforms increases the number of proteins encoded relative to the total number of genes in the genome. Pre-mRNA can be spliced in many different ways before it reaches its final state as mRNA, each variation is considered a different spliceform. These AS types include exon skipping (SE), alternative 3' and 5' splice cites, (A3SS and A5SS), mutually exclusive exons (MXE), and retained introns (RI) among others (Fig. 2) (Zahler, 2005).

RBPs can act as splicing factors to silence or enhance these splice forms, change their longevity via shortening poly-A tails, or adapt their function, sometimes even promoting mRNA decay via poison exons (Wang et al., 2015). In *C. elegans*, The 3' UTR is one of the primary locations at which mRNA is regulated, via stability or translation. These regions are often more associated with regulating gene expression than promoters in the germline, with the exception of spermatogenesis (Merritt et al., 2008).

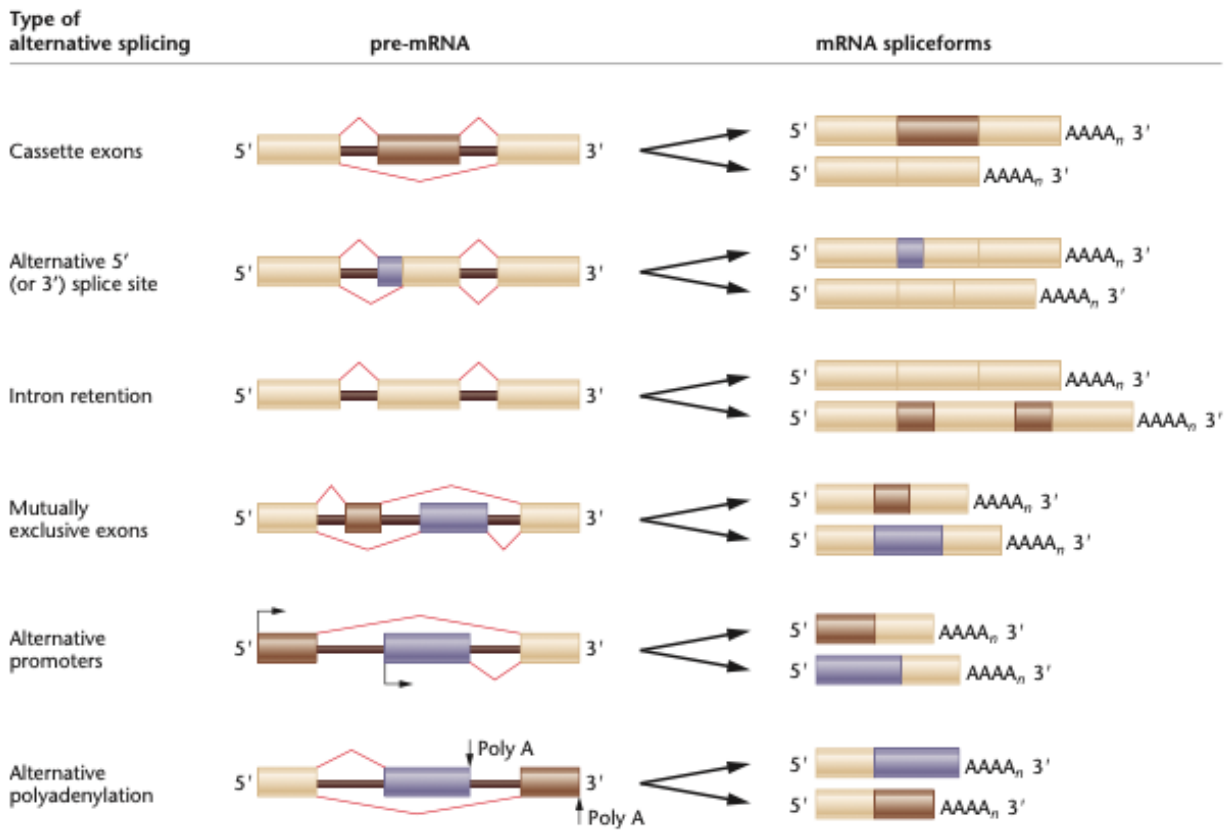


Figure 2: Types of alternative splicing

Boxes represent exons and the thin black lines represent introns. Red lines represent where exons will be connected following alternative splicing (from *Concepts of Genetics, 12th ed*)

Experimental Aims

The goal of this study was to characterize *rbm-39* functionality and expression. This included identifying whether the somatic or germline gonad is failing in *rbm-39(cnj4)* mutants and which part. This also included determining how *rbm-39* acts as an RNA-binding protein in terms of what genes it regulates and how, what biological functions are associated with these genes, and what sort of patterns does the RBM39 protein acts with.

MATERIALS AND METHODS

C. elegans strains

Strains were derived from the Bristol strain N2, grown at 20C, and constructed using standard procedures (Brenner 1974).

An overview of the *C. elegans* crossing strategy is outlined in figure 3. *tmC25 [unc-5; myo-2::Venus]* worms were crossed with *him-5* males to create males carrying the *tmC25* balancer chromosome. This was confirmed as *tmc25* heterozygotes as they had the *tmC25* GFP-tagged pharynx marker, and were not uncoordinated. These males were crossed seperately with the DTC-tagged strain JK2049 (*qIs19 [lag-2[::GFP::unc-54 3'UTR + rol-6(su1006]*) (Hall et al., 1999) and the sheath cell-tagged DG1575 (*tnIs6 [lim-7::GFP + rol-6(su1006]*) (Blelloch et al., 1999). Both these transgenes carry the dominant roller marker *rol-6(su1006)*, with high expression of the roller phenotype in homozygotes. The next generation was verified by the presence of gonad and pharynx markers, and were then self-crossed to obtain homozygotes, which were verified by allowing them to self-fertilize and checking the next generation for any missing markers. *rbm-39(cnj4)* worms, which are balanced with the *tmC25* balancer chromosome, were crossed with *him-5* males. Worms were picked based on the presence of the pharynx marker that moved well, as uncoordinated worms would be homozygous for *tmC25* and have lost the *rbm39(cnj4)* mutation. Male individuals in the next generation are either *rbm-39(cnj4)/+* or *tmC25/+*, so *rbm-39(cnj4)* males were selected by the lack of a pharynx marker. These males were crossed with *tmC25/tmC25 ; GFP/GFP* worms. The next generation can be either *tmC25/+ ; GFP/+* or *tmC25/rbm-39(cnj4) ; GFP/+*. The *rbm-39* mutants were identified by isolating individuals in the next generation that lack the pharynx marker and seeing if they can reproduce. *tmC25/rbm-39(cnj4) ; GFP/GFP* Individuals from confirmed populations were then selected based on the presence of a pharynx marker, a lack of uncoordination, bright gonadal markers, and a rolling phenotype. The offspring of these individuals were checked to make sure they were not all uncoordinated and there were some *rbm-39(cnj4)* homozygotes (lacking the pharynx marker, which is now the method to identify sterile tagged *rbm-39(cnj4)* mutants), and that all the offspring were rollers with the respective gonadal GFP marker. These DTC (*qIs19*) and sheath cell-tagged (*tnIs6*) *rbm-39(cnj4)* worm strains were named DJK342 and DJK343, respectively. Some worms were frozen at -70°C and the rest were kept at 20°C until needed for staining and microscopy.

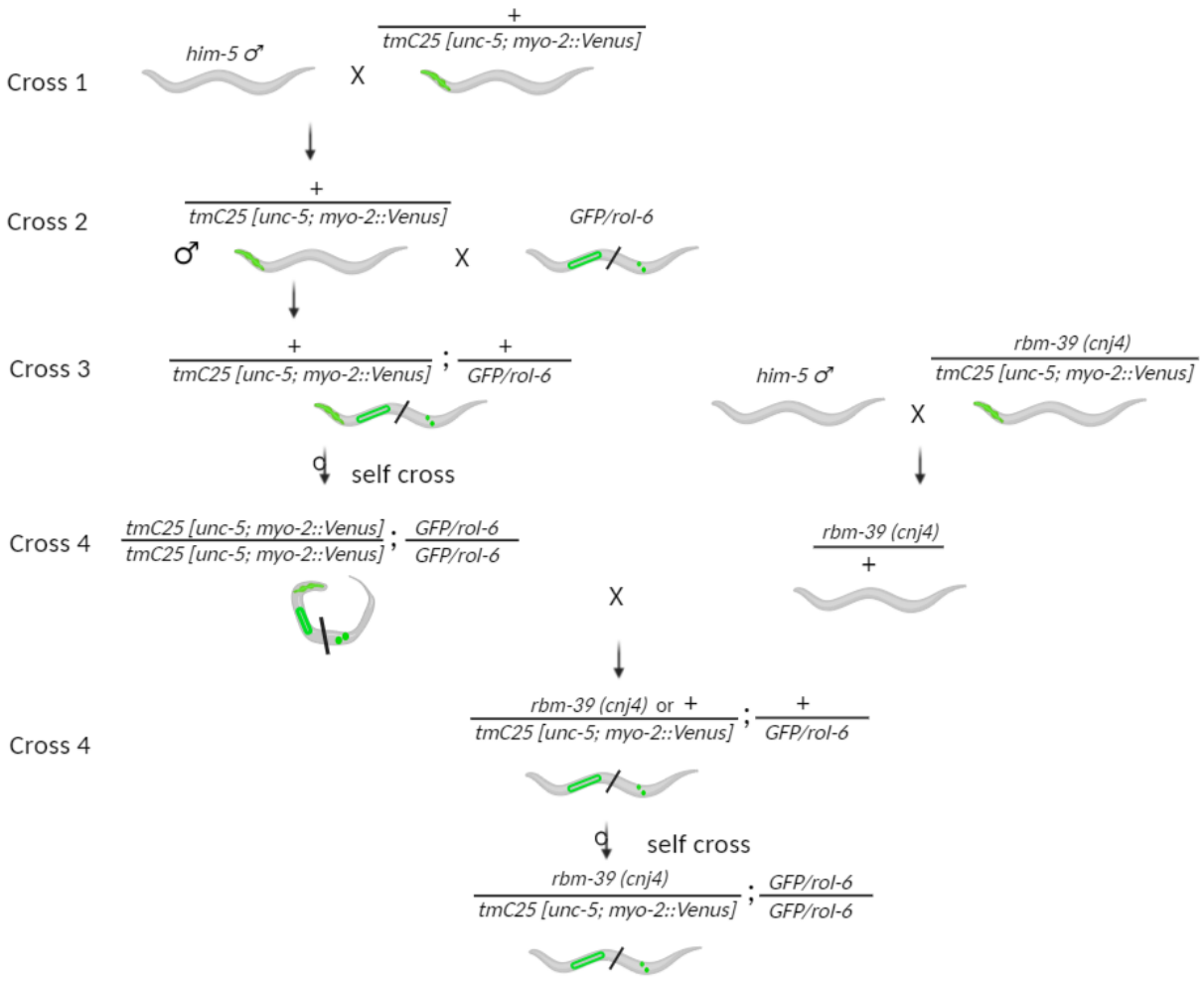


Figure 3: Outline of the *rbm-39(cnj4)* GFP cross

Staining and Microscopy

L4 stage control and experimental *rbm-39(cnj4)* worms, each carrying the relevant GFP transgenes, were transferred directly to a depression slide loaded with 50 µL 0.25mM levamisole. Once worms were fully paralyzed, they were dissected by removed their heads at the pharynx with syringes. 150 µL 3% paraformaldehyde in 0.1 M K_2HPO_4 was added to the well and the worms were fixed for 30 minutes. Gonad arms were transferred directly to a plate with a layer of 2% agarose as a cushion between them and the slide cover. Imaging was done with a Zeiss Axioskop, Zeiss AxioCam, and Zen software at 100X for measuring cytoneme length and distal mitotic zone with the DJK342 strains and at 40X for observing general gonad health and gonadal sheath cells with the DJK343.

RNA Extraction

The *rbm-39(cnj4)* mutants are balanced with *tmC25 [unc-5; myo-2::Venus]*. This allowed homozygous *tmC25* mutants to be identified by the *unc-5* phenotype and heterozygotes by their GFP fluorescence in the pharynx. Homozygous *rbm-39(cnj4)* mutants were identified by their lack of a fluorescent pharynx.

Roughly 100 *rbm-39(cnj4)* homozygotes were picked and added to a 1.5 mL microcentrifuge tube in DI water, as well as an equal quantity of control worms in a separate tube. Worms were spun down at 4k rpm for 5 minutes and supernatant was removed. Water was added again, and this step was repeated. 500 µL Trizol was added to each tube, and then tubes were vortexed and frozen at 70°C. Tubes were thawed, vortexed, and frozen again. This freezing/thawing step was repeated 4 times and then tubes were allowed to incubate at room temperature for 10 minutes. Tubes were spun at 14K for 10 minutes at 4°C. The supernatant was moved to a fresh tube to which 100 µL chloroform was added. Tubes were vortexed and let sit for 3 minutes before being spun at 12K for 15 minutes at 4°C. At least 250 µL of the upper aqueous phase was moved to a clean tube along with 250 µL isopropanol and 1 µL GlycoBlue (ThermoFisher) was added. The solution was mixed by inversion and let sit for 10 minutes at room temperature. Tubes were spun at 12K for 10 minutes at 4°C. The supernatant was poured off, removing the final portion via micropipette. 1 mL 75% ethanol was added to the tube and carefully mixed by pipetting up and down. Tubes were then spun down at 7.5K for 5 minutes at 4°C before removing the supernatant. This ethanol washing step was repeated and pellets were allowed to air-dry for roughly 30 minutes. 30 µL RNase-free water was added to each tube, which were put on a 60°C heat block for 10 minutes to put the RNA into solution. The concentration of RNA was determined via nanodrop and RNA samples were frozen at -70°C until sequencing.

RNA-Sequencing and Data Analysis

Sequencing of the mRNA libraries and preliminary genomic analysis was performed by Novogene. This included data quality control, quantification of gene expression and differential expression, enrichment analysis, and alternative splicing analysis, among others. This allowed for verification that RNA extraction was successful. RNA-seq coverage is lacking any presence of the *rbm-39* gene for all four mutant samples (Fig. 4). All RNA-sequencing samples had sufficient clean reads and sufficiently low error rates (Table 1). Additional GO analysis was

performed using the Gene Ontology GO Enrichment Analysis resource utilizing the Panther Classification System (version 16.0).

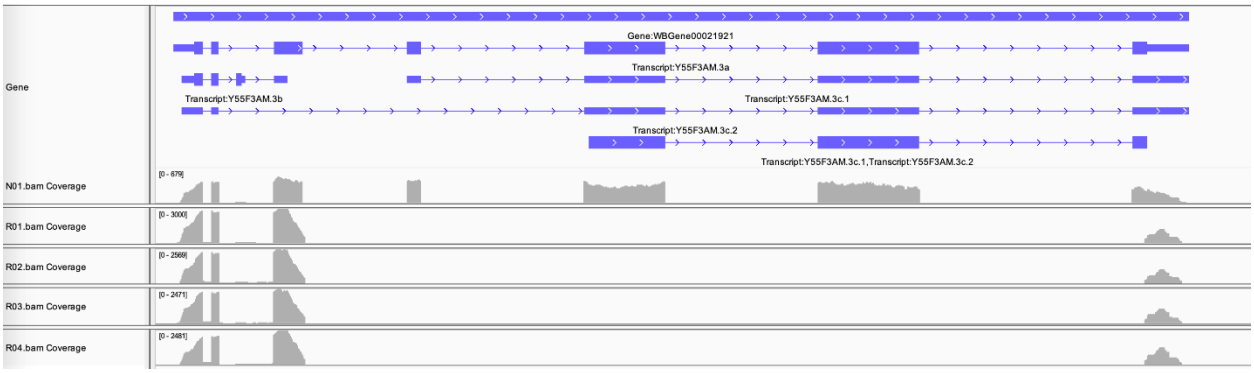


Figure 4: *rbm-39* coverage map shows successful RNA extraction

RNA-seq coverage for one N2 replicate and all four mutants. At the top the *rbm-39* gene is shown. There is no mRNA found in the *rbm-39(cnj4)* mutants for the *rbm-39* gene outside of the beginning, which was not deleted.

Sample name	Raw reads	Clean reads	Raw bases	Clean bases	Error rate(%)	Q20(%)	Q30(%)	GC content(%)
R01	21660058	21326014	6.5G	6.4G	0.03	97.69	93.16	47.46
R02	22942764	22695067	6.9G	6.8G	0.03	97.59	92.98	46.62
R03	21787587	21482317	6.5G	6.4G	0.03	97.75	93.40	47.82
R04	21812045	21357013	6.5G	6.4G	0.03	97.70	93.28	47.08
N01	22369007	22137030	6.7G	6.6G	0.03	97.46	92.55	46.81
N02	22762605	22483494	6.8G	6.7G	0.03	97.43	92.53	47.66
N03	22235610	22045457	6.7G	6.6G	0.03	97.49	92.58	46.86
N04	23185702	23005629	7.0G	6.9G	0.03	97.46	92.57	47.07

Table 1: Data quality summary

Data for the *rbm-39(cnj4)* and N2 mRNA libraries. Shows raw reads and base-counts, clean reads and base-counts post-filtering from raw data (used for all statistics), error rate of the raw reads, percentage of bases with a Q Phred value >20% and >30%, and the percentage GC content.

RESULTS

Gonadal Arm Morphology

The somatic gonad is important for germline development (Pepper et al., 2003; McCarter et al., 1997), thus we analyzed the somatic gonad of *rbm-39* mutants to see if any morphological defects of the somatic gonad might help explain why the worms are sterile. *rbm-39(cnj4)* worms crossed into strains with either their DTC (*qIs19*) or gonadal sheath cells (*tnIs6*) 1-4 tagged with GFP were dissected. Microscopy was performed on the gonadal arms to measure distal mitotic zone and cytoneme length, as well as check the general morphology of the tagged somatic germline (see materials and methods). DTC cytonemes are thin cytoplasmic arms that extend from the DTC and helps keep distal reproductive germ cells in an undifferentiated state by inhibiting meiosis and promoting mitosis (Worm Atlas 'Reproductive System: The Somatic Gonad'). A morphological change to these cytonemes could lead to earlier stem cell determinism and subsequent insufficient oogenesis. There was found to be no significant difference in the length of the DTC cytonemes when comparing controls (Fig 5A) to *rbm-39* mutants (Fig 5B). Despite the lack of change in DTC morphology, the distal mitotic zone was found to be shorter in the *rbm-39* mutants (Fig. 5). Microscopy of the sheath cell-tagged strain revealed that all four pairs of gonadal sheath cells that are tagged in this strain are present (Fig. 6). Gonads generally appear less developed and were more difficult to dissect due to their reduced internal pressure. Stem cells beginning mitosis earlier could contribute to sterility in *C. elegans* and may be caused by a misregulation in the cell differentiation pathway. Following the "seesaw" model of stem cell differentiation outlined in Shu et al., 2015, earlier differentiation could be caused by an imbalance in the complex network of signaling molecules that keep stem cells in their undifferentiated state.

Something other than morphological changes to the DTC or sheath cells was leading to sterility in *rbm-39(cnj4)* worms and a reduced germline stem cell population. This could be due to defects in the signaling pathways from the somatic gonad to the germ cells, or due to cell-autonomous defects within the germ cells.

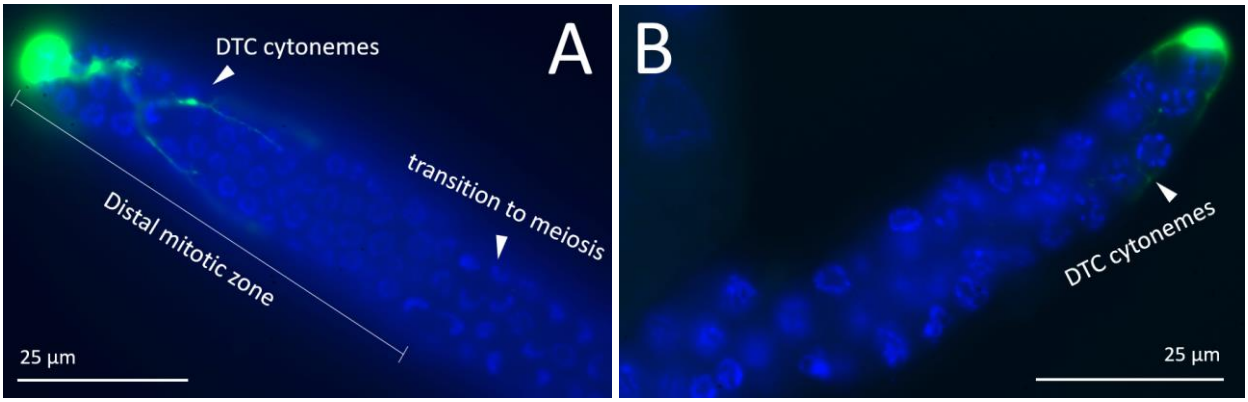


Figure 5: *rbm-39* knockout has no significant impact on distal tip cell cytoneme length but leads to earlier mitosis in developing oocytes
Distal ends of the gonadal arms of worms carrying the *qIs19* DTC-GFP marker (A) and *qIs19*; *rbm-39* mutants (B).

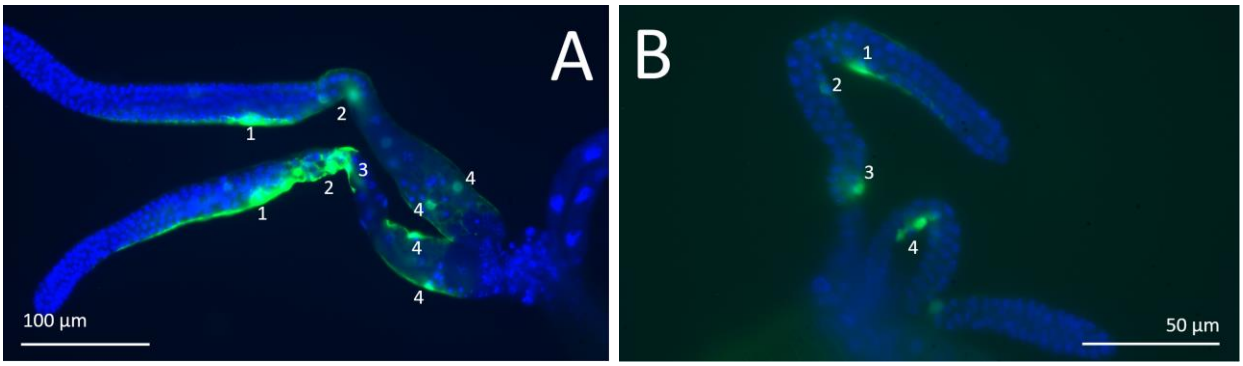


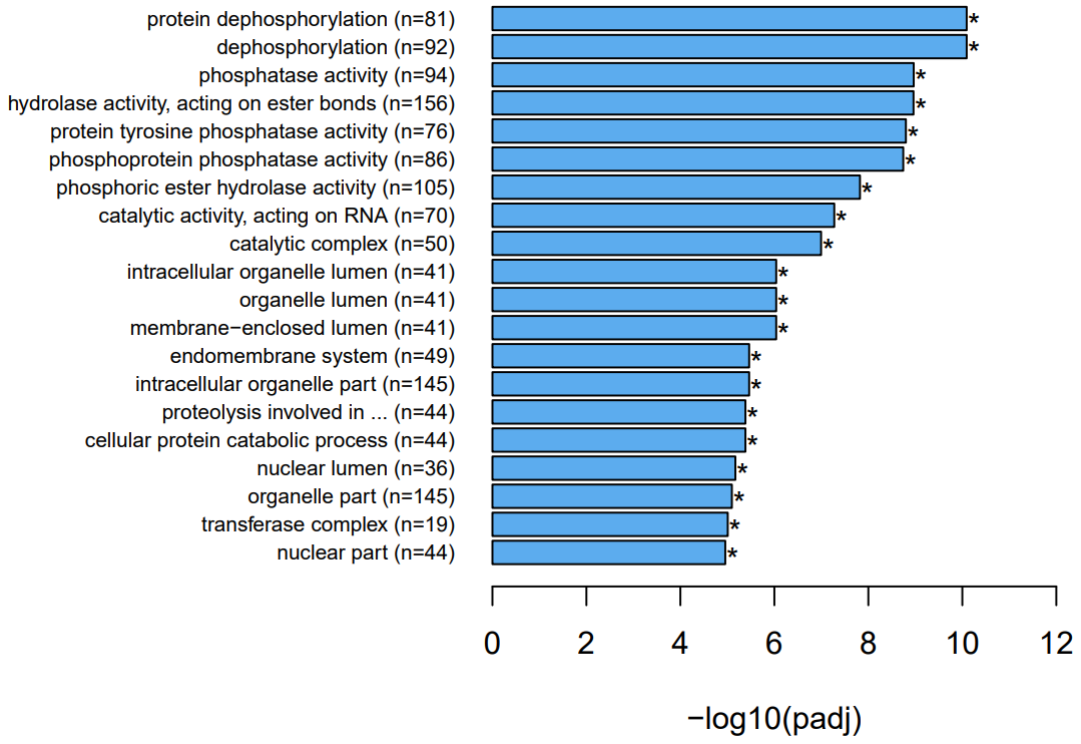
Figure 6: Despite appearing underdeveloped, the *rbm-39* knockout causes no major change to gonadal sheath cell pairs 1-5
Gonads in worms carrying the *tnIs6* gonadal sheath-GFP marker with sheath cells numbered in GFP-tagged controls (A) and the *rbm-39* knockout mutant (B). Despite appearing underdeveloped, all four of the gfp-tagged pairs of sheath cells are present in *rbm-39* mutants. This and the DTC cytoneme data indicate that sterility in *rbm-39* mutants is not caused by morphological changes to either of these somatic gonad tissues.

GO Analysis of Regulation Pathways

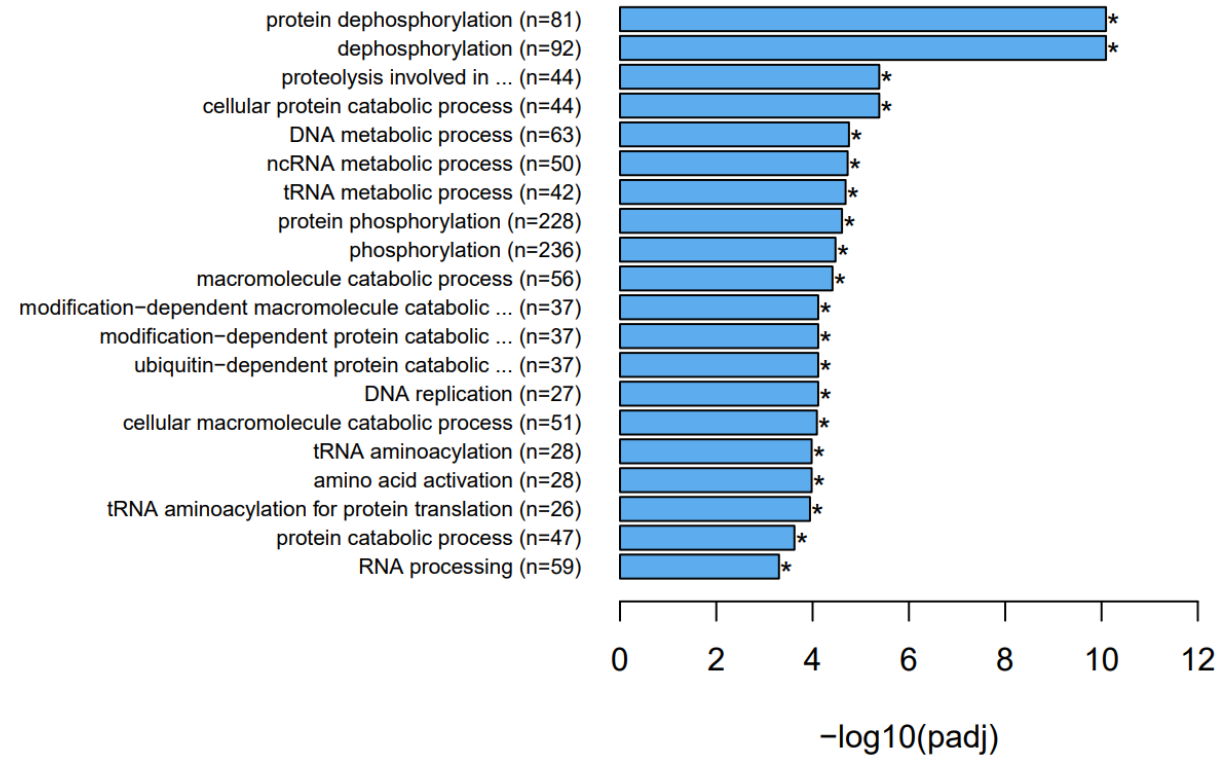
Because RBM39 is an RBP and thus likely participates in the regulation of gene expression, we used RNA-sequencing to determine if there are changes in gene expression that might help explain the germline development defects we observe and that would help shed light on the molecular role of RBM-39. To categorize the transcriptional data and determine what biological, cellular, and molecular functions were transcriptionally altered, gene ontology (GO) analysis was performed. This compared relative levels of statistically significant differences in the

translation of genes between the control and experimental *rbm-39(cnj4)* worms. There were 8600 total significant differentially expressed genes, 3948 of them upregulated and 4652 downregulated in mutants. There was an average fold change in differential expression of 1.8513. The largest was *col-165* with a fold change of -11.4214, with no expression in any of the *rbm-39(cnj4)* mutants. GO enrichment on the RNA-seq differential expression dataset revealed significant changes to many categories. Primarily acting in dephosphorylation, phosphatase, and hydrolase activity, *rbm-39* is well suited to act on RNA. Serine-/arginine-rich RBPs are densely phosphorylated, and the dephosphorylation of splicing factors can lead to increased interactions (Kundinger et al., 2021). Additionally, ribonucleoproteins have been shown to undergo cycles of phosphorylation/dephosphorylation to modify binding activity with pre-mRNA (Fung et al., 1997).

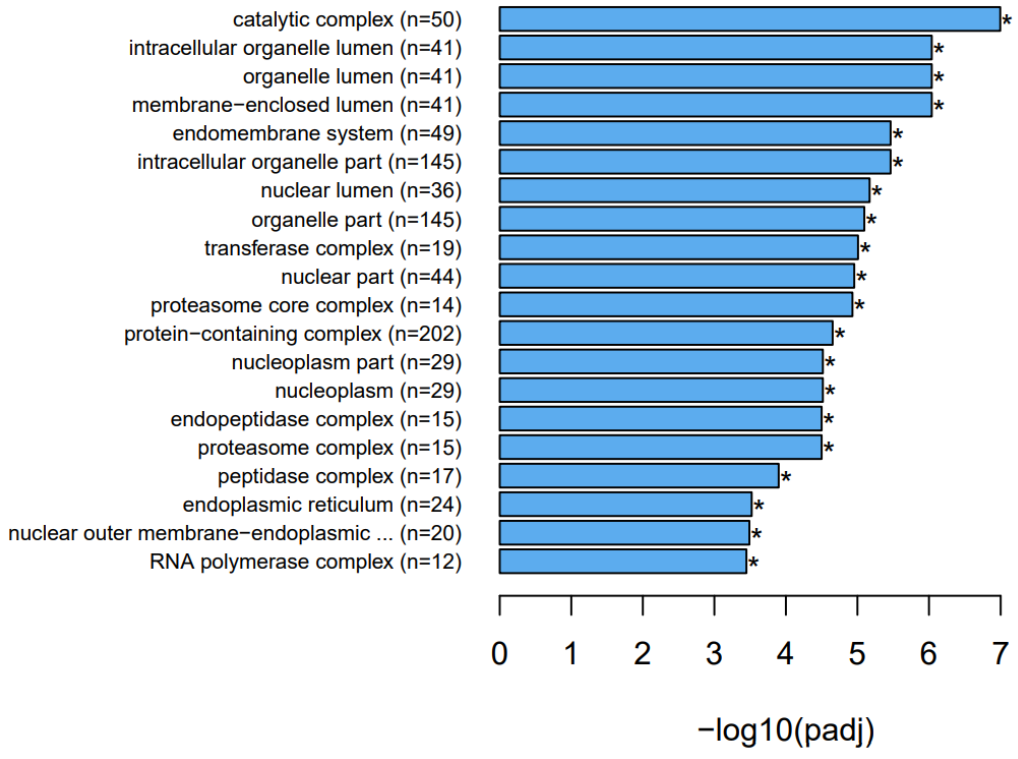
A.



B.


protein dephosphorylation (n=81)
dephosphorylation (n=92)
proteolysis involved in ... (n=44)
cellular protein catabolic process (n=44)
DNA metabolic process (n=63)
ncRNA metabolic process (n=50)
tRNA metabolic process (n=42)
protein phosphorylation (n=228)
phosphorylation (n=236)
macromolecule catabolic process (n=56)
modification−dependent macromolecule catabolic ... (n=37)
modification−dependent protein catabolic ... (n=37)
ubiquitin−dependent protein catabolic ... (n=37)
DNA replication (n=27)
cellular macromolecule catabolic process (n=51)
tRNA aminoacylation (n=28)
amino acid activation (n=28)
tRNA aminoacylation for protein translation (n=26)
protein catabolic process (n=47)
RNA processing (n=59)
0
2
4
6
8
10
12
−log10(padj)

C.


catalytic complex (n=50)
intracellular organelle lumen (n=41)
organelle lumen (n=41)
membrane−enclosed lumen (n=41)
endomembrane system (n=49)
intracellular organelle part (n=145)
nuclear lumen (n=36)
organelle part (n=145)
transferase complex (n=19)
nuclear part (n=44)
proteasome core complex (n=14)
protein−containing complex (n=202)
nucleoplasm part (n=29)
nucleoplasm (n=29)
endopeptidase complex (n=15)
proteasome complex (n=15)
peptidase complex (n=17)
endoplasmic reticulum (n=24)
nuclear outer membrane−endoplasmic ... (n=20)
RNA polymerase complex (n=12)
0
1
2
3
4
5
6
7
−log10(padj)

D.

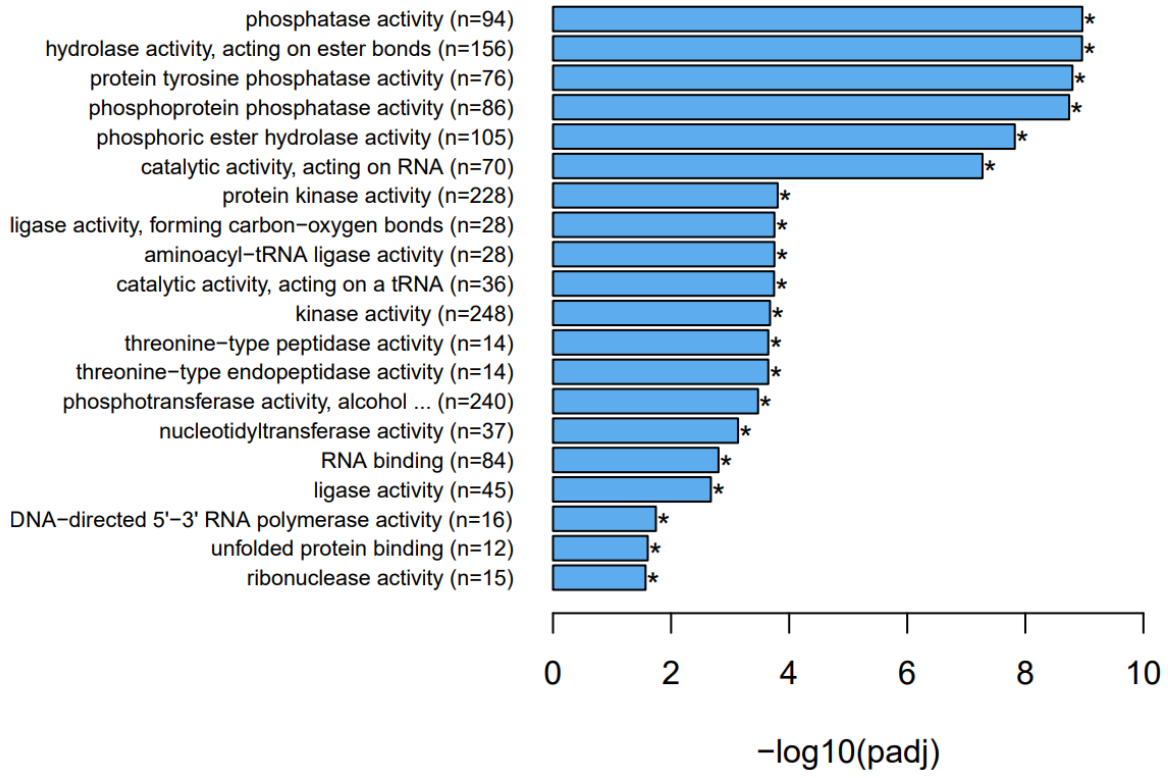


Figure 7: *rbm-39* is highly involved in dephosphorylation and phosphatase activity
Gene ontology enrichment analysis comparing the *rbm-39* mutant and N2 RNA-seq libraries. Shows GO enrichment for all categories (a), biological processes (b), cellular components (c), and molecular functions (d).

RBM-39 Directs 3' Alternative Splicing

In addition to differential expression analysis and broad GO analysis on that expression, there was also analysis of alternative splicing. This is generally performed by observing what splice forms for different genes appear in statistically different quantities. There were 1,591 total significant AS events split into 5 categories (Table 2) (See background for types of alternative splicing). We repeated the GO enrichment analysis on the alternative splicing datasets, which revealed several findings (Fig. 8). For some AS types there was no significant enrichment, nor was there any for the combined datasets of all alternative splicing genes. For other categories there was enrichment, but with lower fold-changes than when analyzing the entire differential expression dataset. For the 3' alternative splicing however, which involves splice site selection at the 3' end of the intron, there were massive fold changes. These ontological categories averaged around a fold change of 5, with some reaching above 30. 3' alternative splicing is

highly involved in the Notch signaling pathway biological process category, which is a key pathway for the regulation of the germ cell mitosis/meiosis decision, and other developmental events. This alternative splicing increased the number of genes regulating phosphatase activity and both protein and non-protein dephosphorylation. This is consistent with GO analysis performed on the original dataset, which returned these three categories as the three most prevalent terms. The massive quantity of dephosphorylating proteins *rbm-39* manages could explain its ability as an effective 3' alternative splicing regulator of germline development and many other areas. There are many 3' alternative splice sites *rbm-39* impacts in the area of germline development, such as the positive regulation of post-embryonic and larval development, negative regulation of vulval development and the developmental process, positive regulation of cell growth and the negative regulation of cell differentiation. It is likely the misregulation of these genes, among others causes *rbm-39* mutant sterility.

AS event type	Events	Sig. events
Exon skipping	7151	1027 (121:906)
Mutually exclusive exons	798	166 (76:90)
Alternative 5' splice sites	274	54 (35:19)
Alternative 3' splice sites	488	286 (46:240)
Retained introns	181	56 (42:14)

Table 2: Differential alternative splicing analysis from replicate RNA-Seq data reveals a comparatively high number of alternative 3' splice sites
The total number of splicing junction (JC) events and statistically significantly events (up:down).

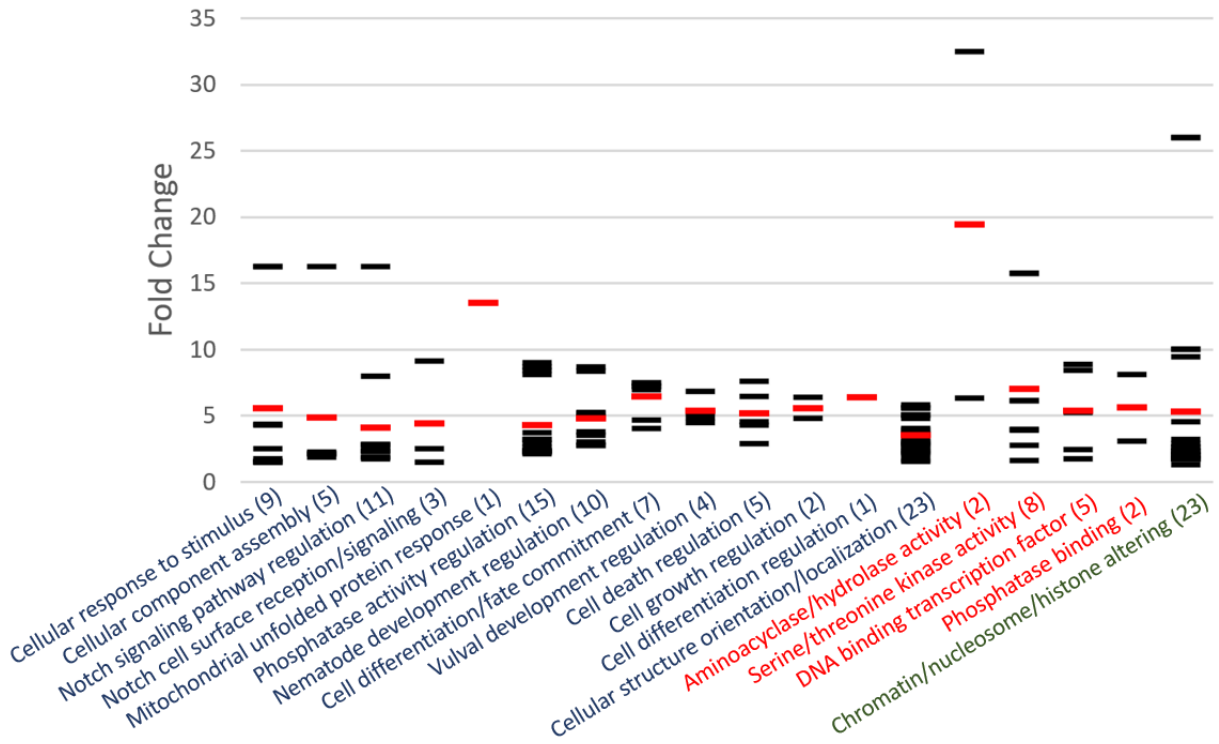


Figure 8: GO analysis reveals ontological similarities for genes with *rbm-39*-affected 3' alternative splicing

Gene ontology of the 3' alternative splicing data. Ontological traits are represented by black bars and are grouped into categories and graphed by fold change increase. The number of traits represented by a category is shown in parenthesis. The red bar indicates the mean fold change of the category. Categories are further divided into biological processes (blue), molecular functions (red), and cellular components (green).

3' Alternative splicing has a pattern regardless of length

3' alternative splice forms change the location of the intron/exon junction and thus can change which sequences are included, or not, in the downstream coding exon. It does so by changing the splice site in transcribed RNA. When comparing controls to *rbm-39(cnj4)* mutants, we found that 3' AS splice sites revealed a consistent pattern in the mutants: a reduction in the expression of both the standard and 3' alternative splice forms, especially reduced in the longer 3' AS form. The spliced section was most often 6 base pairs, with 9 as the second most common and 12 as the third (Fig. 9).

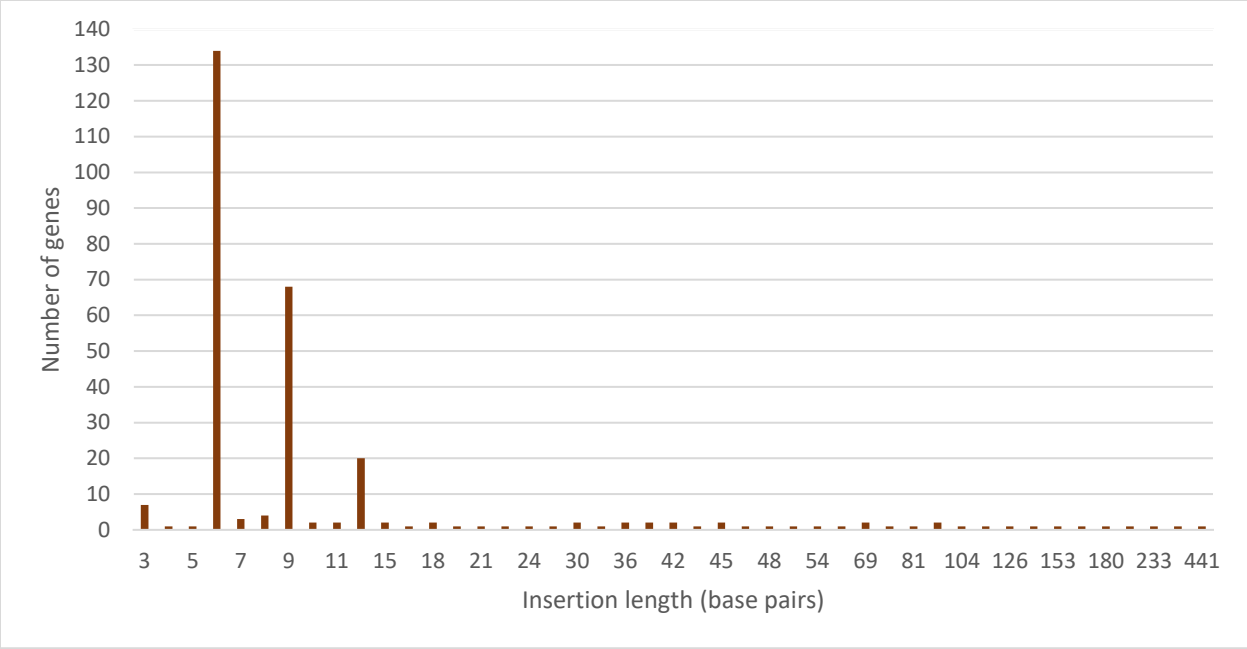
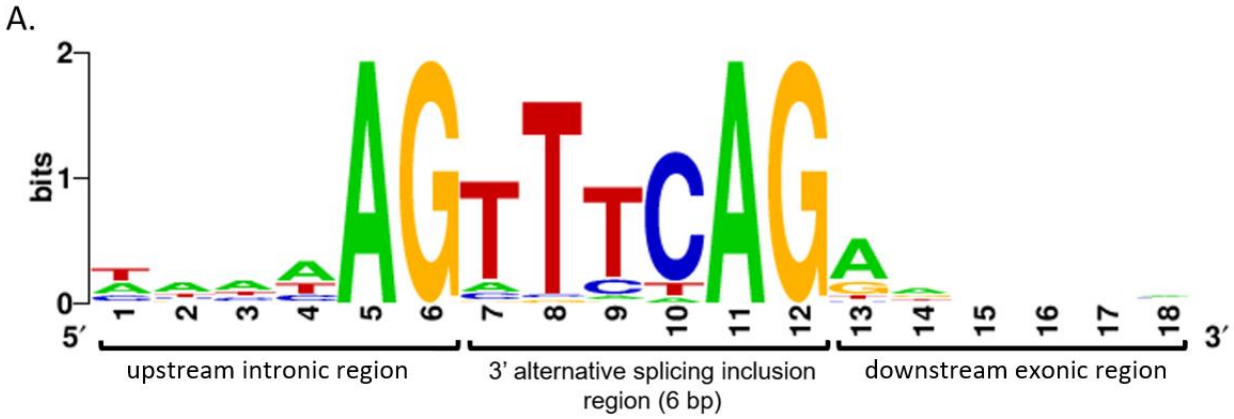


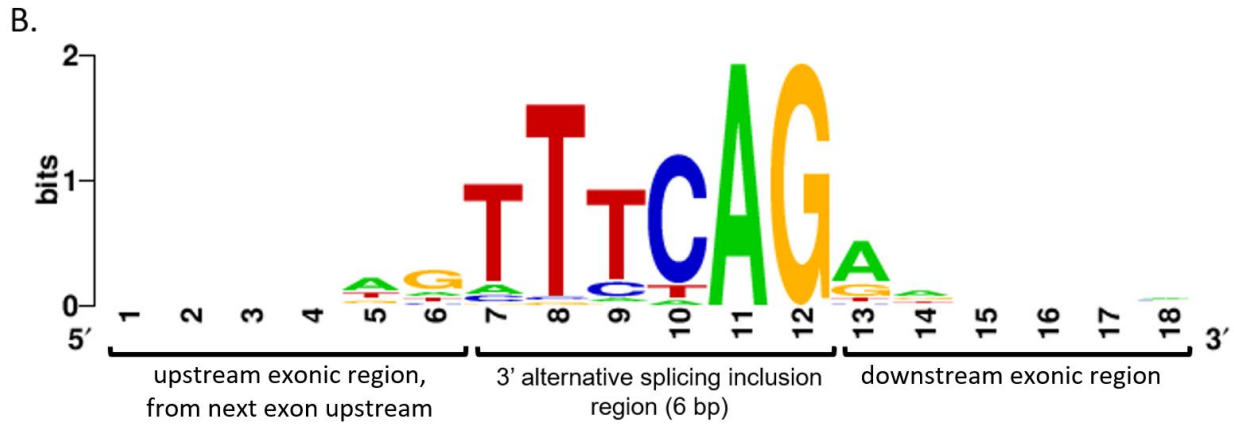
Figure 9: Exonic inclusions for 3' alternative splicing are most often six or nine base pairs, and usually do not cause frameshift mutations

Rates of occurrence for each 3' alternative splice site by the number of base pairs upstream of the canonical site. 8.4% of these alternative splice sites lead to frameshift mutations.

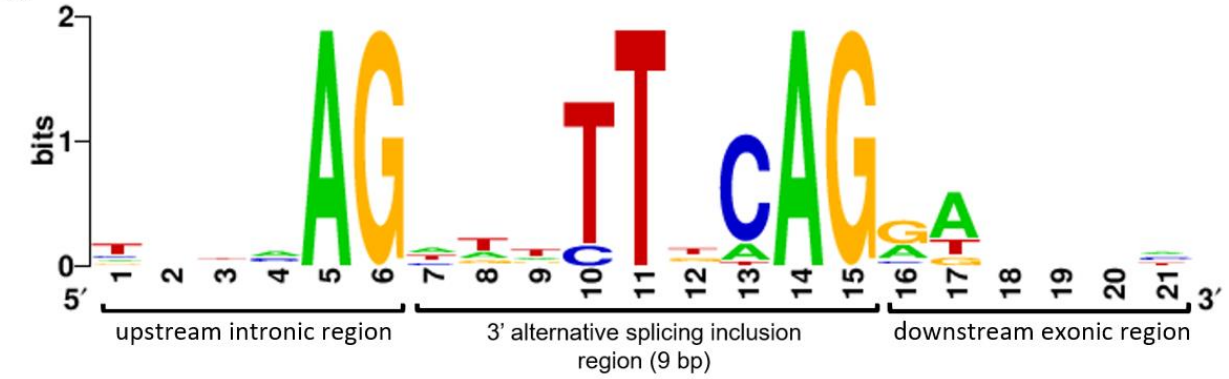
Due to the strong ontological similarity for 3' AS gene set, as well as the consistent length of the altered splice forms, I set out to determine if there is a pattern to the base pairs of or around the 3' insertions to determine if there could be a unifying molecular mechanism. The 3' AS genes were organized by the length of the additional exonic sequence, and a position weight matrix (PWM) was performed on each set using the Berkeley weblogo tool (https://weblogo.berkeley.edu/logo.cgi). This revealed highly significant patterns to the 3' end of the intronic sequence before the exon, as well as in 5' inclusion site for the exon (Fig. 10). Similarly to comparing the gene ontology enrichment of 3' splicing to the other splice forms, only the 3' AS set of has a consistent pattern to the length or nature of the included spliced segment. In all cases, the 3' end of the intron that appears right before the spliced exon ends with an AG, and the splicing region of the exon is ended with the presence of another AG. This AG is usually preceded by a cytosine, and two thymines are usually two nucleotides downstream. These findings are particularly interesting when considering the canonical AG 3' intron sequence (Burset et al., 2000). It appears that the presence of the TTTCAG sequence causes the post-transcriptional splicing mechanisms to skip over the previous AG 3' intron sequence

and instead use the following one downstream, and it is the *rbm-39* gene that is directly or indirectly responsible for upregulating this mechanism. This pattern is also present in the case of frameshift mutations, and given their lower rate of occurrence, they are likely an error in the application of the *rbm-39* gene or mutations made to the splicing after.

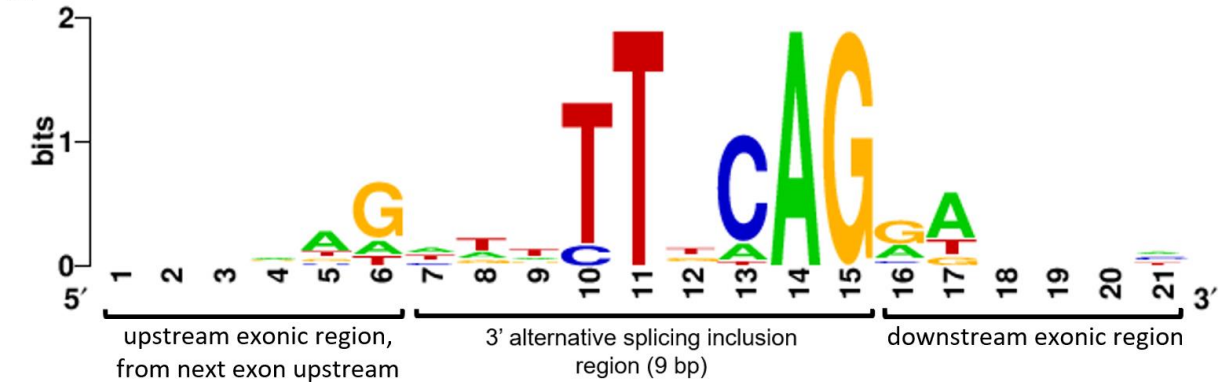




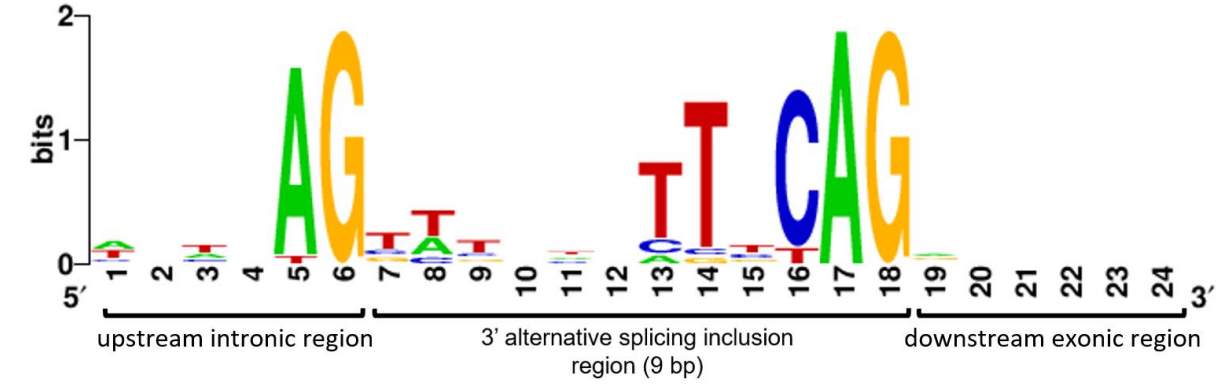
C.


bits
2
1
0
5′
3′
upstream intronic region
3’ alternative splicing inclusion region (9 bp)
downstream exonic region

D.


bits
2
1
0
5′
3′
upstream exonic region, from next exon upstream
3’ alternative splicing inclusion region (9 bp)
downstream exonic region

E.


bits
2
1
0
5′
3′
upstream intronic region
3’ alternative splicing inclusion region (9 bp)
downstream exonic region

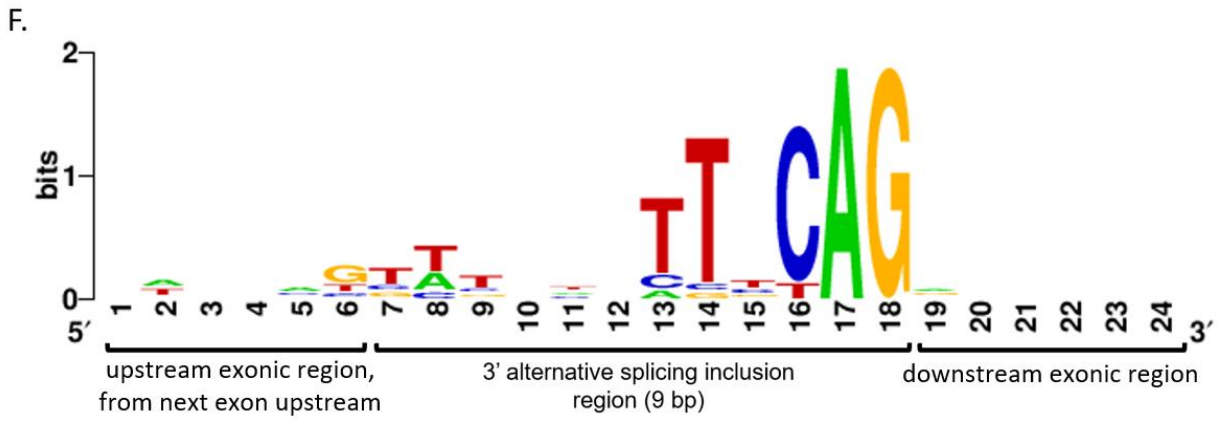


Figure 10: When performing 3' alternative splicing, *rbm-39* acts upon an intronic AG to recognize the 5' end of the splicing region, and ends the splice when encounters a CAG.
PWM for the 3' end of the alternatively spliced exon for all genes with the same inclusions size as well as six base pairs downstream, in the neighboring intron and 6 base pairs upstream of the exon past the alternative splicing region. This was performed on the three largest sets of sizes, 6 base pairs (A), 9 base pairs (C), and 12 base pairs (E). A another set of PWMs were made for the post-splicing mRNA with the intronic regions cut out on the same sets, 6 (B), 9 (D), and 12 (F) base pairs.

This analysis reveals more about RBM-39's function. In *rbm-39(cnj4)* mutants, expression of both splice forms were highly reduced, particularly the longer splice form that included more of the 5' end of the exon. This tells us that *rbm-39* is directly or indirectly responsible for the necessary levels of transcription for both alternative splice forms, and this also suggests that specific levels of these alternative spliceforms are necessary for various biological processes, including process that underlie fertility.

DISCUSSION

The role of *rbm-39* on the *C. elegans* germline and somatic gonad

Microscopy performed on *rbm-39* mutants revealed no obvious defects to gonadal sheath cells 1-4, nor to the distal tip cell. Therefore, sterility and germline development defects are not caused by a gross anatomical problem of the distal portion of the somatic gonad. However, it is still possible sterility is caused by morphological and functional changes to the uterus, or to the spermathecal-uterine valve though this is unlikely because these somatic gonadal tissues are not documented to play a role in stem cell regulation and early oogenesis. It is possible that

even if the somatic gonad cells appear morphologically normal that they are defective in signaling to the germ cells. It is also possible that the germline defects are due to misregulation of processes within the germ cells themselves. The RNA-sequencing analyses strongly suggest that there are many genes that are misregulated and misspliced in *rbm-39* mutants and these data provide many candidate genes that are implicated in germline development.

The Role of *rbm-39* in Neuron Development

Mutations to RBPs often show tissue-specific phenotypes. *rbm-39* was identified as a gene of interest due to its role in *C. elegans* neuron development but was later found to be important for germline development as well, which is the focus of this thesis. Analysis of neuronal roles for *rbm-39* is the focus of another senior thesis (Cade Thumann, 2022).

3' Alternative Splicing

The discoveries of this research hint at the mechanisms of *rbm-39*, but there are fewer findings with regards to why *rbm-39(cnj4)* homozygotes are sterile. The alternative splicing analysis revealed that *rbm-39* upregulates four genes that are 3' UTR-based primary regulators of germline development as RNA or DNA-binding transcription factors. These genes are germline progenitor *nos-3,* oocyte gene *pal-1,* and mixed genes *cye-1* and *mex-3* (Merritt et al., 2008). *nos-3* is predicted to enable RNA binding, and is involved in upregulating meiotic nuclear division. *pal-1* enables DNA-binding TF activity and RNA polymerase II transcriptional regulation. It is involved in embryonic pattern specification and cell fate commitment. *mex-3* enables single-stranded RNA binding activity in the DTC and embryos, and is involved in various processes such as gene regulation. *cye-1* is involved in various processes of reproduction and development (gene information from wormbase.org). Due to these gene's involvement with transcriptional regulation and reproduction, and their roles as 3' UTR germline development regulators, they are key genes for future research into why *rbm-39(cnj4)* incurs sterility. The link between 3' UTRs and 3' alternative splicing involving germline development is currently unknown, but future research into these genes and *rbm-39* could reveal the mechanisms for sterility. Another key gene for future research is *lag-1*, which has a 3' AS splice form upregulated by *rbm-39*, and is part of the Notch pathway by enabling Notch binding. This gene presets the best opportunity to learn the role *rbm-39* plays in Notch activity, a key source

of signalling in the *C. elegans* embryo. Any of these genes could be knocked out to determine if their misregulation leads to sterility, to explain where sterility in *rbm-39(cnj4)* worms originates.

Future Directions with the *Drosophilia* Caper gene

There is currently no full knockout of the fruit fly ortholog to *rbm-39*, Caper. There is, however, a knockdown, and preliminary RNA-sequencing has been performed. GO analysis was performed on the differential expression dataset using the same amiGO tool as was used for the *rbm-39(cnj4)* differential expression, and very little overlap in controlled terms was found. I have assisted in designing a true knockout of the Caper gene with CRISPR/Cas9, with a few different versions to get around poison exons. When this is completed a more accurate RNA-seq can be performed. Additionally, alternative splicing analysis separated by type has yet to be performed on the knockdown strain. It is currently unknown if the Caper gene is also primarily a 3' AS regulator, and if so, if it has a similar pattern. If these genes have similar molecular mechanisms, it could shed light onto how these genes have evolved, and tell us more about how it might work in the human ortholog RBM39.

Human Ortholog RBM39

The human gene RBM39 also encodes an RNA binding protein and is predicted to be involved in alternative splicing as well. The findings of this paper corroborate this possibility. RBM39 may also found to be related to oogenesis, as it acts as a transcriptional coregulator for estrogen receptors in the mice ortholog. rbm-39 could be highly similar in molecular mechanisms to RBM39, which would shed light on this gene and could help with treating genetic diseases associated with RBM39 or genes close to it on the transcriptional pathway.

Future Directions

To determine if RBM39 acts directly on the nucleotide pattern identified by the position weight matrix, a RIP-seq will be used to determine what sequences of RNA the RBM39 protein binds to. This will be done by creating an RBM39::GFP transgene.

REFERENCES

Merritt, C., Rasoloson, D., Ko, D., & Seydoux, G. (2008). 3′ UTRs are the PRIMARY regulators of gene expression in the C. elegans Germline. Current Biology, 18(19), 1476–1482. https://doi.org/10.1016/j.cub.2008.08.013

Zahler, A. M. (2005). Alternative splicing in C. elegans. In WormBook : the online review of C. elegans biology. https://doi.org/10.1895/wormbook.1.31.1

Olesnicky, E. C., Bono, J. M., Bell, L., Schachtner, L. T., & Lybecker, M. C. (2017). The RNA-binding protein caper is required for sensory neuron development in drosophila melanogaster. Developmental Dynamics, 246(8), 610–624. https://doi.org/10.1002/dvdy.24523

Wang, E., Lu, S. X., Pastore, A., Chen, X., Imig, J., Chun-Wei Lee, S., Hockemeyer, K., Ghebrechristos, Y. E., Yoshimi, A., Inoue, D., Ki, M., Cho, H., Bitner, L., Kloetgen, A., Lin, K. T., Uehara, T., Owa, T., Tibes, R., Krainer, A. R., … Aifantis, I. (2019). Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia. Cancer Cell, 35(3), 369-384.e7. https://doi.org/10.1016/j.ccell.2019.01.010

Zahler AM. Pre-mRNA splicing and its regulation in Caenorhabditis elegans. In: WormBook: The Online Review of C. elegans Biology [Internet]. Pasadena (CA): WormBook; 2005-2018. Figure 1, [Types of alternative splicing. In...]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK116073/figure/splicing_Zahler_figure1/

Gebauer, F., Schwarzl, T., Valcárcel, J. *et al.* RNA-binding proteins in human genetic disease. *Nat Rev Genet* 22, 185–198 (2021). https://doi-org.coloradocollege.idm.oclc.org/10.1038/s41576-020-00302-y

Hentze, M., Castello, A., Schwarzl, T. *et al.* A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* 19, 327–341 (2018). https://doi-org.coloradocollege.idm.oclc.org/10.1038/nrm.2017.130

Simona Antonacci, Daniel Forand, Margaret Wolf, Courtney Tyus, Julia Barney, Leah Kellogg, Margo A Simon, Genevieve Kerr, Kristen L Wells, Serena Younes, Nathan T Mortimer, Eugenia C Olesnicky, Darrell J Killian, Conserved RNA-Binding Proteins Required for Dendrite Morphogenesis in *Caenorhabditis elegans* Sensory Neurons, *G3 Genes*|*Genomes*|*Genetics*, Volume 5, Issue 4, 1 April 2015, Pages 639–653, https://doi.org/10.1534/g3.115.017327

Vong YH, Sivashanmugam L, Leech R, Zaucker A, Jones A, Sampath K. The RNA-binding protein Igf2bp3 is critical for embryonic and germline development in zebrafish. PLoS Genet. 2021 Jul 2;17(7):e1009667. doi: 10.1371/journal.pgen.1009667. PMID: 34214072; PMCID: PMC8282044.

Albarqi MMY, Ryder SP. The endogenous mex-3 3´UTR is required for germline repression and contributes to optimal fecundity in C. elegans. PLoS Genet. 2021 Aug 23;17(8):e1009775. doi: 10.1371/journal.pgen.1009775. PMID: 34424904; PMCID: PMC8412283.

Xu Y, Nijhuis A, Keun HC. RNA-binding motif protein 39 (RBM39): An emerging cancer target. Br J Pharmacol. 2020 Nov 25. doi: 10.1111/bph.15331. Epub ahead of print. PMID: 33238031.

Xu C, Chen X, Zhang X, Zhao D, Dou Z, Xie X, Li H, Yang H, Li Q, Zhang H, Di C. RNA-binding protein 39: a promising therapeutic target for cancer. Cell Death Discov. 2021 Aug 13;7(1):214. doi: 10.1038/s41420-021-00598-7. PMID: 34389703; PMCID: PMC8363639.

Prigge JR, Iverson SV, Siders AM, Schmidt EE. Interactome for auxiliary splicing factor U2AF(65) suggests diverse roles. Biochim Biophys Acta. 2009 Jun-Aug;1789(6-8):487-92. doi: 10.1016/j.bbagrm.2009.06.002. Epub 2009 Jun 18. PMID: 19540372; PMCID: PMC2748126.

Stratil A, Knoll A, Horák P, Bílek K, Bechynová R, Bartenschlager H, Van Poucke M, Peelman LJ, Svobodová K, Geldermann H. Mapping of the porcine FBN2, YWHAQ, CNN3, DCN, POSTN, SPARC, RBM39 and GNAS genes, expressed in foetal skeletal muscles. Anim Genet. 2008 Apr;39(2):204-5. doi: 10.1111/j.1365-2052.2007.01678.x. Epub 2007 Dec 19. PMID: 18162102.

Kimble J, Seidel H. *C. elegans* germline stem cells and their niche. 2013 Nov 15. In: StemBook [Internet]. Cambridge (MA): Harvard Stem Cell Institute; 2008–. PMID: 24354021.

Rose KL, Winfrey VP, Hoffman LH, Hall DH, Furuta T, Greenstein D. The POU gene ceh-18 promotes gonadal sheath cell differentiation and function required for meiotic maturation and ovulation in Caenorhabditis elegans. Dev Biol. 1997 Dec 1;192(1):59-77. doi: 10.1006/dbio.1997.8728. PMID: 9405097.

Killian, D. J., & Hubbard, E. J. A. (2004). C. elegans pro-1 activity is required for soma/germline interactions that influence proliferation and differentiation in the germ line. In *Development* (Vol. 131, Issue 6). https://doi.org/10.1242/dev.01002

Klug, W. S., Cummings, M. R., Spencer, C. A., Palladino, M. A., & Killian, D. (2020). *Concepts of Genetics* (12th ed.). Pearson.

Albert Hubbard, E. J., & Schedl, T. (2019). Biology of the caenorhabditis elegans germline stem cell system. *Genetics*, *213*(4). https://doi.org/10.1534/genetics.119.300238

Pepper, A. S. R., Lo, T. W., Killian, D. J., Hall, D. H., & Hubbard, E. J. A. (2003). The establishment of Caenorhabditis elegans germline pattern is controlled by overlapping proximal and distal somatic gonad signals. *Developmental Biology*, *259*(2). https://doi.org/10.1016/S0012-1606(03)00203-3

McCarter, J., Bartlett, B., Dang, T., & Schedl, T. (1997). Soma-germ cell interactions in Caenorhabditis elegans: Multiple events of hermaphrodite germline development require the somatic sheath and spermathecal lineages. *Developmental Biology*, *181*(2). https://doi.org/10.1006/dbio.1996.8429

Kundinger, S. R., Dammer, E. B., Yin, L., Hurst, C., Shapley, S., Ping, L., Khoshnevis, S., Ghalei, H., Duong, D. M., & Seyfried, N. T. (2021). Phosphorylation regulates arginine-rich RNA-binding protein solubility and oligomerization. *Journal of Biological Chemistry*, *297*(5). https://doi.org/10.1016/j.jbc.2021.101306

Fung, P. A., Labrecque, R., & Pederson, T. (1997). RNA-dependent phosphorylation of a nuclear RNA binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(4). https://doi.org/10.1073/pnas.94.4.1064

Burset, M., Seledtsov, I. A., & Solovyev, V. v. (2000). Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Research*, *28*(21). https://doi.org/10.1093/nar/28.21.4364

Gonadal sheath tagged GFP marker strain:

Hall, D.H., Winfrey, V.P., Blaeuer, G., Hoffman, L.H., Furuta, T., Rose, K., Hobert, O., Greenstein, D., 1999. Ultrastructural features of the adult hermaphrodite gonad of Caenorhabditis elegans: relations between the germ line and soma. Dev. Biol. 212, 101–123.

DTC tagged GFP marker strain:

Blelloch, R., Anna-Arriola, S.S., Gao, D., Li, Y., Hodgkin, J., Kimble, J., 1999. The gon-1 gene is required for gonadal morphogenesis in Caenorhabditis elegans. Dev. Biol. 216, 382–393.