Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Binta Jalloh Date

The *Drosophila* Nab2 RNA-Binding Protein Regulates Splicing Neuronal RNAs

By

Binta Jalloh
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Genetics and Molecular Biology

_____
Kenneth H. Moberg, Ph.D.
Advisor

_____
Anita H. Corbett, Ph.D.
Advisor

_____
Peng Jin, Ph.D.
Committee Member

_____
Yue Feng, Ph.D.
Committee Member

_____
Gary J. Bassell, Ph.D.
Committee Member

_____
Tamara Caspary, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

The *Drosophila* Nab2 RNA-Binding Protein Regulates Splicing Neuronal RNAs

By

Binta Jalloh
B.S., Montclair State University, 2012

Advisor: Anita H. Corbett, Ph.D. & Kenneth H. Moberg, Ph.D.

An abstract of
a dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Genetics and Molecular Biology
2021

Abstract

A large number of RNA binding proteins (RBPs) play important roles in RNA processing and in regulating gene expression at the post-transcriptional level. It is therefore not surprising that dysregulation of key RNA processing steps due to mutations in genes that encode these RBPs are often linked to tissue-specific diseases. A majority of which affects the highly specialized neuronal cell, resulting in neurodevelopmental and or neurodegenerative disorders. One such RBP that is essential for neurons and required for proper brain development and function is the polyadenosine RNA-binding protein, ZC3H14 (Zinc finger Cysteine-Cysteine-Cysteine-Histidine containing protein 14). Mutations in the gene encoding ZC3H14 has been linked to a heritable and a non-syndromic, autosomal recessive form of intellectual disability. To investigate the role(s) of ZC3H14 in the brain, we employed the *Drosophila melanogaster* ortholog of ZC3H14, Nab2, to study which RNAs are affected upon Nab2 loss. We show here that *Drosophila* Nab2 is implicated in splicing. Using RNA from adult *Drosophila* heads, we performed the first global transcriptomic analysis on Nab2 heads using RNA sequencing experiments. The *Nab2* mutants show defects in RNA splicing compared to controls. Further analysis comparing mutant females to controls reveal dysregulation of several pathways that are important for proper brain function and sheds light into which RNA transcripts are most affected by the loss of Nab2 in neuron enriched tissues. This high-throughput and unbiased RNA sequencing dataset provides a novel functional role for *Drosophila* Nab2 in brain tissue by regulating the splicing of a specific subset of RNAs. Of the ~150 affected neuronal RNAs, the most significant splicing defects observed is the female-specific retention of a male-specific exon in the sex determination factor gene, *Sex lethal*.

The *Drosophila* Nab2 RNA-Binding Protein Regulates Splicing Neuronal RNAs

By

Binta Jalloh
B.S., Montclair State University, 2012

Advisor: Anita H. Corbett, Ph.D. & Kenneth H. Moberg, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Genetics and Molecular Biology
2021

Acknowledgements

This work was funded by the National Institute of Health NRSA Fellowship (F31 NS 103595 to Binta Jalloh). I would like to thank my advisors Kenneth Moberg and Anita Corbett for taking me on as their first co-mentored Black graduate student to work on this project. All my achievements over the years during my graduate career at Emory is proof of the ample mentoring and constant support I have received from my iron-willed mother, Habibatu Bah Jalloh and my life-long mentor, Dr. Nick Ingoglia. This manuscript is the product of the joint collaborative efforts of the Moberg and Corbett labs. I am eternally grateful to all members of my thesis committee for their insightful and probing questions over the years. I am also thankful for Chris Rounds, Drs. Rick Bienkowski and Milo Fasken for all their intellectual and scientific contributions to my thesis work.

Table of Contents

List of Figures
Chapter 1:

Chapter 2:

Tables

Chapter 3:

Chapter 4:

Chapter 1

General Introduction

1.1 The Central Dogma of Molecular Biology

The discovery of the structure of deoxyribonucleic acid, DNA, the molecule that carries all genetic information, was a revolution in the field of molecular biology. The description of the right-handed DNA double helix structure by James Watson and Francis Crick, building on the insights from Rosalind Franklin's X-ray image of DNA, paved the way for the concept known as the Central Dogma of Molecular Biology (1). The Dogma describes the linear transfer of genetic information from DNA to ribonucleic acid, RNA, to proteins (2). However, at that time, little was known about the machinery involved in the transfer process, or how the transfer of this information was regulated. Today, we know the process through which DNA is converted to the intermediate RNA as transcription, which takes place in the nucleus of a eukaryotic cell. The process where the RNA is converted to the final protein product as translation, which takes place in the cytoplasm of the cell.

In all eukaryotic cells (except mature reticulocytes or red blood cells), DNA is stored in the nucleus (1). Nuclear DNA is typically present in protein/DNA complexes called chromatin. Chromatin can exist as easily accessible (de-condensed or 'open' euchromatin) or condensed 'closed' heterochromatin. The 'open' conformation of chromatin is often associated with actively transcribed DNA and 'closed' with silenced genes that are not being expressed (3). The state of chromatin impacts the rate of DNA transcription, impacting which genes are expressed.

The nucleosome, the fundamental unit of the chromatin, is composed of core histone proteins with N-terminal histone tails that can be post-translationally modified to regulate gene expression (4). Modifications such as acetylation, methylation, and phosphorylation can occur at

one or multiple sites on the histone tails, with each modification associated with a specific function (4). The fields of molecular genetics and epigenetics are based on modification of histone tails and the DNA wrapped around these histone proteins. Modification of histones can directly affect transcription and nucleosome stability to regulate gene expression (5). Though regulation of gene expression begins at the DNA level, the experiments presented in this thesis are focused on the next level of gene expression, the processing of RNAs with a focus on messenger RNAs (mRNAs).

1.2 Regulation of Gene Expression at The RNA Level

The completion of the Human Genome Project revealed that only a small percentage (1.5%) of the human genome consist of protein-coding genes (i.e. mRNAs), while the remaining 98.5% contains sequences that are not translated into protein (non-coding regions) or "junk" DNA (6). Although only about 20,000 protein-coding genes exist in the human genome, the complexity associated with humans and other higher order organisms is attributed to the variety of protein isoforms, produced as a result of alternative splicing and 3'-end processing of RNAs (7-9). A variety of small and long non-protein coding RNAs are also involved in protein synthesis. The non-coding RNAs such as ribosomal RNAs, transfer RNAs, micro RNAs and small interfering RNAs (involved in RNA interfering pathway to silence genes), piwi-interacting RNAs, small nucleolar RNAs, small nuclear (spliceosomal) RNAs, all have specific expression patterns relevant for regulating gene expression (6, 7, 10). Dysregulation of many of these regulatory RNAs is implicated in cancers and other complex human diseases. Genetic alteration of the small non-coding micro RNA (miRNA) is estimated to effect about 60% of human genes involved in regulatory pathways including cancer (11). miRNAs regulate several mRNAs using the RNA-induced silencing complex by binding to specific sequences on the 3'-untranslated region (3'UTR) of the mRNA transcript to block translation initiation or induce mRNA degradation as shown in **Figure 1—1** (12-14).

Unlike the stable, double stranded, and three-dimensional structure of DNA, mRNA is more unstable, is predominantly single stranded, and can assume secondary and tertiary structures. The secondary and tertiary structures of mRNA contain specific sequences and motifs that are important for post-transcriptional regulation of gene expression (15). The process of producing mature mRNAs that can direct the production of functional proteins in the cytoplasm is not an easy feat. The differential expression of mRNAs by small non-coding miRNAs encoded by the same genetic blueprint (DNA) in eukaryotic cells and alternative splicing of these mRNAs, are all essential post-transcriptional regulation mechanisms that give rise to protein diversity (see **Figure 1—1**), (16). Recent advances in the field of molecular genetics, biochemistry, and RNA sequencing have revealed huge complexity in the mechanisms through which mRNAs are post-transcriptionally processed.

1.3 Gene Expression Is Modulated By Post-Transcriptional Regulation

Proper transcription of RNA from the DNA template in the cell nucleus is vitally important. However, it is the post-transcriptional processing of RNA to form mature mRNA that dictates the protein isoforms expressed. As shown in **Figure 1—1**, post-transcriptional regulation of gene expression at the mRNA level involves a variety of processing steps that begin in the nucleus of eukaryotes. First, as the newly synthesized nascent RNA molecule is synthesized and released from the RNA polymerase, a protective cap is immediately added after the ~20 nucleotides are synthesized at the 5 prime-end (5'-end) of the RNA transcript. The addition of this 7-methylguanosine cap protects the premature mRNA (pre-mRNA) from 5'exonucleolytic degradation in the nucleus. Provides stability to the mature mRNA once it enters the cytoplasm for translation (17). This 5'-7-methylguanosine cap (5'-m^7G) is so essential, that it is evolutionarily conserved across all eukaryotes, and it is the first modification that takes place co-

transcriptionally in the nucleus (18). Moreover, the 5'-m^7G cap is required in eukaryotic mRNAs for pre-mRNA splicing, polyadenylation, and ultimately for translation in the cytoplasm (19).

Another key nuclear mRNA processing step that occurs at the 3'-end of the pre-mRNA is cleavage/polyadenylation. This involves the cleavage and the addition of a polyadenylated tail or poly(A) tail to the mRNA (20). The polyadenylation step is integral for proper export of mRNAs from the nucleus to the cytoplasm, and stability of the mRNA (20). This two-step nuclear 3'-end processing machinery is an important biological regulator of RNA metabolism as dysregulation of this complex results in degradation of the RNA by miRNA or by deadenylation proteins (see **Figure 1—1**).

The advances in technologies, and decades of research dedicated to understanding the apparatus involved in the 3'-end processing, has shed light on an important phenomenon. Human genes can contain more than one poly(A) site which results in alternative polyadenylation producing several RNA transcripts with alternative 3' ends (21). This means a single gene can generate multiple variants of the same RNA transcript based on which poly(A) site is used. In addition, the length of the poly(A) tail also adds a layer of functional regulation during nuclear pre-mRNA processing (22). In budding yeast for example, the poly(A) tail is ~70-80 adenosine nucleotides (As), while in human cells the average poly(A) tail consists of ~250-300 adenosines (21). Like a zip code to an address, the length of the RNA's poly(A) is very important for reaching the final destination, which is translation of the transcript. **Figure 1—1** illustrates the importance of polyadenylation and deadenylation of transcripts to regulate poly(A) tail lengths.

A key RNA processing step beyond capping and cleavage/polyadenylation, is splicing. The advances in high-throughput sequencing technologies has made genome-wide profiling of dynamic spicing patterns, such alternative splicing (see **Figure 1—1**) in complex organisms

possible, and provide insight into the protein isoform diversity observed in different cell types and tissues (16). Splicing involves the removal of parts of the pre-mRNA called introns from the exons. Proper splicing depends on the assembly and formation of the large macromolecular complex called the spliceosome, which is initiated by the consensus sequences found at the intron-exon junctions or splice sites (23). Depending on the splice sites employed by the spliceosome, a single mRNA molecule can produce a variety of mRNA products, an important contributor to the diversity of gene expression (23). The excision of the introns followed by the joining of exons is carried out by the spliceosome complex. Introns are removed during the splicing reaction through the catalysis of two transesterification steps. These processes are facilitated by a set of five small nuclear ribonucleoproteins (snRNPs) in the spliceosome complex along with other accessory proteins that assembles onto the introns for removal (23). Different components of these snRNPs (U1, U2, U4, U5 and U6) in the spliceosome interact with splicing factor RNA-binding motifs to regulate splicing (24). The mRNAs that are not properly spliced are targeted for destruction.

The final mRNA processing step could be thought of as quality control, including turnover/degradation. Aberrantly processed RNAs (coding and non-coding) are targeted by the evolutionarily conserved, nuclear RNA exosome complex for degradation (25). The RNA exosome is a complex multi-purpose decay machinery involved in RNA surveillance and quality control, while also managing numerous RNA processing reactions (26). The barrel-shaped RNA/protein complex targets and threads aberrant RNAs for destruction in both the nucleus and the cytoplasm through a 3' to 5' exonuclease activity (27). Precise processing and complete degradation of target RNAs by the RNA exosome is critically important, as mutations in any of the genes that encode the nine structural subunits that comprise the RNA exosome (EXOSC1-EXOSC9) cause tissue-specific diseases (25, 27-29). Mutations in the human *EXOSC3* gene are linked to an autosomal

recessive neurological disorder called Pontocerebellar Hypoplasia Type 1b (28). Chapter 3 of this dissertation specifically explores the role of EXOSC3 in neurons.

The role of the RNA exosome in mRNA quality control, regulation of mRNA levels, and turnover depends on co-factor proteins, which provide the versatility required to monitor the various RNAs constantly undergoing co-transcriptional and post-transcriptional processing events (25, 27). All these post-transcriptional processing steps require the coordination and appropriate expression of RNAs and proteins (14). The major nuclear processing events of these pre-mRNAs to form mature mRNAs are modulated by RNA-binding proteins (RBPs).

1.4 Post-Transcriptional Regulation By RNA-Binding Proteins (RBPs)

An RNA-binding protein (RBP) is defined as a protein that binds RNA by using one or more RNA-binding domain to regulate the fate and function of that bound RNA (30). However, there are exceptions to this rule as recent discoveries of the structure of the complex molecular machine called the spliceosome has revealed the existence of protein-RNA complex interactions that expand the functional definition of RNA-binding domains (30). Over the decades, several hundred RBPs have been discovered and studied extensively. **Figure 1—2** illustrates some well-established functions associated with RBPs in RNA metabolism from transcription in the nucleus to degradation in the cytoplasm (31).

Many RBPs associate with the pre-mRNA from transcription initiation to termination (32). As described earlier, the nascent mRNA molecule is modified at the 5'end by a 7-methylguanosine cap enzyme. The cap structure and cap-binding proteins are evolutionarily conserved and play important roles in the regulation of gene expression (19, 33). The 5'-end capping and 3'-end processing with cleavage/polyadenylation to add a polyadenosine (poly(A)) tail occur co-transcriptionally. The 3'end-bound poly(A) tail-binding proteins control the addition of the

appropriate length of As at the tail, which is essential for stability and translational regulation (22). Short poly(A) tail length is associated with repressed translational states, whereas elongation of the tails is thought to promote translation (34). However, long vs. short poly(A) tail length is relative (transcript and organism dependent). Thus, two major determinants of mRNA stability are the m^7G cap and the poly(A) tail (35).

The mechanism through which poly(A) tail length is controlled, is understood in some detail. The process involves a specific type of RBP called Poly(A) Binding Proteins (PABPs), which protect the transcript from the many cellular RNA exonucleases. A major challenge with studying RNA binding proteins is that they may have multiple critical functions, as illustrated in **Figure 1—2** (31, 36, 37). Thus, multifunctional PABPs control fate of mRNAs by promoting poly(A) tail synthesis, regulating tail length, and stimulating mRNA maturation, which are all critical in regulating gene expression (35, 38).

1.5 Zinc Finger (ZnF) Poly(A) Binding Proteins (PABPs)

In yeast and metazoans, synthesis and regulation of the poly(A) tail length is controlled by poly(A) polymerase (PAP) and poly(A) RNA-binding proteins (PABPs), which bind their target RNA via specific RNA recognition motifs (RRMs) (38, 39). Most PABPs regulate target RNAs via an RRM. However, in 1993, a new class of PABP that bind polyadenosine RNA via a Cys-Cys-Cys-His-type (CCCH) zinc finger (ZnF) domain, was identified, initially in *Saccharomyces cerevisiae* (40). Using ultraviolet (UV) light to covalently cross-link proteins directly bound to RNAs in living yeast cells, investigators were able to detect, isolate, and purify a variety of nuclear polyadenylated RNA-binding (Nab) proteins (40). Further characterization of the second Nab protein, Nab2, revealed three fundamental facts. The yeast Nab2 protein (1) is localized predominantly in the nucleus, (2) Nab2 associates with poly(A) RNAs, and (3) the *NAB2* gene is essential for viability (40).

The Nab2 protein belongs to an evolutionarily conserved family of polyadenosine RNA binding ZnF proteins, with functional roles in regulating the length of the poly(A) tail of mRNAs and export from the nucleus (40-42). The yeast Nab2 protein is composed of an N-terminal Proline-Tryptophan-Isoleucine (PWI)-like domain followed by a Glutamine-rich (QQQP) and Arginine-Glycine (RGG) motifs. The RGG motif is a nuclear localization signal (NLS) and after the RGG are seven CCCH ZnFs that bind to polyadenosine RNAs (43). In budding yeast, Nab2 zinc fingers 5-7 comprise the major domain that mediates high binding affinity to polyadenosine RNAs (43). These domains are all evolutionarily conserved across different species as shown in **Figure 1—3**.

Poly(A) tail length control is an important evolutionarily conserved function for Nab2 (44). Compared to mammals with a maximum of 200-300 adenine nucleotides long poly(A) tails, yeast cells contain considerably shorter (70-90 adenine nucleotides) poly(A) tail lengths (41). The link between poly(A) tail length and export from the nucleus was elegantly demonstrated using a standard nitrocellulose filter binding assay to first show Nab2 binds poly(A) RNA in a similar manner to Pab1, which is about ten times more abundant than Nab2 (41). Then, using poly(U)-Sephadex chromatography, Nab2 depleted cells were shown to have longer poly(A) tails (41). Finally, by using fluorescence *in situ* hybridization or FISH, an increased accumulation of poly(A) RNAs in the nucleus of Nab2 depleted cells was observed compared to mRNA export mutants (41). These experiments implicate Nab2 protein in controlling mRNA poly(A) tail lengths as well as export from the nucleus.

The ability of the yeast Nab2 protein to recognize and bind poly(A) RNAs with high affinity due to its ZnF binding domains instead of an RRM was further demonstrated and characterized using gel-shift assays (GST) and fluorescence correlation spectroscopy (FCS) (45). The requirement for nuclear Nab2 and cytoplasmic Pab1 in poly(A) tail binding and length control for stability as

well as export are all essential functions, as deletion of *Nab2* or *Pab1* results in turnover/degradation of mRNAs and cell death (35). In yeast, metazoan, and mice, these defects often lead to lethality, but in humans, defects in these processes often result in abnormal behaviors and tissue-specific diseases.

1.6 PABPs in Human Diseases

In humans, the ubiquitously expressed PABP with a PAP binding site and a single RRM for poly(A) RNA binding, is the polyadenylate-binding nuclear protein 1 (PABPN1) (46). PABPN1 cooperates with the PAP to regulate poly(A) tail length (46). With both nuclear and cytoplasmic roles in RNA quality control, PABPN1 may have a distinct function in different cell types. Thus, a mutation in the human *PABPN1* gene that causes an expansion of alanine residues in the N-terminal domain of the protein results in a muscle specific disorder called Oculopharyngeal Muscular Dystrophy (OPMD) (46). OPMD is a late-onset, autosomal dominant, skeletal muscle disease affecting the eyelids, proximal limbs, and pharyngeal muscle tissues often causing difficulty in swallowing (47). PABPN1 is essential in mice, and no mouse model of the cytoplasmic PABPC exists yet, presumably because PABPC is also essential. The current view of the molecular mechanism of OPMD tissue-specific pathology is described in a model where already low PABPN1 protein expression levels under normal conditions become even lower due to the alanine expansion in OPMD muscle tissues (33, 46, 48). This model is consistent with the differential expression of genes in different cells and tissues, and perhaps can be applied in other tissue-specific disease conditions.

Another major nuclear PABP also ubiquitously expressed and associated with a tissue-specific disease, is the human orthologue of the yeast Nab2 protein, Zinc finger Cysteine-Cysteine-Cysteine-Histidine-type containing 14 (ZC3H14) protein. Unlike PABPN1 which contains a conventional RRM, human ZC3H14, like yeast Nab2 contains a CCCH zinc fingers RNA binding

motif (45). Mutations in the human gene encoding the ZC3H14 protein are linked to an autosomal recessive form of intellectual disability (49). This association was described in a large-scale study to define genes linked to intellectual disability (50). Intellectual disability is a developmental disorder that manifests before the age of 18, and affected individuals have significant limitations in both intellectual functioning and adaptive behaviors (51, 52).

Like many RBPs, ZC3H14 is ubiquitously expressed, but mutations cause a neurological disorder (49, 53, 54). The tissue-specific disease phenotype, due to a mutation in a single gene encoding an RBP, broadly expressed in all cells pose the same phenomenon as described for OPMD. This implies loss of these RBPs affects one tissue or cell type differently. A catalog of over 1,500 human RBPs and their links to human genetic diseases, show a tissue-specific requirement for RBPs (31, 32, 53). A continuously growing list of RBPs linked to different human diseases, including ZC3H14, is illustrated in **Figure 1—4** (32). A major question is what specific function these RBPs play in different tissues and are these proteins and their functions evolutionarily conserved.

1.7 ZC3H14 Is An Evolutionarily Conserved PAB Protein

The ZC3H14 protein is evolutionarily conserved across species, from budding yeast to mammals. The structural domains of this PABP across different model organisms are shown in **Figure 1—3** (43). The budding yeast *Saccharomyces cerevisiae* Nab2 (*Sc*Nab2), the fission yeast *Schizosaccharomyces pompe* (*Sp*Nab2), the *Chaetomium thermophilum* (CtNab2), the fruit fly *Drosophila melanogaster* (*Dm*Nab2), the round worm *Caenorhabditis elegans* (*CeSUT*-2), and *Homo sapiens* (HsZC3H14) all have similar domain structures. This overall conservation of structure is important for protein function.

1.7.1 Structure & Function of ZC3H14/Nab2 Protein

The N-terminal domain of the mammalian ZC3H14 (see **Figure 1—3**) protein contains a conserved Proline-Tryptophan-Isoleucine (PWI) fold, which is predicted to serve as a protein-protein interacting domain (43). This PWI-like domain is required for proper nuclear export of RNAs from the nucleus in budding yeast (45). Although this domain is highly conserved, little is known about the function of this domain in mammals. Removing this domain in yeast results in impaired cell growth and accumulation of poly(A) RNAs, consistent with its function in nuclear export (55). However, studies to explore the role of this domain have not been performed beyond the yeast model, so the conservation of this function requires further investigation.

There are two predicted classical nuclear localization signals (cNLSs) in higher eukaryotes that may function in nuclear import, while the RGG motif within budding yeast Nab2 serves as a nuclear localization signal, and they are all located in about the same place (all in about the middle, see **Figure 1—3**) (43). The ZC3H14 protein contains five tandem zinc fingers (ZnFs) at the C-terminus, the ZnFs are marked red in **Figure 1—3**. The total number of ZnFs varies across different species as shown in **Figure 1—3**. These ZnFs are necessary and sufficient for binding polyadenosine RNA (43, 45). In budding yeast, mutations within ZnFs 5-7 disrupts polyadenosine RNA binding and defects in poly(A) tail length (56).

1.7.2 ZC3H14 Controls Poly(A) Tail Length

In mice, ZC3H14 is required for proper poly(A) tail length based on analysis of bulk RNA isolated from cortex and hippocampus tissues in the brain of normal (control) and loss of function (mutant) *ZC3H14* mice (Rha et al., 2017). Mutant *ZC3H14* mice exhibited a significant increase in the length of the poly(A) tails in brain hippocampus tissue relative to control animals and relative to other tissues such as the liver (Rha et al., 2017). Also ZC3H14 knock-down in cultured cells show an

increase in poly(A) tail length (44).These experiments strongly supported a tissue-specific requirement for ZC3H14 as well as a functional role in restricting poly(A) tail length (Rha et al., 2017). Although the exact mechanism through which ZC3H14 controls poly(A) tail length is unknown, depletion of ZC3H14 is not lethal in worms or mammals (as it is in yeast and metazoan). Another possibility is that in yeast and metazoan, Nab2, is involved in more biological processes important for cellular development and viability, while in worms (SUT-2) and mammals (ZC3H14) the RNAs bound and regulated by this RBP are required for neurons but are not essential for viability. This idea is consistent with the *ZC3H14* patients, mouse model of *ZC3H14* mutations, as well as worm *SUT-2* mutants, which all exhibit neuropathic phenotypes (57).

1.8. A Role For ZC3H14 In RNA Splicing

Previous experiments on mammalian mouse cells show that ZC3H14 co-localize with poly(A) RNAs in nuclear speckles, which imply a role for ZC3H14 in nuclear RNA processing (49). Nuclear speckles are dynamic subnuclear structures located in the interchromatin regions of mammalian cells that are enriched in pre-mRNA splicing factors (58). This discovery is consistent with yeast Nab2 functions in the nucleus, as ZC3H14 orthologues in yeast models have been implicated in transcription, RNA poly(A) tail length restrictions, export from the nucleus, and splicing of pre-mRNAs (39, 41, 55, 59). Evidence supporting a role for mammalian ZC3H14 in splicing comes from a proteomic analysis revealing an interaction between ZC3H14 and components of the spliceosome, specifically the U2AF2/U2AF (39).

These finding reflect a major challenge in studying RBPs as they may have many functions. They may partner with different proteins. In addition, they may cause both direct and indirect effects. While less is known about the biochemical functions of ZC3H14/Nab2 in metazoans than in yeast, there are several lines of evidence that implies this RBP has a role in modulating splicing events. Though ZC3H14/Nab2 can interact with splicing factors, little is known about which RNAs, particularly in neurons, might be regulated by ZC3H14/Nab2 (39, 42). Furthermore, the tissue-specific requirement for ZC3H14 in the brain also poses an interesting question. Why are mutations in the human *ZC3H14* gene linked to a form of autosomal recessive, non-syndromic intellectual disability? Why is the brain affected upon the loss of this ubiquitously expressed poly(A) binding RBP? **The experiments described in this dissertation exploit the model organism *Drosophila melanogaster* (fruit fly) to investigate the role(s) of ZC3H14/Nab2 in post-transcriptional regulation in the fly brain**. The work presented in Chapter 2 investigates roles of

Nab2 in processing mRNAs in neurons and in Chapter 3 to model RNA exosome linked human diseases.

1.9 *Drosophila Melanogaster* as A Model to Study Neurological Diseases

The fruit fly, *Drosophila melanogaster*, is a well-established system to model human neurological diseases. The human central nervous system (CNS) is comprised of 100 billion neuronal cells (neurons) and there are 200,000 neurons in the fruit fly brain (60). In both humans and flies, neurons are interconnected by their axons and dendrites forming a precise neuronal network that is essential for brain function (61). Therefore, any genetic mutations in genes, important for neurons, may have devasting effects on CNS function.

Despite having a brain that is significantly smaller, the whole fly brain is the size of the tip of a fountain pen. The CNS of the fruit fly is comprised of similar neuronal and glial cell types as the human CNS; thus making it a potentially powerful system to study and identify important molecular mechanisms in neurodevelopment and neurological diseases (61). Besides being small and having a complex brain, the fly is extremely fertile (females lay ~200 eggs a day), with a short generation time (10-day life cycle), making genetic experiments (over several generations) fast and easy (62). The fruit fly also shares ~75% of its genome with humans, making it possible to study ~75% of human genetically linked diseases in this model (61, 62). In fact, key features of several neurodegenerative diseases (Parkinson's and Alzheimer's) and neurodevelopmental disorders (Fragile-X Syndrome and Intellectual Disability) have been discovered using the *Drosophila* model system (63).

Current investigations using classic genetic tools in the *Drosophila* system to study multifaceted disease conditions, including heritable neurodevelopmental disorders such as Fragile-X Syndrome, have provided essential mechanistic insights into the molecular and cellular

pathogenesis of monogenic (single gene) disorders (64). A perfect and a well-developed example of *Drosophila* modeling of heritable neurodevelopmental disorder in humans is the Fragile X Mental Retardation gene 1 (*FMR1*) in Fragile-X Syndrome (FXS) (63). This hereditary neurodevelopmental disorder is caused by a mutation in the *FMR1* gene, where a CGG expansion in the 5'UTR (untranslated region) results in the silencing of the *FMR1* gene. This causes the loss of function of the Fragile X Mental Retardation Protein (FMRP) (65). FMRP is a ubiquitously expressed RBP enriched in neurons of the hippocampus, which is the center for learning and memory in mammals (66). FMRP plays roles in RNA metabolism important for local mRNA regulation and protein synthesis in dendritic and axon neurons (localization and transport, see **Figure 1—1**) (14, 67).

The hallmark of FXS in humans is intellectual disability and developmental delays such as autism (68). Intellectual disability is the most common early-onset developmental disorder characterized by impairments in cognitive functions, adaptive behaviors, and is associated with an intelligence quotient (IQ) of less than 75 (52, 69-71). In the fly model, all these functions and behaviors can be recapitulated. For example, some FXS patients have delayed motor development and functions that can be compared to uncoordinated behaviors in flies, which can be measured using a locomotion or climbing assay (63). In addition, olfactory learning and courtship behaviors in flies are another way to test the effects of impaired CNS functions often associated with intellectual disabilities (63).

1.9.1 A *Drosophila* Disease Model to Study ZC3H14 Linked Disease

Like FMRP, the ZC3H14 protein is important for brain development and function. Mutations in the *ZC3H14* gene result in a non-syndromic, heritable form of autosomal recessive intellectual disability (49). One pathogenic *ZC3H14* mutation creates a nonsense mutation located in exon 6 of *ZC3H14* gene (see R154X as shown in **Figure 1—5C**). The other *ZC3H14* gene mutation is a 25-

bp deletion within the 3'UTR as shown in **Figure 1—5C** (49). The pedigree of affected individuals is shown as Family-1 and Family-2 in **Figure 1—5A**, **5B**, respectively. Although these affected family members are all male, the intellectual disability linked gene *ZC3H14* is not an X-linked disease as in FXS. The *ZC3H14* gene is located on chromosome 14 in mammals (49). To date, there are other *ZC3H14* patients who are not male and that are not from consanguineous families that have been identified.

The individuals affected with mutations in the *ZC3H14* gene shown in **Figure 1—5** is all from consanguineous marriages and autosomal recessive intellectual disability is most common in consanguineous families (49, 72). This means finding these mutations in multiple family members in the general population is very unlikely. In fact, the *ZC3H14* gene was identified after a large-scale mapping and linkage association study investigating the molecular causes of autosomal recessive intellectual disability in 200 consanguineous families in rural Iran (49, 50).

Following the discovery of *ZC3H14* mutations as a cause for intellectual disability, motivated generation of a fly model to study the role of Nab2 in a multicellular organism (49). A *Nab2* mutant fly was generated using imprecise excision of a P-element within the *Nab2* gene, causing a 1.5 kilo base deletion in the gene resulting in a loss of function Nab2 in the flies (49). An isogenic control, with a precise excision of the P-element was also generated to serve as an 'wild-type' control (49). This model was employed to provide key initial insights into the function of Nab2 in *Drosophila*.

First, Nab2 is essential for viability as observed in budding yeast, and maternal depletion of *Nab2* is embryonic lethal. However, genomic deletion of *Nab2* in the developing embryo results in ~3% viable adult flies called "escapers" (49). Second, mutant *Nab2* flies exhibit longer poly(A) tails, indicating that *Drosophila* Nab2 is required for proper poly(A) tail length control (49). Finally,

and most importantly, Nab2 is required in neurons for flight, locomotion, and courtship behaviors (49). The neurological deficits observed are important for movement and learning new behaviors, however, there may be many other deficits that include sensory functions such as vision and olfaction that Nab2 is also regulating.

One of the most important pieces of data from fly Nab2 to date is its role in neurons. Expressing Nab2 protein in *Nab2* mutant neurons only shows improved locomotion and viability, serving as the only evidence that Nab2 has a key role in neurons (49). This was the first link of a poly(A) RNA binding protein required for neuronal function and provided new insights into the human brain disorder in the *ZC3H14* patients diagnosed with intellectual disability (49). These data indicate a really key role for *Drosophila* Nab2 in neurons. A major question that remained was why only a particular subset of brain neurons are affected in ZC3H14/Nab2 loss of function conditions?

1.10 Role & Requirement For *Drosophila* Nab2 In Neurons

The ZC3H14 orthologue in *Drosophila* is essential for viability, and poly(A) tail length control, but most importantly, depletion of Nab2 in neurons causes defects in short-term memory with intact learning ability (73). As intellectual disability is the only manifestation of disease when *ZC3H14* is altered, it is likely that human ZC3H14 could play a role in CNS development and function. The fact that patients with only mutations in *ZC3H14* have intellectual disability links ZC3H14 to intellectual disability. The effect of a single gene mutation on memory associated behaviors was described initially by Seymour Benzer (74, 75). His work led to the neurogenetic dissection of animal behavior affecting the ability of an organism to store and acquire newly formed information in their nervous system (short-term memory) (76). These works, over five decades ago, laid the foundations and provided tools to investigate the molecular and cellular

mechanism(s) underlying learning and memory in the *Drosophila* model and other mammalian (mice) model systems (76).

Neuron-specific re-expression of fly Nab2 or expression of human ZC3H14 in *Nab2* mutant flies is sufficient to rescue the behavioral defects (locomotion, courtship), cementing the idea that ZC3H14 is a functional orthologue of *Drosophila* Nab2 (73). The impaired neuronal functions in *Drosophila Nab2* mutants has been linked directly to a specific set of mushroom body (MB) neurons in the fly brain (73). The MBs originate from a bundle of axons extending from a group of ~200 neurons (Kenyon cells) to form symmetrical (twin neuropil structures) with alpha, alpha prime, beta, beta prime, and gamma lobes (α/α', β/β', and γ-lobes), and these structures require proper developmental cues and guidance to perform learning and memory functions (77, 78). The axon projections from the MBs are implicated in learning and memory in the developing fly brain (77). MB developmental defects are present in *Drosophila* models of FXS, Down's Syndrome, and other neurological diseases (79, 80). However, the genes involved in MB development and why these neurons are the only ones affected in single gene mutations, is still unclear.

Drosophila Nab2 mutant flies show mis-projection of the axons that comprise the alpha(α) and beta (β) lobes of the MBs (73). **Figure 1—6** shows the two distinct MB defects associated with Nab2 loss: (1) overgrowth/over projection of β axons across the midline resulting in fused β lobes and (2) missing or thinned α lobes, (see **Figure 1—6B**) (73). The penetrance of these MB defects in *Nab2* mutant adult flies is close to 80% based on the scoring criteria used (see **Figure 1—6C**) (73). The MB structures are imaged using the neuronal adhesion protein, FasII (Fasciclin-II) because these proteins are highly enriched in the alpha and beta axons of the MBs (73, 81). These discoveries in *Drosophila* MBs were later complemented by studies in the mouse model of ZC3H14 and cultured ZC3H14 hippocampal neurons (66, 82), supporting an evolutionarily conserved role for ZC3H14/Nab2 in CNS neurons.

Defining the nuclear and cytoplasmic roles of Nab2 is key to explaining the molecular mechanism(s) of how and perhaps why ZC3H14/Nab2 is required for neuronal development of MBs and brain function. Previous work on cytoplasmic control of neuronal mRNAs by *Drosophila* Nab2, show this RBP can genetically interact with *Drosophila* fragile X protein ortholog, dFMRP, which is a translational repressor, to co-regulate MB neurons (66). This was the first insight to defining a cytoplasmic role for ZC3H14/Nab2 in translation control of mRNAs of neurons by interacting with proteins associated with the translation machinery and with dFMRP. However, the nuclear role(s) of ZC3H14/Nab2 are still undefined. Chapter 2 of this dissertation explores the transcriptomic landscape of *Drosophila Nab2* mutants and provides a novel insight into a possible nuclear role for ZC3H14/Nab2 in regulating specific changes in levels of mRNAs in brain tissues.

1.11 ZC3H14/Nab2 Show Specificity For Target RNAs

As a PABP, *Drosophila* Nab2 might interact with other RBPs to regulate expression of a specific set of mRNAs to modulate gene expression in MB neurons required for proper brain development. Previous work investigating the role of ZC3H14/Nab2 in modulating target mRNAs implicated in neuronal functions was done by a transcriptomic wide-analysis of ZC3H14 depleted cells (83). These experiments showed only approximately 1% of expressed mRNA transcripts were affected in ZC3H14 depleted cells compared to controls, indicating only a small and a specific subset of mRNAs are regulated by ZC3H14 (83). Although PABPs are known to bind poly(A) RNAs, the knockdown experiment clearly supports the notion that only a select group of mRNAs are impacted by ZC3H14. The question then becomes: **What mRNAs are regulated by ZC3H14/Nab2?** The answer to this question for fly Nab2 is addressed in Chapter 2 of this dissertation. The sequence, structure, and emerging role for RNA modifications such methylation, specifically m^6A RNA methylation, have been implicated in neuronal functions.

1.12 Roles & Requirement of m^{6}A RNA Methylation In Neurons

Like DNA, RNA is also subject to different modifications that affect gene expression. To date, there are over 150 different types of RNA modifications known to regulate RNA fate (84, 85). One of the best characterized and most abundant mRNA modification is N6-methyladenosine (m^{6}A), which is a chemical modification that adds a methyl group to the 6th position of an adenosine (4, 84). Although the m^{6}A RNA modification was identified in the early 1970s, little was known about the biological role of this chemical modification. The development and availability of an anti-m^{6}A antibody that recognize and bind m^{6}A molecules made it possible to detect and quantify m^{6}A modified mRNAs (86, 87). Using the m^{6}A antibody and high-throughput RNA sequencing to map the m^{6}A transcriptome landscape, it became clear that m^{6}A RNA methylation is an abundant and reversible modification on many mRNAs. RNA stability, splicing, export, and protein translation can all be regulated by m^{6}A methylation (86-94).

1.13 The m^{6}A RNA Methylation Machinery Is A Dynamic Complex

The dynamic m^{6}A RNA modification is catalyzed by the core methyltransferases (“writers”) METTL3 (Methyltransferase-like protein 3) and METTL14; and the m^{6}A binding-proteins (“readers”) such as the YTH domain family proteins (TYHDFs) (86, 87, 91, 92, 94, 95). Demethylases (“erasers”) are extremely controversial, however, m^{6}A modified RNAs are likely turned over/degraded eventually. A series of the m^{6}A “writer” enzymes, and “reader” RBPs are responsible for RNA processing fate have been identified and their binding sites mapped across different models including mice and flies (86, 89, 96-98). The biological relevance of m^{6}A modification in neuronal cells during development of the CNS and association with human diseases in which this methylation machinery is disrupted are areas of very active research.

m^6A RNA modification in the nervous system is thought to be an important modification as m^6A is highly enriched in the brain (84, 86, 95). Complete deletion of *METTL3* is embryonic lethal, however, depletion of *METTL3* revealed a tissue-specific requirement for METTL3 in development of the cerebellum and hippocampal neurons, as well as a requirement for regulating alternative splicing of pre-mRNAs (99, 100). Thus, m^6A RNA modification appears to be essential for nervous system development and function (84, 86, 101).

This methylation machinery has been implicated in every aspect of RNA metabolism from transcription initiation to transcription termination and ultimately degradation. Interestingly, in mammalian cells, m^6A (METLL3) methylation at the 5'-end of transcripts has been shown to promote cap-independent translation, by somehow stabilizing the transcript long enough for translation without being degraded (102). This discovery alone puts m^6A RNA modification on the forefront as an essential regulator of RNA biology. **Figure 1—7** illustrates the regulatory roles of m^6A in RNA processing and RNA fate in both the nucleus and cytoplasm (85).

In fruit flies, the m^6A 'writers' (Mettl3 & Mettl14) and readers' (YTHDF) have been implicated in alternative splicing, sex determination, and nervous system development and function (88-90, 98, 103). Consistent with the controversy regarding "erasers" in mammals, there are no known "erasers" in *Drosophila*. Recent work has linked m^6A RNA modification to learning and memory in *Drosophila*, consistent with evidence from mice (98, 104). Depletion of Mettl3 protein in a tissue-specific manner from the MBs in the fly brain caused mutants impaired associative odor avoidance learning (Kan et al., 2020 bioRxiv preprint). Chapter 2 of this dissertation presents molecular and genetic data linking *Nab2*, *Mettl3*, and *Nab2*/*Mettl3* double mutants providing an exciting possible connection to m^6A RNA methylation. Together with the other roles of m^6A, this new link of m^6A modification to MB defects affecting learning and memory, and splicing all supports a model where RBPs, like Nab2 could be working directly or

indirectly with the m^6A machinery to regulate gene expression, by controlling post-transcriptional processing and mRNA translation in specific brain neurons (103).

1.14 Summary & Scope

Regulation of gene expression at the mRNA level by RBPs in a tissue-specific manner is an area of active investigation. Although RBPs play key roles in gene expression and are often ubiquitously expressed, dysregulation of genes encoding RBPs have been linked to numerous tissue-specific diseases. A new class of PABPs which utilize zinc fingers to bind target RNAs, is linked to intellectual disability and is conserved from yeast to humans. The yeast, fruit fly, and mouse model systems have been exploited to study the functions of this class of Nab2/ZC3H14 protein. Using these model systems, a body of knowledge in the regulation of mRNA has accumulated over the years. The yeast ZC3H14 orthologue of *Drosophila* Nab2 is essential for viability and functions in regulating the poly(A) tail lengths of RNAs (40-42). These systems established the importance of these PABPs in conserved biological processes that control mRNA fate (105).

Both mouse and fruit fly model systems have been employed in an effort to investigate the functions of the ZC3H14 protein in relevant cell types, to understand the mechanism through which loss of ZC3H14 results in intellectual disability. In both these systems, ZC3H14/Nab2 is necessary for proper brain development and function. In the mice, ZC3H14 is not essential for viability, but ZC3H14 is required for working-memory and for proper development of the hippocampus (66, 82). Before the mouse model was developed, the fly was extensively used to study ZC3H14/Nab2 function in the brain, as this model system has been thoroughly exploited in other neurological disorders such as Fragile X Syndrome (63). Most importantly, loss of fly Nab2 results in short-term memory defects, presumably caused by aberrant mushroom body (MB) formation during development (73). The neurons that make up the MB structures are critical for learning and memory function for the flies; loss of Nab2 in these neurons results in abnormal

courtship behaviors as well as other behavioral defects. These findings prompted the following question: Why are neurons susceptible to loss of this particular RBP? Using the fly model, the experiments presented in this dissertation provide insight into this question by testing the hypothesis that *Drosophila* Nab2, post-transcriptionally regulates gene expression of a subset of brain RNAs, by modulating the levels of m^6A methylation by virtue of a direct or indirect interaction, with the m^6A methylation machinery (see **Chapter 2**).

To define the RNA targets of *Drosophila* Nab2 and the *Drosophila* RNA exosome, a high-throughput RNA sequencing experiment was utilized to analyze the RNAs affected in *Nab2* mutant (**Chapter 2**) and *Drosophila* RNA exosome mutants (**Chapter 3**) fly heads. From the RNA sequencing experiments, a small percentage of brain RNAs were identified as Nab2 and RNA exosome regulated targets. These discoveries established a novel role for the RNA-binding protein Nab2 (**Chapter 2**) and the *Drosophila* RNA exosome subunit Rrp40 in regulating a small set of brain RNAs (**Chapter 3**). **Specifically, loss of *Drosophila* Nab2 in the nervous system altered the splicing pattern of mRNAs implicated in different biological processes**. The functional roles of the post-transcriptional RNA modification of m^6A in fly brain development and function were largely unknown, until the development of the m^6A antibody and the m^6A methyltransferase (*Mettl3*) fly mutants (86, 97). Experiments presented in this dissertation present the first link between Nab2, a polyadenosine RNA binding protein, and Mettl3, an enzyme that modifies A residues in mRNA.

1.15 Figures

Figure 1—1

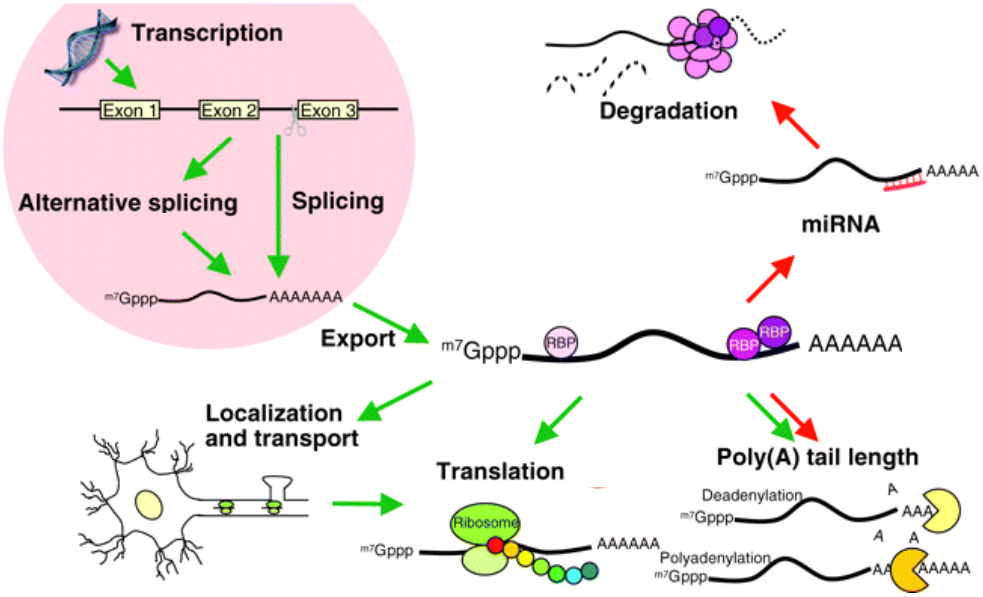


Figure 1—1: Post-transcriptional regulation of gene expression at the RNA level.

Post-transcriptional processing of pre-mRNAs involves splicing, alternative splicing, 5'-capping (5'-m^7G cap), and 3'-end cleavage/polyadenylation in the nucleus. Exported mRNA transcripts are then transported, translated, silenced, or degraded in the cytoplasm. The transcripts also undergo further regulation of their poly(A) tail length, which affects transcript stability and translation initiation. **Green** arrows indicate pathways leading towards translation, and **red** arrows indicate mRNA degradation pathways. RBP, RNA-binding protein, and miRNA for micro RNA. Figure is adapted from Kojima *et al.*, (2011).

Figure 1—2

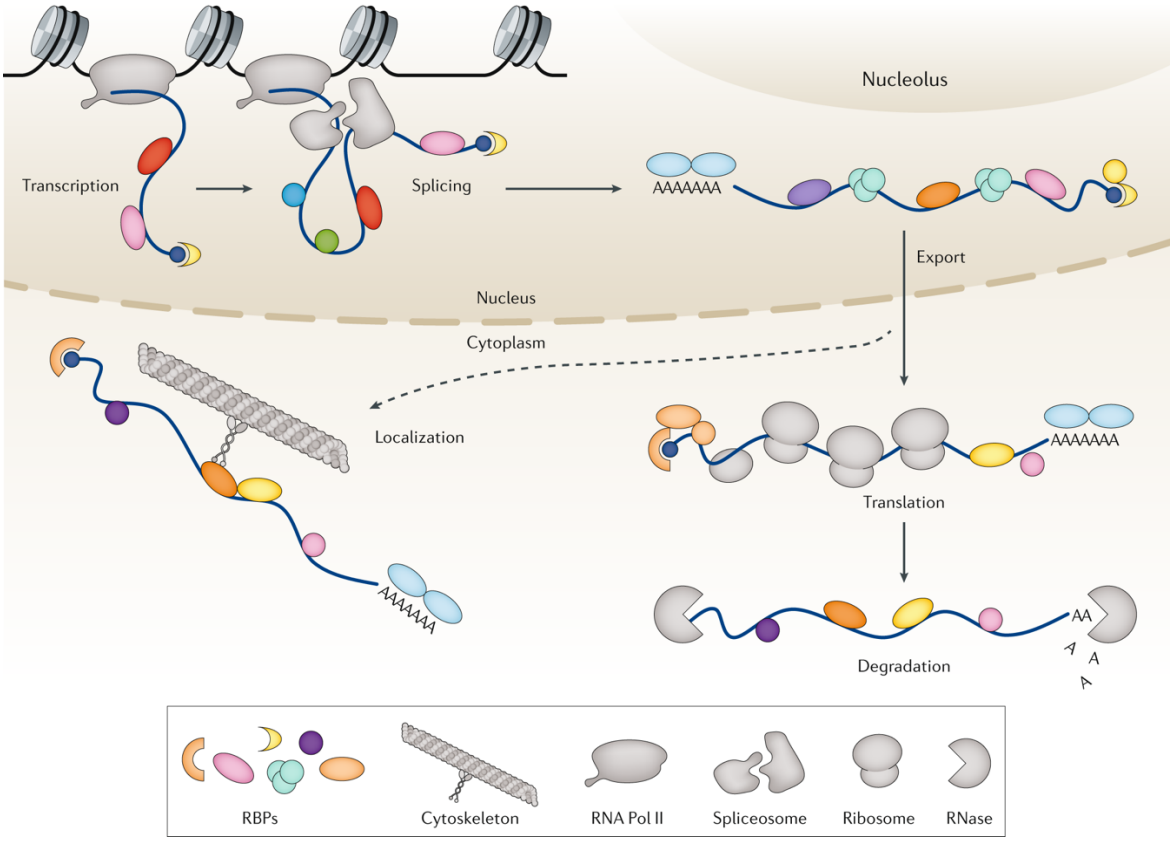

Nucleolus
Transcription
Splicing
AAAAAAA
Export
Nucleus
Cytoplasm
Localization
AAAAAAA
Translation
AA
A
A
A
Degradation
RBPs
Cytoskeleton
RNA Pol II
Spliceosome
Ribosome
RNase

Figure 1—2: RNA-binding proteins and their functions in RNA biology.

Nascent RNA transcripts emerging from RNA polymerase II in the nucleus are immediately bound by RNA binding proteins to control RNA life. Nuclear RBPs control transcription, capping, polyadenylation, and splicing by the spliceosome. Cytoplasmic RBPs are involved in translation and degradation by the ribosome, and RNase respectively. RBPs, RNA-Binding Proteins, and RNA Pol II, RNA polymerase II. This Figure is from *Gebauer et al*., (2020).

Figure 1—3

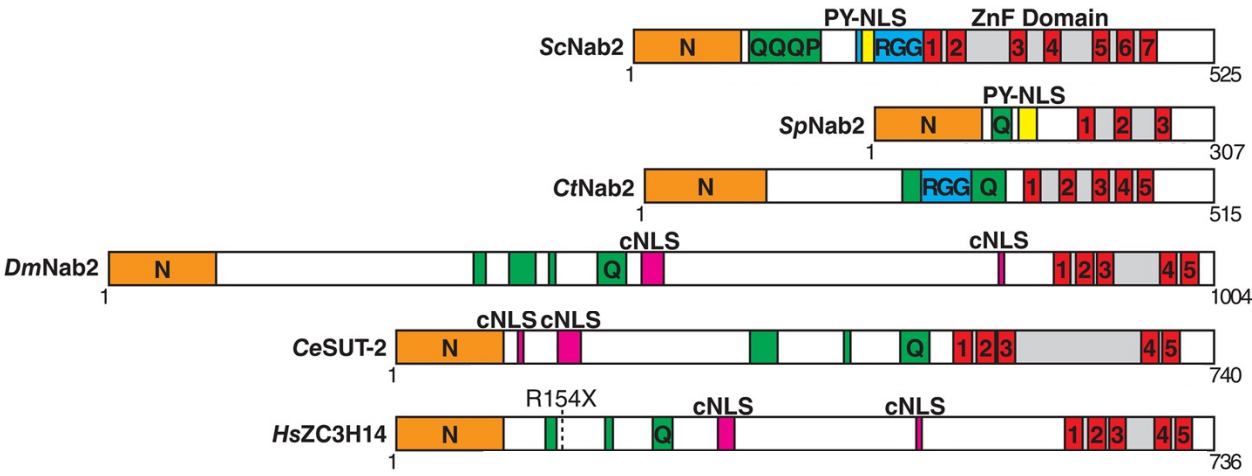

PY-NLS
ZnF Domain
ScNab2
N
QQQP
RGG
1
2
3
4
5
6
7
1
525
PY-NLS
SpNab2
N
Q
1
2
3
1
307
CtNab2
N
RGG
Q
1
2
3
4
5
1
515
cNLS
cNLS
DmNab2
N
Q
1
2
3
4
5
1
1004
cNLS cNLS
CeSUT-2
N
Q
1
2
3
4
5
1
740
R154X
cNLS
cNLS
HsZC3H14
N
Q
1
2
3
4
5
1
736

Figure 1—3: Domain structure of the ZC3H14/Nab2 protein family.

Members of the Nab2/ZC3H14 zinc finger polyadenosine RNA-binding protein family share conserved domains. The *S. cerevisiae* (*Sc*Nab2), *S. pombe* Nab2 (*Sp*Nab2), *C. thermotolerans* (*Ct*Nab2), *Drosophila melanogaster* (*Dm*Nab2), *Caenorhabditis elegans* SUT-2 (*Ce*SUT-2), and *Homo sapiens* (*Hs*ZC3H14) all share common domain structures. An N-terminal PWI-fold domain (**orange**) serves as a protein-protein interaction domain in *Sc*Nab2, followed by Q-rich region (**green**), and a C-terminal zinc finger (ZnF) domain (**gray**) containing a series of ZnFs (**red**), which bind to polyadenosine RNA. *Sc*Nab2 and *Ct*Nab2 contain an RGG domain (**blue**), which functions in nuclear import in *Sc*Nab2. In addition, *Sc*Nab2 contains a Pro-Tyr nuclear localization signal (PY-NLS) in **yellow** and *Sp*Nab2 contains a predicted PY-NLS that functions in nuclear import in *Sc*Nab2. Nab2 orthologues from higher Eukaryotes contain two predicted classical nuclear localization signals (cNLS) in **magenta** that function in nuclear import. The nonsense mutation R154X in ZC3H14 patients with autosomal recessive intellectual disability is highlighted. Figure adopted and modified from *Fasken et al*., (2019).

Figure 1—4

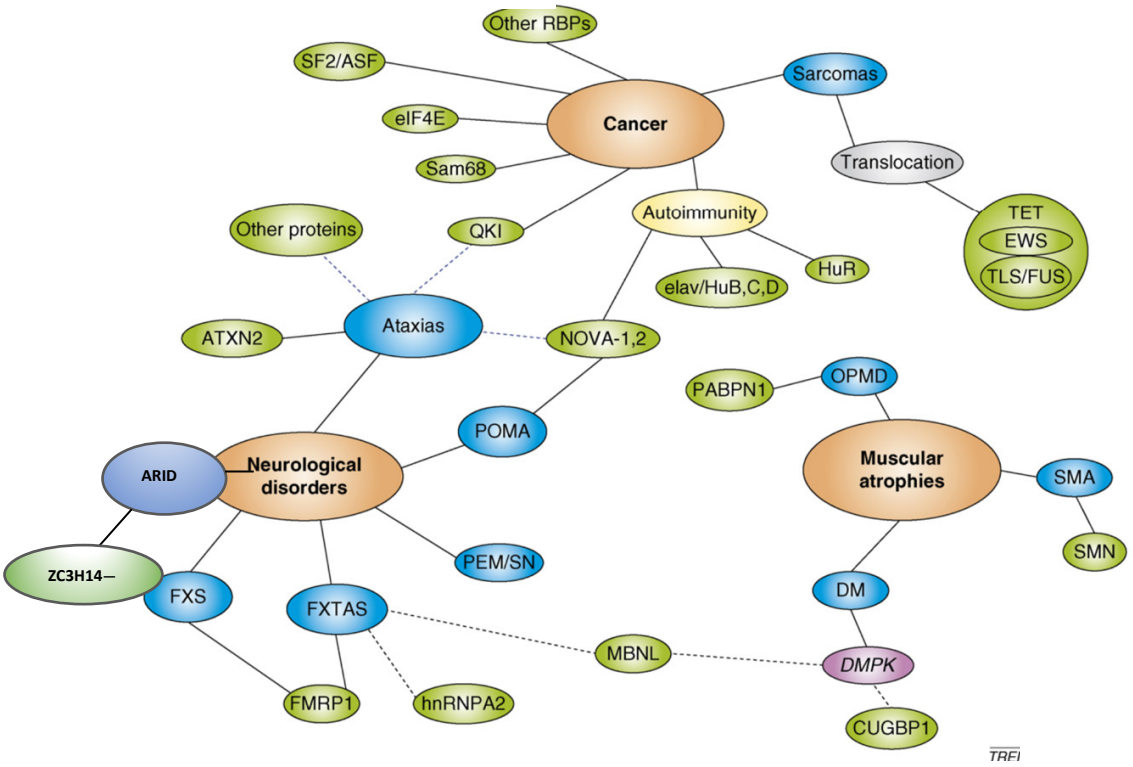

Other RBPs
SF2/ASF
Sarcomas
eIF4E
Cancer
Sam68
Translocation
Autoimmunity
TET
EWS
TLS/FUS
Other proteins
QKI
HuR
elav/HuB,C,D
Ataxias
ATXN2
NOVA-1,2
PABPN1
OPMD
POMA
ARID
Neurological disorders
Muscular atrophies
SMA
SMN
ZC3H14
FXS
FXTAS
PEM/SN
DM
MBNL
DMPK
FMRP1
hnRNPA2
CUGBP1
TRFI

Figure 1—4: RNA-binding proteins (RBPs) including ZC3H14 and human diseases.

A network of RNA-binding proteins (green) with aberrant expression or functions have been identified in several major classes of human diseases (orange) sometimes directly (— solid lines) or indirectly (---dashed lines) associated with a specific disease (blue) or gene (pink). Neurological disorders directly associated with Ataxias, ARID, and FXS, are blue with solid lines. The anti-neuronal antibodies in brain tumors are in yellow and chromosomal translocations in grey. ARID, autosomal recessive intellectual disability. Figure is adopted and modified from *Lukong et al*., (2008).

Figure 1—5

A. Family-1 **B. Family-2** **C.** Schematic of human *ZC3H14* with mutations

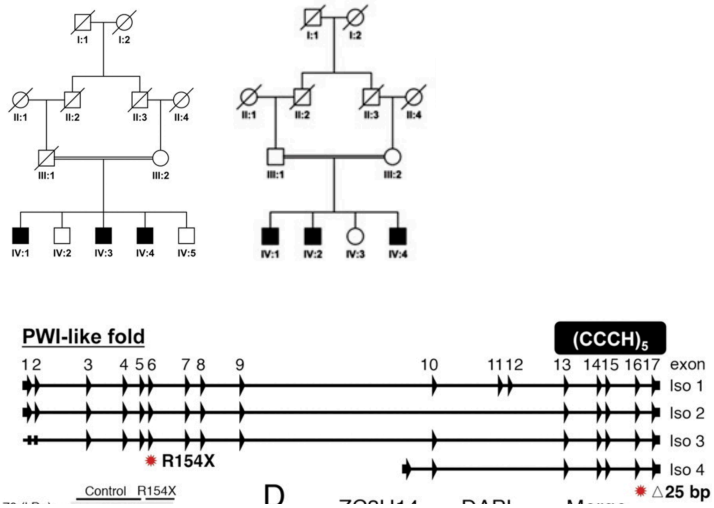


Figure 1—5: Pedigrees of families with ZC3H14 mutations linked to disease.

Pedigree of Family-1 (**A**) and Family-2 (**B**) with mutations in *ZC3H14* gene. Schematic of four *ZC3H14* splice variants indicating exons encoding the N-terminal PWI-like domain and C-terminal CysCysCysHis zinc-finger (ZnF) RNA binding motif $(CCCH)_5$ domain (**C**). Positions of patients with R154X nonsense mutation and 25bp deletion are indicated by red stars. Figure adopted and modified from *Pak et al*., (2011).

Figure 1—6

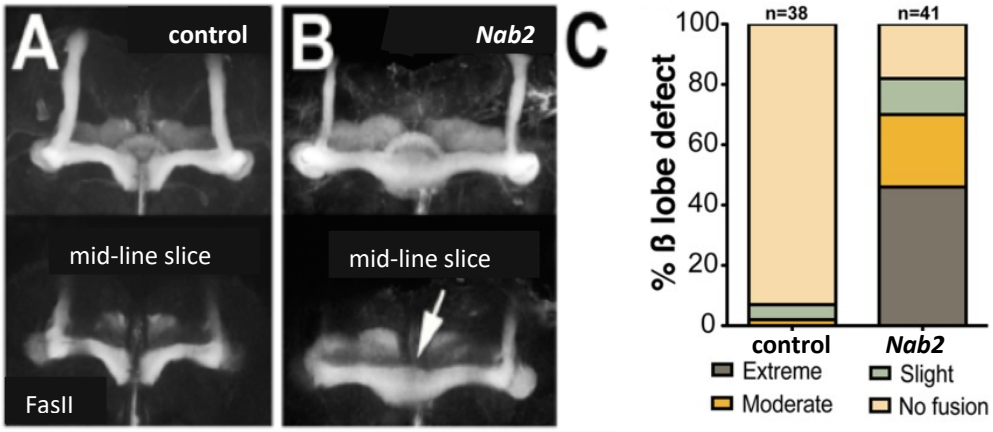

A
control
mid-line slice
FasII
B
Nab2
mid-line slice
C
% ß lobe defect
100
80
60
40
20
0
n=38
n=41
control
Nab2
Extreme
Moderate
Slight
No fusion

Figure 1—6: *Drosophila* Nab2 is required for development of mushroom body neurons.

Fasciclin II (FasII) staining of mushroom body axons depicting medially projecting beta (β) lobes and vertical alpha (α) lobes in control (**A**) and *Nab2* mutant (**B**) brains. Maximum intensity Z-stack projections of single sections (top) and mid-line slices (bottom) are shown. β-lobes of control brain rarely cross the mid-line and have well-formed α-lobes, while *Nab2* mutant brain appears to have fused β-lobes (white arrow) and or missing α-lobes. Quantification of the frequency of β-lobe defects as extreme, moderate, slight or no fusion (**C**). Figure was adopted and modified from *Kelly et al*., (2016).

Figure 1—7

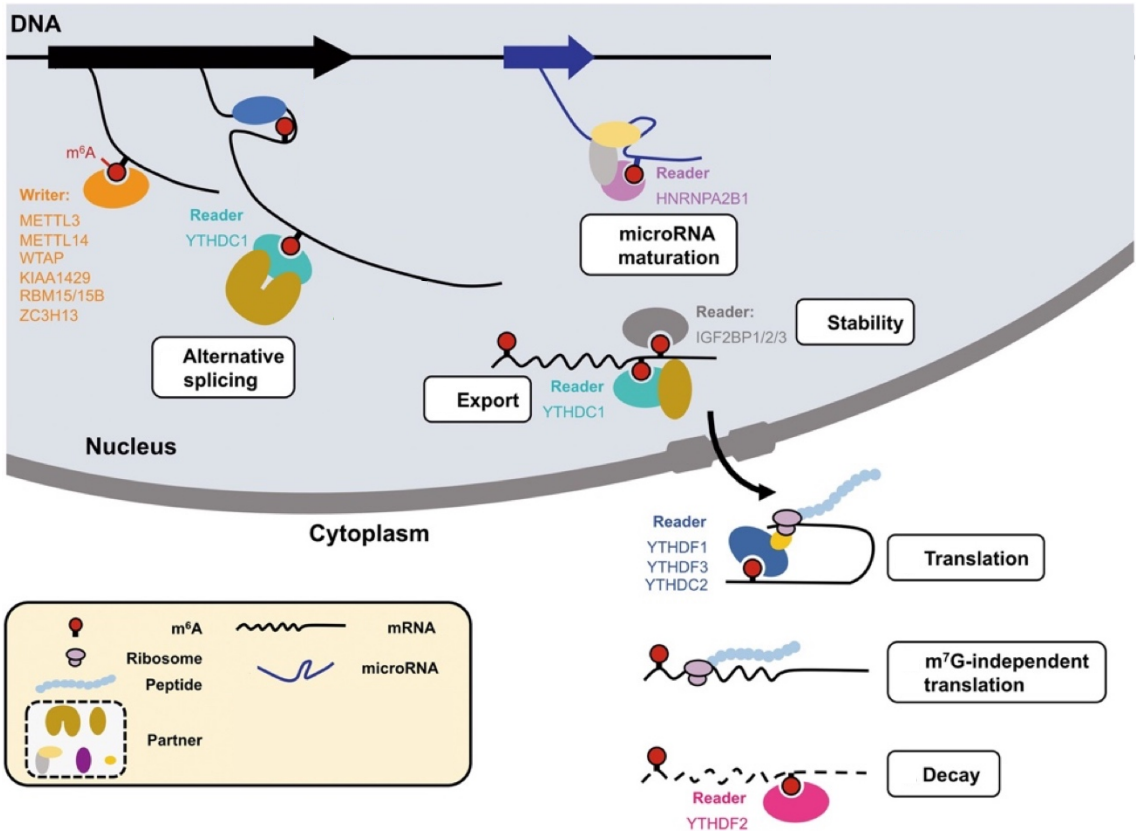

DNA
m6A
Writer:
METTL3
METTL14
WTAP
KIAA1429
RBM15/15B
ZC3H13
Reader
YTHDC1
Alternative splicing
Reader
HNRNPA2B1
microRNA maturation
Reader:
IGF2BP1/2/3
Stability
Export
Reader
YTHDC1
Nucleus
Cytoplasm
Reader
YTHDF1
YTHDF3
YTHDC2
Translation
m7G-independent translation
Reader
YTHDF2
Decay
m6A
mRNA
Ribosome
Peptide
microRNA
Partner

Figure 1—7: Dynamic role of m^6A in RNA processing paths.

There are diverse molecular functions of m^6A RNA methylation. In eukaryotic cells, m^6A RNA levels are dynamically regulated by “writers” and recognized by “readers” directly or indirectly. The diversity of cellular RNA processing involves various m^6A “readers”. Nuclear m^6A modulates mRNA alternative splicing, mRNA export, micro-RNA maturation, and mRNA stability. Cytoplasmic m^6A enhances processes involved in mRNA translation, including 5’-m^7G-cap-independent translation and mRNA decay. mRNA, messenger RNA. Figure adopted and modified from *Yang et al*., (2018).

Chapter 2

The Nab2 RNA binding protein promotes sex-specific splicing of *Sex lethal* in *Drosophila* neuronal tissue

***Please Note*:**

J. Christopher Rounds, a listed co-author of the manuscript, analyzed the RNA Sequencing data I generated in the current chapter (Figures 1-2, Figure 3A and 3G, Table 1, Supplemental Figures 1-4, and Supplemental Table 1-3). The genetic and behavioral experiments were done by the undergraduate student, Brianna Brown under the direct supervision of B. Jalloh. The m^6A RNA pull downs were performed by both Brianna Brown and B. Jalloh.

This chapter has been submitted for review as:

Binta Jalloh, J. Christopher Rounds, Brianna E. Brown, Isaac J. Kremsky, Ayan Banerjee, Derrick J. Morton, Rick S. Bienkowski, Milo B. Fasken, Anita H. Corbett, and Kenneth H. Moberg, "The Nab2 RNA binding protein promotes sex-specific splicing of *Sex lethal* in *Drosophila* neuronal tissue" (2020). This chapter is also posted online in biorxiv.

https://www.biorxiv.org/content/10.1101/2020.11.13.382168v3

2.1 Abstract

The *Drosophila* polyadenosine RNA binding protein Nab2, which is orthologous to a human protein lost in a form of inherited intellectual disability, controls axon projection, locomotion, and memory. Here we define an unexpectedly specific role for Nab2 in regulating splicing of ~150 exons/introns in the head transcriptome and link the most prominent of these, female retention of a male-specific exon in the sex determination factor *Sex-lethal* (*Sxl*), to a role in m^6A-dependent mRNA splicing. Genetic evidence indicates that aberrant *Sxl* splicing underlies multiple phenotypes in *Nab2* mutant females. At a molecular level, Nab2 associates with *Sxl* pre-mRNA and ensures proper female-specific splicing by preventing m^6A hypermethylation by Mettl3 methyltransferase. Consistent with these results, reducing Mettl3 expression rescues developmental, behavioral and neuroanatomical phenotypes in *Nab2* mutants. Overall, these data identify Nab2 as a required regulator of m^6A-regulated *Sxl* splicing and imply a broader link between Nab2 and Mettl3-regulated brain RNAs.

2.2 Introduction

RNA binding proteins (RBPs) play important roles in guiding spatiotemporal patterns of gene expression that distinguish different cell types and tissues within organisms. There are an estimated ~1500 RBPs that distribute between the nucleus and cytoplasm (106), and each has the potential to interact with RNAs to modulate post-transcriptional gene expression. Such regulation is particularly critical in highly specialized cells such as neurons (107) where highly regulated alternative splicing of coding regions and 3'UTRs, cleavage/polyadenylation, trafficking and local translation are contribute to precise regulation of gene expression (108). The critical roles of RBPs in neurons is highlighted by many functional studies that reveal the importance of this class of proteins in brain development and function (109) and by the prevalence of human neurological diseases linked to mutations in genes encoding RBPs (108). Many of these RBPs are ubiquitously expressed and play multiple roles in post-transcriptional regulation. Thus, defining the key neuronal functions of these proteins is critical to understanding both their fundamental roles and the links to disease.

Among the RBPs linked to human diseases are a group of proteins that bind with high affinity to polyadenosine RNAs, which are termed poly(A) RNA binding proteins or PABs (56). Functional studies of classical nuclear and cytoplasmic PABs, which utilize RNA recognition motifs (RRMs) to recognize RNA, have uncovered diverse roles for these proteins in modulating mRNA stability, alternative cleavage and polyadenylation and translation (105). A second, less well-studied, group of PABs uses zinc-finger (ZnF) domains to bind target RNAs. Among these is the Zinc Finger Cys-Cys-Cys-His-Type Containing 14 (ZC3H14) protein, which binds with high affinity to poly(A) RNAs via a set of C-terminal tandem Cys-Cys-Cys-His type zinc-finger domains (110). ZC3H14 is broadly expressed in many tissues and cell types but mutations in the human *ZC3H14* gene are associated with a heritable form of intellectual disability (49), implying an important

requirement for this protein in the central nervous system. ZC3H14 has well-conserved homologs in eukaryotes, including *S. cerevisiae* Nuclear poly(A)-binding protein 2 (Nab2), *Drosophila melanogaster* Nab2, *C. elegans* SUT-2 and murine ZC3H14 (43). Zygotic loss of *ZC3H14* in mice and *Drosophila* impairs neuronal function (49, 111), while neuron-specific depletion of *Drosophila* Nab2 is sufficient to replicate these effects (49). Reciprocally, expression of human ZC3H14 in Nab2-deficient neurons rescues this defect, demonstrating a high degree of functional conservation between human ZC3H14 and *Drosophila* Nab2 (44). Collectively, these data focus attention on what are critical, but poorly understood, molecular roles for ZC3H14/Nab2 proteins in neurons.

Neuronal ZC3H14/Nab2 can be divided into two pools, a nuclear pool that accounts for the majority of ZC3H14/Nab2 in the cell, and a small cytoplasmic pool of protein detected in mRNA ribonucleoprotein particles (mRNPs) of axons and dendrites (66, 110, 111). Depletion of both pools in *Drosophila* neurons cause defects in axon genesis within the brain mushroom bodies (73), a pair of twin neuropil structures involved in learning and memory (77, 78). This requirement has been linked to a physical association between cytoplasmic Nab2 and the *Drosophila* Fragile-X mental retardation protein homolog, Fmr1 (112), and translational repression of shared Nab2-Fmr1 target RNAs in the cytoplasm (66). Despite this insight into a cytoplasmic function of Nab2, molecular roles of the abundant population of Nab2 in neuronal nuclei remain undefined.

Here, we employed a broad and an unbiased RNA sequencing approach to identify transcriptome-wide changes in the heads of *Nab2* loss-of-function mutant flies. While the steady-state levels of most transcripts were not significantly changed, we uncovered a striking effect on splicing of a subset of neuronal RNA transcripts. We focused our analysis on a well-characterized sex-specific alternative splicing event in the *Sex-lethal* (*Sxl*) transcript. Results reveal that *Nab2* plays a novel role in regulating the alternative splicing of *Sxl* in a sex-specific manner. Recent

works has revealed a role for m^6A RNA methylation by the enzyme Mettl3 in modulating this splicing event (98, 113). Similar to *Mettl3*, the requirement for *Nab2* in alternative splicing of *Sxl* is only essential for neuronally-enriched tissues. Genetic and biochemical experiments support a functional link between m^6A methylation and Nab2 function. These results demonstrate the role for *Drosophila Nab2* in RNA alternative splicing as well as RNA methylation and sex determination in the neurons.

2.3 Results

2.3.1 Nab2 loss affects levels and processing of a subset of RNAs in the head transcriptome.

To assess the role of Nab2 in regulating the central nervous system transcriptome, a high-throughput RNA Sequencing (RNA-Seq) analysis was carried out in triplicate on *Nab2* null mutant heads ($Nab2^{ex3}$ imprecise excision of *EP3716*) (49) and isogenic control heads ($Nab2^{pex41}$ precise excision of *EP3716*). To capture sex-specific differences, heads were collected from both male and female flies of each genotype. Briefly, total RNA from 1-day old adults was rRNA-depleted and used to generate stranded cDNA libraries that were sequenced (150 cycles) on a NextSeq 500 High Output Flow Cell. This generated a total of approximately 1.1 billion 75 base-pair (bp) paired-end reads (91 million/sample) that were mapped onto the Dmel6.17 release of the *Drosophila* genome using RNA STAR (114). Read annotation and per-gene tabulation was conducted with featureCounts (115) and differential expression analysis was performed with DESeq2 (116).

RNA sequencing reads across the *Nab2* gene are almost completely eliminated in $Nab2^{ex3}$ mutants, confirming the genetic background and integrity of the analysis pipeline (**Supplemental Figure 1**). Principal component analysis (PCA) performed with DESeq2 output data confirms that the 12 RNA-seq datasets distribute into four clusters that diverge significantly from one another based on genotype ($Nab2^{ex3}$ vs. $Nab2^{pex41}$ control; PC1 58% variance) and sex (male vs. female; PC2 26% variance) (**Figure 1A**). The DESeq2 analysis detects 3,799 and 1,545 genes in females and males, respectively, that exhibit statistically significant differences in RNA abundance between $Nab2^{ex3}$ and control (Benjamini-Hochberg [BH] adjusted *p*-value/false discovery rate (FDR)<0.05). Comparison of fold-changes ($Nab2^{ex3}$ vs. control) among these significantly different RNAs reveals a high degree of correlation in female vs. male samples (R=0.79), particularly among RNAs whose levels are most elevated upon Nab2 loss (**Figure 1B**). Applying a 2-fold change cutoff ($|\log_2$[fold-

change]|≥1) trims these sets to 453 significantly changed RNAs in females (294 'up', 159 'down'), and 305 significantly changed RNAs in males (150 'up', 155 'down') (**Figure 1C**), which merge into a combined set of 570 significantly affected RNAs that trend similarly in heatmap analysis of mutant vs. control samples (**Figure 1D**). A majority of the 453 affected 'female' RNAs are mRNAs (439) and the remaining are snoRNAs (8), snRNAs (1), pre-rRNAs (1), and tRNAs (4) (**Figure 1E**). A similar distribution occurs in male heads: a majority of the affected RNAs are mRNAs (297) and the remainder are snoRNAs (4), snRNAs (1), pre-rRNAs (1), and tRNAs (2) (**Figure 1E**). Overall, the number of significantly changed RNAs ((|$\log_2$[fold-change]|≥1 and FDR<0.05) in *Nab2^{ex3}* females and male heads represents a small fraction of RNAs detected in heads (2.2% and 3.7% in males and females, respectively), suggesting that Nab2 normally contributes to RNA-specific regulatory mechanisms in *Drosophila* head tissue.

2.3.2 Nab2 loss alters levels of transcripts linked to mRNA processing.

To screen Nab2-regulated RNAs for enriched functions, Gene Set Enrichment Analysis (GSEA) (117, 118) was carried to identify enriched gene ontology (GO) terms (119, 120) among the significantly changed female and male RNAs identified by DESeq2. This filtering uncovers significant enrichment (p<0.05) for "RNA splicing" GO (GO:0008380) within the upregulated group of RNAs in both sexes (**Figure 2A**). In *Nab2^{ex3}* females, 32 of 155 genes annotated under this term are present among upregulated RNAs; whereas in males, 75 of 159 genes annotated under this term are present among upregulated RNAs (**Figure 2A**). This enrichment for upregulated splicing-related factors indicates that Nab2 loss could shift splicing patterns in the adult head. Consistent with this hypothesis, MISO (mixture of isoforms) analysis (121) of annotated alternative splicing events confirms that Nab2 loss significantly alters splicing patterns within a small number of transcripts in female (50) and male (51) heads (**Supplemental Table 1**) that fall into a variety of

GO terms (**Supplemental Figure 2**). These MISO-called alternative splicing events include 5' and 3' alternative-splice site usage, intron retention events, and previously annotated exon skipping events, some of which are detected in the same transcripts (**Figure 2B**).

To test whether Nab2 loss results in unannotated or aberrant splicing events, DEXSeq analysis (122) was performed to scan for differential abundance of individual exons relative to other exons within the same transcript. This analysis detects 151 affected RNAs in *Nab2*^{ex3} females and 114 in *Nab2*^{ex3} males (**Table 1**), with many top-ranked transcripts encoding factors with roles in behavior, neurodevelopment, and/or neural function (**Supplemental Table 2**).

The most statistically significant exon usage change in either sex is female-specific inclusion of exon 3 in the *Sex lethal* (*Sxl*) mRNA (2.86-fold increase, p=3.08x10^{-235}). This effect on *Sxl* mRNA in *Nab2*^{ex3} females is followed in rank order of significance by enhanced inclusion of exons 1 and 2 of the MIF4GD homolog transcript *CG13124*, exons 1 and 2 of the voltage-gated ion channel transcript I_h *channel* (I_h), and exon 1 of the synaptic enzyme transcript *Acetylcholine esterase* (*Ace)*. In *Nab2*^{ex3} males, the top four events are enhanced inclusion of exon 1 of the *Ace* transcript, exon 1 of the *Protein kinase C at 53E* (*Pkc53E*) transcript, exons 1 and 2 of the Rab GTPase *pollux* (*plx*) transcript, and exons 1 and 2 of *Protein kinase N* (*Pkn*) transcript. In some cases, identical exons are affected in both *Nab2*^{ex3} sexes and accompanied by retention of the intervening intron (e.g., see *CG13124* and I_h traces in **Supplemental Figure 2**). The robust increase in *Sxl* exon 3 in *Nab2*^{ex3} females is noteworthy both for the central role that differential inclusion of exon 3 plays in *Drosophila* sex determination (123), but also because DEXSeq did not detect changes in exon 3 inclusion or abundance in *Nab2*^{ex3} males. In light of this sex-specific effect of Nab2 loss on alternative splicing of *Sxl* exon 3, subsequent analyses focused on the role of Nab2 in *Sxl* mRNA splicing in female heads.

2.3.3 *Nab2*^{ex3} females exhibit masculinized *Sxl* splicing in neuron-enriched tissues.

The Sex lethal (Sxl) protein is a female-specific, U-rich RNA binding protein that acts through the *tra-dsx* and *msl-2* pathways to promote female somatic and germline identity (124, 125). *Sxl* pre-mRNA is expressed in both males and females, but alternative splicing regulated by m^{6}A RNA methylation and several RBPs leads to female-specific skipping of exon 3 during splicing (113, 126, 127). Because exon 3 includes an in-frame translation 'stop' codon, full-length Sxl protein is only made and active in female cells (128). The inclusion of *Sxl* exon 3 in *Nab2*^{ex3} mutants would thus implicate Nab2 as a novel component of molecular machinery that controls *Sxl* pre-mRNA splicing in female heads.

Visualizing *Sxl* RNA-Seq reads with IGV Viewer (129) confirms a large increase in exon 3 reads in *Nab2*^{ex3} females (*Nab2*^{ex3}-F) relative to control females (control-F), and also shows retention of ~500 bases of intron 3 sequence in *Nab2*^{ex3} females (**Figure 3A**). Quantification of reads across the entire *Sxl* locus detects an ~1.5-fold increase in the overall abundance of the *Sxl* mRNA in *Nab2*^{ex3} females compared to control females. Normal splicing patterns are detected across all other *Sxl* intron-exon junctions in both genotypes of males and females, including female-specific exon 9 inclusion (**Figure 3A**). Reverse transcription polymerase chain reaction (RT-PCR) on fly heads using *Sxl* primers (see arrows in **Figure 3A** schematic) that detect exon 2-exon 4 (control females) and exon 2-exon 3-exon 4 (control males) confirms the presence of the mis-spliced exon 2-exon 3-exon 4 mRNA transcript in *Nab2*^{ex3} females (**Figure 3B**). The exon 2-exon 3-exon 4 mRNA transcript appears to be more abundant in *Nab2*^{ex3} female heads than in female heads lacking *Mettl3*, which encodes the catalytic component of an m^{6}A methyltransferase complex that promotes exon 3 skipping in nervous system tissue (98, 113, 127). RT-PCR also reveals a ~1kb band in *Nab2*^{ex3} females (arrowhead, **Figure 3B**) that sequencing identifies as aberrantly spliced transcript that incorporates 503 bases of intron 3 leading up to a cryptic 5'

splice site (i.e., exon 2-exon 3-intron 3^{503}-exon 4); this matches the *Sxl* intron 3 sequencing reads observed in IGV (see **Figure 3A**). Significantly, RT-PCR analysis of *Sxl* mRNA in dissected control and *Nab2^{ex3}* females detects exon 3 retention in *Nab2^{ex3}* thoraxes, but not in abdomens or ovaries (**Figure 3C**). This result implies that Nab2 is only necessary to direct *Sxl* exon 3 exclusion in specific tissues or cell types such as neurons, which are enriched in the head (brain) and thorax (ventral nerve cord). In sum, these data reveal a tissue-specific role for Nab2 in blocking *Sxl* exon 3 inclusion in females and regulating 5'-splice site utilization at the exon 3-exon 4 junction.

Sxl exon 3 inclusion in *Nab2^{ex3}* female head RNAs indicates that insufficient levels of the exon 2-exon 4 splice variant contribute to *Nab2^{ex3}* phenotypes in females. To test this hypothesis, the constitutively female-spliced *Sxl^{M8}* allele (130) was placed as a heterozygote into the background of *Nab2^{ex3}* animals. *Sxl^{M8}* contains a 110 bp deletion covering the 5'-end of intron 2 and 3'-end of exon 3 and consequently undergoes constitutive splicing to the feminized exon 2-exon 4 variant regardless of sex (top panel, **Figure 3D**). Heterozygosity for this *Sxl^{M8}* allele produces strong rescue of *Nab2^{ex3}* mutant female viability from ~4% to 71% (*Sxl^{M8}/+;;Nab2^{ex3}*) (**Figure 3D**). Female *Nab2^{ex3}* siblings that did not inherit the *Sxl^{M8}* allele also exhibit elevated viability (64%), perhaps due to maternal loading of *Sxl* mRNA (**Figure 3D**). Surviving *Sxl^{M8}/+;;Nab2^{ex3}* females also show improved locomotion in a negative geotaxis assay (**Figure 3E**) and lengthened lifespan (**Figure 3F**) relative to *Nab2^{ex3}* females. Consistent with the original report describing *Sxl^{M8}* (130), the allele is male-lethal in control (*Nab2^{pex41}*) and *Nab2^{ex3}* backgrounds. This female-specific rescue of *Nab2^{ex3}* by *Sxl^{M8}* indicates that restoring Sxl expression can compensate for Nab2 loss in some developing tissues.

As Sxl is itself an RBP with roles in alternative splicing (124, 128), the rescuing effect of the *Sxl^{M8}* allele prompted a bioinformatic scan for RBP motifs enriched in proximity to the Nab2-

dependent alternative splicing events identified by MISO analysis (see **Figure 2B**). Input sequences were composed of retained introns plus 25bp extending into each flanking exon, and alternative splice sites with 25bp of exon plus 1kb of adjacent intron (see schematic, **Figure 3G**). This unbiased scan detected predicted Sxl binding sites as the single most abundant RBP motif within the Nab2-regulated MISO events in females (**Figure 3G**). Notably, Sxl motifs were not detected as enriched in the male *Nab2*^{ex3} MISO dataset, which otherwise strongly resembles the remaining group of female-enriched RBP motifs (e.g., the HNRNPL homolog *smooth* (*sm*), *RNA binding protein-9* (*Rbp9*), the U1-SNRNPA homolog *sans fille* (*snf*), and the U2-SNRNP component *U2AF50*). The female-specific enrichment for Sxl binding sites implies that Nab2 may regulate alternative splicing events indirectly via control of a *Sxl*-regulated splicing program. Intriguingly, the Sxl target *transformer* (*tra*) and the Tra target *double-sex* (*dsx*) (131, 132) were not recovered in the *Nab2*^{ex3} MISO or DESeq2 datasets, and IGV reads show no evidence of altered structure of their RNAs relative to *Nab2*^{pex41} controls (**Supplemental Figure 4**). Together these data suggest that Sxl may not control the *tra-dsx* pathway in the adult head, or that *tra* and *dsx* splicing are only altered in a subset of *Nab2*^{ex3} head cells and thus not detectable by bulk RNA-Seq analysis.

2.3.4 The dosage compensation complex contributes to phenotypes in *Nab2*^{ex3} mutant females.

The lack of evidence that *tra* and *dsx* mRNAs are affected by *Nab2* loss prompted analysis of the other major role of Sxl, which is to bind to the *male-specific lethal-2* (*msl-2*) mRNA and inhibit its translation in female somatic and germline tissues (133, 134). As a result, Msl-2 protein is only expressed in male cells, where it promotes assembly of a chromatin modifying complex termed the Dosage Compensation Complex (DCC; composed of Msl-1, Msl-2, Msl-3, Mof, Mle and *roX1* and *roX2* non-coding RNAs), which is recruited to the male X chromosome to equalize X-linked gene expression between males and females (133, 134). A number of DCC components are

expressed highly in the adult nervous system (135), which correlates with the tissue-restricted link between Nab2 and *Sxl* splicing (as in **Figure 3B**). As a functional test of interactions between Nab2 and the DCC pathway, a loss-of-function allele of *msl-2* (*msl-2*$^{killer\ of\ males\text{-}A}$ or *msl-2*^{kmA}) (136) was tested for dominant effects on *Nab2*^{ex3} female phenotypes. Remarkably, a single copy of *msl-2*^{kmA} significantly rescues defects in viability (**Figure 4A**), lifespan (**Figure 4B**), and locomotion (**Figure 4C**) among *Nab2*^{ex3} females. *roX1* and *mle* loss-of-function alleles were also able to rescue *Nab2*^{ex3} phenotypes (**Supplemental Figure 5**). Given that Msl-2 is not normally active in adult female tissues (135, 137) and that forced *msl*-2 expression reduces female viability (138), rescue by *msl-2*^{kmA} heterozygosity provides strong evidence that the DCC pathway is inappropriately activated in *Nab2*^{ex3} females. Of note, the *msl-2* and *mle* RNAs appear similar in IGV reads from control and *Nab2*^{ex3} adults (**Supplemental Figure 4**), indicating that genetic interactions between these loci are not likely due to direct effects of Nab2 loss on abundance and structure of these RNAs.

2.3.5 Nab2-regulated splicing of *Sxl* exon3 is dependent upon the Mettl3 m^{6}A methyltransferase.

Genetic interactions between *Nab2*, *Sxl*, and *msl-2* alleles are consistent with a role for Nab2 protein in regulating sex-specific splicing of *Sxl* exon 3. One mechanism that promotes exon 3 exclusion in females is based on N^6-methylation of adenosines (m^{6}A) in *Sxl* pre-mRNA by the Methyltransferase like-3 (Mettl3)-containing methyltransferase complex (reviewed in 88). Inactivating mutations in components of this m^{6}A 'writer' complex masculinize the pattern of exon 3 splicing in female flies (98, 113, 127) in a manner similar to *Nab2*^{ex3}. Molecular studies indicate that the Mettl3 complex promotes exon 3 exclusion in females by depositing m^{6}A within *Sxl* exon 3 and flanking introns (98, 113, 127).

To assess Nab2-Mettl3 functional interactions, the null allele *Mettl3^{null}* (formerly known as *Ime4^{null})* (Lence et al., 2016) was recombined into the *Nab2^{ex3}* background (**Supplemental Figure 6**; the loci are 281kb apart on chr3R), and then used to test for effects on phenotypes in *Nab2^{ex3},Mettl3^{null}* double mutant females. As described previously (98, 113, 127), loss of Mettl3 in an otherwise *wildtype* genetic background reduces adult viability, shortens lifespan and decreases locomotion in a negative geotaxis assay (**Figure 5A**-**C**). However, removing *Mettl3* has the inverse effect of suppressing each of these defects in *Nab2^{ex3}* females (**Figure 5A-C**): *Nab2^{ex3},Mettl3^{null}* double mutant females show approximately 5-fold higher viability, 1.5-fold longer lifespan, and 2-fold greater locomotion activity (at the 30sec time point; **Figure 5C**) than *Nab2^{ex3}* mutants alone. Significantly, qPCR analysis confirms that *Mettl3* loss causes aberrant *Sxl* exon 3-exon 4 splicing (as reported in 98, 113, 127) but reduces it in *Nab2^{ex3},Mettl3^{null}* double mutant females relative to *Nab2^{ex3}* females (**Figure 5D-E**). Within the adult brain, removing *Mettl3* also rescues structural defects in α-and ß-lobes in the mushroom bodies (MBs) that are otherwise highly penetrant in *Nab2^{ex3}* adults (**Figure 5F**). *Nab2^{ex3}* brains normally show 60-80% penetrance of thinned or missing α-lobes and fused ß-lobes (66, 73), but *Nab2^{ex3},Mettl3^{null}* double mutants exhibit a reduction of α-lobe defects and complete suppression of the ß-lobe fusion defect (**Figure 5F**). As Nab2 and Mettl3 each act autonomously within MB neurons to pattern α/ß-lobe structure (73, 90), these genetic interactions imply that Nab2 and Mettl3 may co-regulate pathways which guide axon projection.

2.3.6 Nab2 binds *Sxl* pre-mRNA and modulates its m^{6}A methylation.

The ability of the *Mettl3^{null}* allele to promote appropriate sex-specific exon 2-exon 4 splicing of *Sxl* in *Nab2^{ex3}* females is significant, both because it identifies the m^{6}A 'writer' Mettl3 as required for *Sxl* mis-splicing in heads lacking Nab2, and because this same *Mettl3* allele normally results in

hypomethylation of *Sxl* mRNA and exon 3 inclusion in female flies (98, 113, 127). This paradox could be explained if exon 3 inclusion in *Nab2*^{ex3} females is due to hypermethylation of the *Sxl* pre-mRNA, which is then suppressed by removing *Mettl3*. To test this hypothesis, a series of primer sets was designed to examine *Sxl* pre-mRNA and mRNA transcripts by RNA-immunoprecipitation (RIP) and anti-m^{6}A-RIP (MeRIP) (**Figure 6A**). As illustrated in **Figure 6A**, the *Sxl* transcript contains candidate binding sites for both Sxl protein (polyuridine tracts=**red** ticks) and Nab2 protein (polyadenosine tracts=**green** ticks), and approximate sites of m^{6}A methylation (**yellow** ticks) (mapped in 98) (see **Supplemental Figure 7** for a complete schematic). To assess the m^{6}A status of total *Sxl* RNA, MeRIP precipitates from female head lysates (control, *Nab2*^{ex3}, *Mettl3*^{null}, and *Nab2*^{ex3},*Mettl3*^{null}) were analyzed by reverse transcription-real time quantitative PCR (RT-qPCR) with the exon 2-exon 2 (E2-E2) primer pair, which amplifies both pre-mRNA and mature mRNA (*Sxl*$^{E2\text{-}E2}$ in **Figure 6B**). This approach detects reduced *Sxl* m^{6}A in *Mettl3*^{null} heads relative to controls, which is consistent with prior studies (98, 113, 127), and an increase in *Sxl* transcript recovered from MeRIP of *Nab2*^{ex3} heads, consistent with *Sxl* hypermethylation. This apparent increase in *Sxl* m^{6}A methylation in *Nab2*^{ex3} heads requires Mettl3, as *Sxl* mRNA recovery in MeRIP is strongly reduced in *Nab2*^{ex3},*Mettl3*^{null} double mutant heads. Two m^{6}A-methylated candidate Mettl3-target RNAs, *Act5c* and *Usp16* (98, 113) were analyzed as additional positive controls for m^{6}A status. MeRIP-qPCR indicates that both mRNAs are hypomethylated in *Mettl3*^{null} and hypermethylated in *Nab2*^{ex3} (**Supplemental Figure 8**). For *Act5c*, this *Nab2*^{ex3} hypermethylation requires *Mettl3* (**Figure 6B**). Shifting this analysis to qPCR with the *Sxl* E2-E4 primer set (*Sxl*$^{E2\text{-}E4}$ in **Figure 6B**), which is predicted to selectively detect spliced *Sxl* mRNAs, supports a very similar pattern of elevated *Sxl* m^{6}A in *Nab2*^{ex3} heads that requires *Mettl3.* These MeRIP-qPCR data argue that Nab2 either inhibits Mettl3-mediated m^{6}A deposition or promotes m^{6}A removal on *Sxl* mRNAs, which in turn controls patterns of exon 3 retention/skipping. A prediction of this model

is that Nab2 loss should result in hypermethylation of the *Sxl* pre-mRNA. Testing this hypothesis in MeRIP precipitates with the I3-E3 primer pair ($Sxl^{I2\text{-}E3}$ in **Figure 6C**) or the I3-E4 primer pair ($Sxl^{I3\text{-}E4}$ in **Figure 6C**) reveals moderate (1.5-fold) enrichment for intron 2-containing *Sxl* RNAs in $Nab2^{ex3}$ heads, and stronger (4.5-fold) enrichment for intron 3-containing RNAs, consistent with elevated m^6A on *Sxl* pre-mRNAs that still contain introns 2 and 3. A parallel anti-Flag IP from head lysates of adult females expressing N-terminally tagged Nab2 specifically in neurons (*elav>Flag:Nab2*), coupled with RT-qPCR with I3-E4 primers, indicates that Nab2 associates with unspliced *Sxl* pre-mRNA (**Figure 6D**). In sum, these data provide a molecular framework to interpret *Nab2-Mettl3-Sxl* interactions in which Nab2 associates with the *Sxl* pre-mRNA, perhaps via the poly(A) sites located in I2 and I3 (**green** ticks; **Figure 6A**), and prevents its Mettl3-dependent hypermethylation, thus ensuring a level of m^6A necessary to guide *Sxl* exon 3 in the developing female nervous system.

2.4 Discussion

He we have employed an unbiased high-throughput RNA sequencing approach to identify head-enriched RNAs whose levels or structure are significantly affected by Nab2 loss. Bioinformatic filtering of this high read-depth dataset reveals changes in levels and structure of a relatively small set of transcripts, with the latter effect on RNA structure traced to splicing defects - intron retention, alternative 5' and 3' splice site usage, and exon skipping - in a small group of approximately 150 mRNAs. Among these, the most significant change is female-specific inclusion of exon 3 in the *Sxl* mRNA and use of a cryptic 5' splice site within the downstream intron. Because *Sxl* exon 3 contains a stop codon, its inclusion is predicted to disrupt female-specific expression of Sxl protein, a U-rich RNA binding protein that controls somatic and germline sexual identity via effects on splicing and translation of target mRNAs (rev. in 124, 139). Bioinformatic and genetic data indicate that *Sxl* mRNA may be an especially significant target of Nab2 in neurons: mis-spliced

RNAs in *Nab2* mutant female heads are highly enriched for predicted Sxl binding motifs, and an allele of *Sxl* that constitutively skips exon 3 (130) substantially reverses neurodevelopmental and behavioral defects in *Nab2* null females. Moving downstream of Sxl, alleles of male-specific dosage compensation complex (DCC) components, including the direct Sxl target *msl-2* (140, 141), also rescue phenotypic defects in *Nab2* mutant females. Given that Msl-2 is not normally expressed or active in females, these data provide evidence that masculinized *Sxl* splicing and *msl-2*/DCC activation are important contributors to phenotypes in *Nab2* mutant female flies. These results imply a fairly specific link between Nab2 and the *Sxl* exon 3 splicing machinery, which is confirmed by strong genetic interactions between *Nab2* and the *Mettl3* methyltransferase that deposits m^6A on *Sxl* pre-mRNA and promotes exon 3 skipping (98, 113, 127). Molecular assays provide key insight into these *Nab2*:*Sxl* interactions. The Nab2 protein associates with unspliced *Sxl* pre-mRNA in head lysates, and its loss results in Mettl3-dependent hypermethylation of mature *Sxl* mRNA and the unspliced *Sxl* pre-mRNA. Given the known role of m^6A in regulating *Sxl* exon 3 splicing, these data imply that Nab2 interacts with the *Sxl* pre-mRNA in the nucleus and opposes excessive m^6A methylation by the Mettl3 complex, thus ensuring a level of m^6A necessary to guide *Sxl* exon 3 skipping in the developing female nervous system.

Our findings are consistent with a fairly specific role for nuclear Nab2 in control of exon-specific splicing patterns within a small subset of head RNAs, most of which are shared between males and females. One Nab2-regulated transcript, the sex determination factor *Sxl*, is a sex-specific target of Nab2 in the female nervous system, and evidence indicates that altered *Sxl* splicing contributes to their phenotypic defects. As Sxl is itself an RBP that can control splicing, some fraction of the mis-sliced mRNAs detected in our high-throughput approach may not be direct Nab2 targets, but rather Sxl targets. This hypothesis is supported by the enrichment for predicted Sxl-binding sites among mis-spliced mRNAs in *Nab2* mutant female heads, and by the

substantial rescue conferred by the *Sxl*^{M8} allele. However, splicing of the Sxl target *tra* is unaffected in the *Nab2* mutant RNA-Seq datasets. This could be due to lack of read depth, although this does not seem to be the case (see **Supplemental Figure 4**), or to Sxl-independent *tra* splicing in adult heads. Unbiased screens for Sxl target RNAs have been carried out in ovaries (142) and primordial germ cells (143), but a similar approach has not been taken in the post-mitotic nervous system, where Sxl targets may differ from other tissue types. In this regard, the subgroup of Nab2-regulated head RNAs that also contain predicted U-rich Sxl binding motifs may be enriched for Sxl targets that contribute to developmental phenotypes in *Nab2* mutants.

The evidence for a *Drosophila* Nab2 role in splicing parallels evidence of accumulation of ~100 intron-containing pre-mRNAs in *S. cerevisiae* lacking the *Nab2* homolog (39), with only a few transcripts affected in yeast, similar to what is observed in the current *Drosophila* study. Rescue of *Nab2* mutants by neuron-restricted expression of human ZC3H14 (44) implies that this specificity may be a conserved element of Nab2/ZC3H14 proteins in higher eukaryotes. Indeed, knockdown of ZC3H14 in cultured vertebrate cells results in pre-mRNA processing defects and intron-specific splicing defects in the few RNAs that have been examined (83, 144). The basis for Nab2 specificity in *Drosophila* heads is not clear, but it could reflect a high degree of selectivity in binding to nuclear pre-mRNAs (e.g., *Sxl*), or to interactions between Nab2 and a second mechanism that defines splicing targets. The convergence of Mettl3 and Nab2 on the *Sxl* pre-mRNA represents the first evidence that Nab2 can modulate m^{6}A-dependent control of pre-mRNA splicing. Hypermethylation of *Sxl* that results from Nab2 loss could reflect a requirement to bind A-rich sequences and block access of the Mettl3 complex or to recruitment of an m^{6}A 'eraser'. Intriguingly the human homolog of *Drosophila* Virilizer, a m^{6}A methyltransferase subunit and splicing cofactor (145, 146), was recovered in an IP/mass-spectrometry screen for nuclear interactors of ZC3H14 (144), suggesting a potential functional interaction between these proteins

on shared target RNAs. Additional evidence that Nab2 restricts m^{6}A methylation of targets beyond *Sxl* (e.g., *Act5c* and *Usp16*; **Figs. 6B and S8**) highlights the possibility that Nab2 may have a broader role in modulating m^{6}A-dependent RNA processing events, such as splicing, turnover, trafficking and translation.

The robust *Mettl3^{null}* suppression of *Nab2^{ex3}* mushroom body (MB) defects is unlikely to be mediated solely by effects on *Sxl*, which has no reported role in MB development. This observation suggests that Nab2 and Mettl3 have an overlapping set of targets RNAs that extends beyond *Sxl*. As almost all *Nab2* mutant phenotypes originate from a Nab2 role in central nervous system neurons (44, 49, 73), suppression by *Mettl3^{null}* also implies that the Nab2-Mettl3 functional interaction is likely to play out in neurons. The broad phenotypic rescue of *Nab2^{ex3}* afforded by *Mettl3^{null}* is consistent with a model in which loss of Nab2 leads to m^{6}A hypermethylation of Nab2-Mettl3 shared targets, which is then suppressed by removing Mettl3. This 'goldilocks' model is consistent with the *Sxl* data presented here and implies that excessive m^{6}A caused by loss of Nab2 perturbs neuronal mRNA processing and/or expression in a manner similar to m^{6}A hypomethylation caused by loss of Mettl3. The precise effect of excess m^{6}A is unknown, but it could result in excessive recruitment of m^{6}A 'reader' proteins that overwhelm the specificity of downstream regulatory steps. Significantly, the ability of human *ZC3H14* to rescue *Nab2^{ex3}* viability and locomotion when expressed in fly neurons indicates that this new m^{6}A inhibitory role of Nab2 may be conserved within the Nab2/ZC3H14 family of RBPs, and that excessive m^{6}A methylation of key RNAs also contributes to neurological deficits in *ZC3H14* mutant mice and patients.

2.5 Figures

Figure 2—1

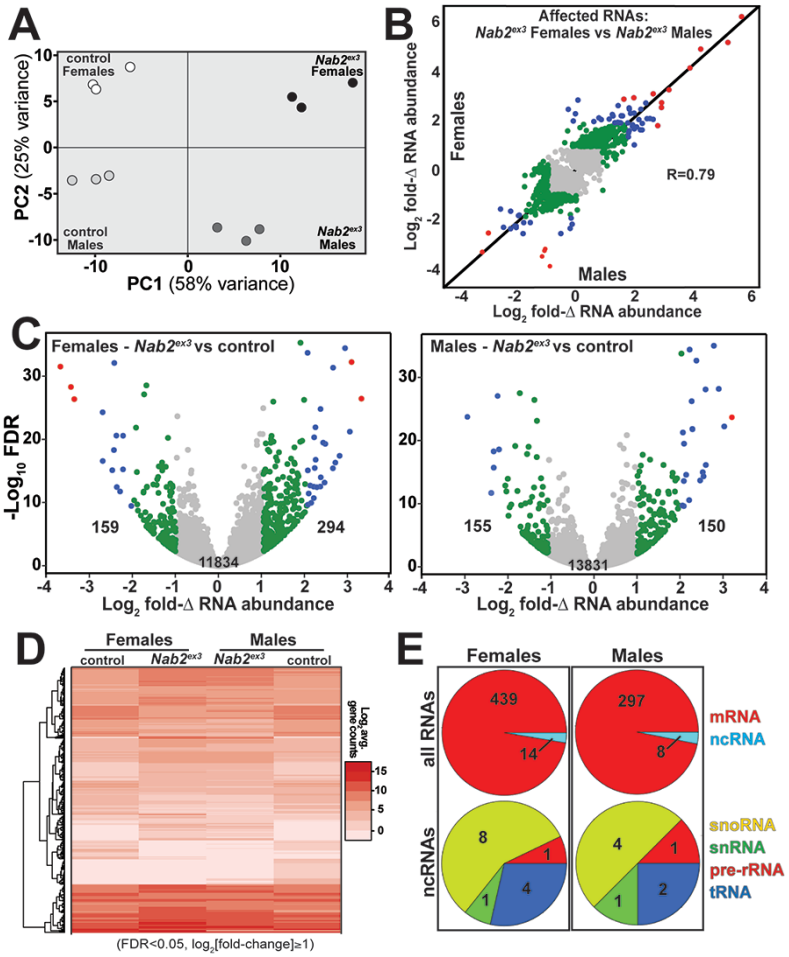


Figure 2—1: RNA sequencing detects effects of Nab2 loss on the head transcriptome. **(A)** Principal component analysis (PCA) of RNA-seq data from three biological replicates of control and *Nab2* mutant (*Nab2*^{ex3}) male and female heads. **(B)** Correlation scatter plot of $\log_2$ fold-change (Δ) in abundance of affected RNAs in males and females ($\log_2$ average gene counts: **grey**<1, 1≤**green**<2, 2≤**blue**<3, **red**≥3). **(C)** Volcano plots of fold-Δ in abundance vs. false discovery rate (FDR -$\log_{10}$) of affected RNAs in *Nab2*^{ex3} females and males (dot plot color coding as in B). Elevated (≥1), reduced (≤-1), and total RNAs are indicated. **(D)** Heatmap comparison of significantly changed gene counts (FDR<0.05; |$\log_2$ fold-Δ|≥1) in *Nab2*^{ex3} females and males vs. sex-matched controls. **(E)** Pie chart shows distribution of RNA classes among significantly affected RNAs detected in **C** and **D**.

Figure 2—2

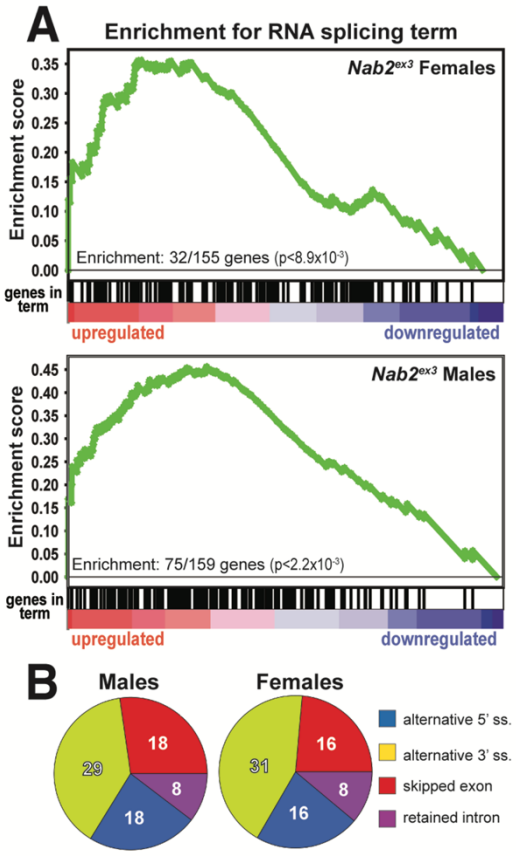

A
Enrichment for RNA splicing term
Nab2ex3 Females
Enrichment score
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Enrichment: 32/155 genes (p<8.9x10-3)
genes in term
upregulated
downregulated
Nab2ex3 Males
Enrichment score
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Enrichment: 75/159 genes (p<2.2x10-3)
genes in term
upregulated
downregulated
B
Males
Females
18
29
8
18
16
31
8
16
alternative 5' ss.
alternative 3' ss.
skipped exon
retained intron

Figure 2—2: Significantly up/down-regulated RNAs in *Nab2*^{ex3} heads are enriched for predicted splicing factors.

(A) Gene set enrichment analysis (GSEA) detects enrichment for the 'RNA splicing' GO term in up- and down-regulated gene sets in both female (top) and male (bottom) *Nab2*^{ex3} datasets. Gene enrichments are indicated with corresponding p-values. **(B)** Pie chart illustrating the distribution of previously annotated alternative splicing RNA splicing events that are significantly altered in *Nab2*^{ex3} mutant female and male heads (ss=splice site).

Table 2—1

Table 1

Alternative exon usage (DEXSeq) in *Nab2*^{ex3} head transcriptomes

	Females	Males
# of alternatively used exons *	**151**	**114**

* |(exon usage fold change)| > ~1.75 BH Adj. p < 0.05

Top misspliced transcripts

Females	foldΔ exon usage	BH adj. p-value
Sex lethal (Sxl)	2.86	3.08×10^{-235}
CG13124	2.45	1.09×10^{-81}
Ih channel	2.29	3.28×10^{63}
Ace	1.81	1.02×10^{-59}
Males	**foldΔ exon usage**	**BH adj. p-value**
Ace	2.02	1.88×10^{-169}
Pkc53E	1.74	6.12×10^{-102}
plx	2.12	9.03×10^{-67}
Pkn	1.84	9.04×10^{-67}

Figure 2—3

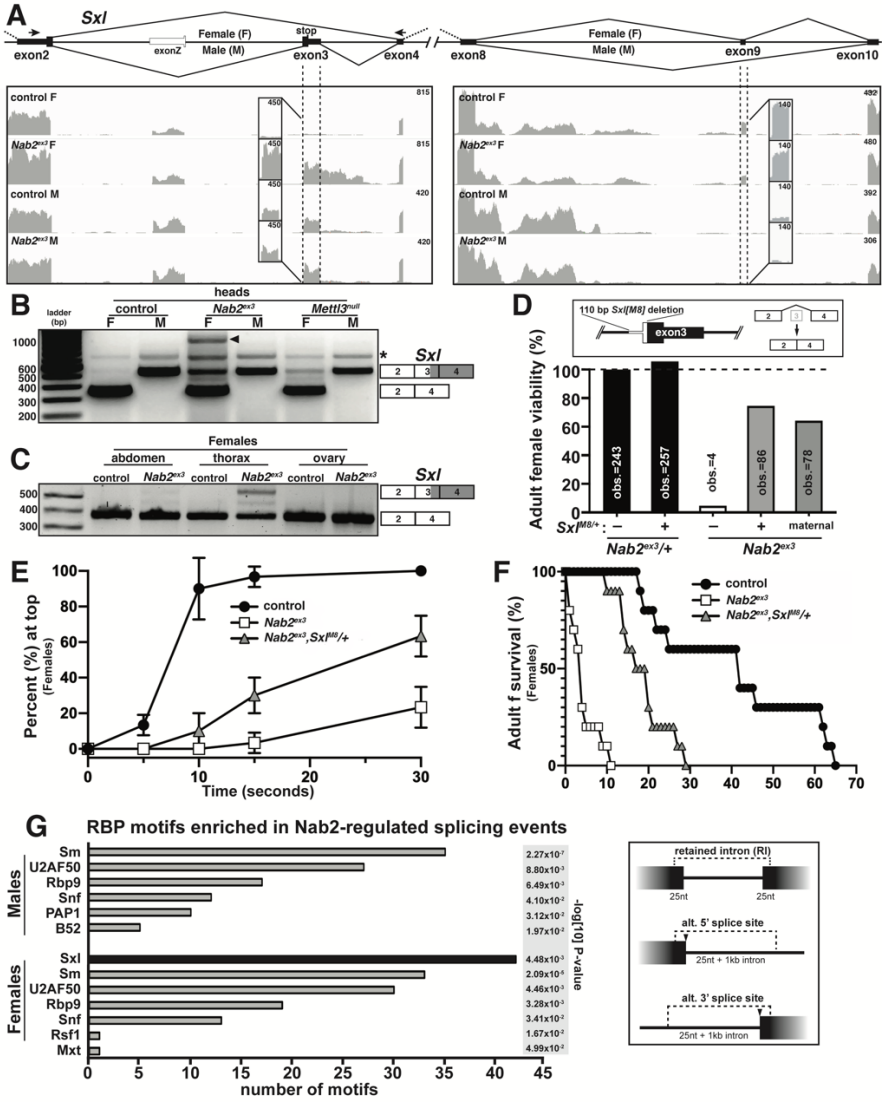

A
Sxl
exon2
exonZ
Female (F)
Male (M)
stop
exon3
exon4
exon8
exon9
exon10
control F
Nab2ex3 F
control M
Nab2ex3 M
B
heads
control
Nab2ex3
Mettl3null
ladder (bp)
Sxl
C
Females
abdomen
thorax
ovary
D
110 bp Sxl[M8] deletion
Adult female viability (%)
obs.=243
obs.=257
obs.=4
obs.=86
obs.=78
maternal
E
Percent (%) at top (Females)
Time (seconds)
control
Nab2ex3
Nab2ex3,SxlM8/+
F
Adult f survival (%) (Females)
G
RBP motifs enriched in Nab2-regulated splicing events
Males
Females
Sm
U2AF50
Rbp9
Snf
PAP1
B52
Sxl
Rsf1
Mxt
number of motifs
-log10[p-value]
retained intron (RI)
alt. 5' splice site
alt. 3' splice site
25nt
25nt + 1kb intron

Figure 2—3: Alternative splicing of *Sxl* is disrupted in *Nab2^{ex3}* female heads.

(A) Top panel: normal *Sxl* alternative splicing patterns across the exon 2-4 and exon 8-10 regions in females (F) and males (M). Arrows indicate location of primers used in **B**. **Bottom panel**: corresponding sequencing reads across the *Sxl* locus in the indicated sexes and genotypes. Dotted lines and boxed insets highlight exon 3 and exon 9 reads. **(B)** RT-PCR analysis of *Sxl* mRNA in control, *Nab2^{ex3}* and *Mettl3^{null}* male (M) and female (F) heads. Exon 2-3-4 and exon 2-4 bands are indicated. Arrowhead denotes exon 2-3-intron-4 product noted in text. Asterisk (*) is non-specific product. **(C)** RT-PCR analysis of *Sxl* mRNA in adult female (F) control and *Nab2^{ex3}* tissues using the same primers as in **B** and with exon 2-3-4 and 2-4 bands indicated. **(D)** A single copy of the *Sxl^{M8}* allele, which harbors a 110 bp deletion that causes constitutive exon2-4 splicing, partially suppresses lethality of Nab2^{ex3} females, both zygotically and maternally. **(E-F)** *Sxl^{M8}* dominantly (i.e., *M8/+*) suppresses previously defined locomotion (as assessed by negative-geotaxis) and life-span defects among age-matched *Nab2^{e3}* females. **(G)** RNA binding protein (RBM) motif enrichment analysis detects predicted Sxl binding sites as the most frequent motif among Nab2-regulated splicing events in female heads. Other enriched motifs are similar between male and female heads. Regions used for motif analysis (retained introns, and alternative 5’ or 3’ splice sites plus flanking sequence) are shown in the schematic.

Figure 2—4

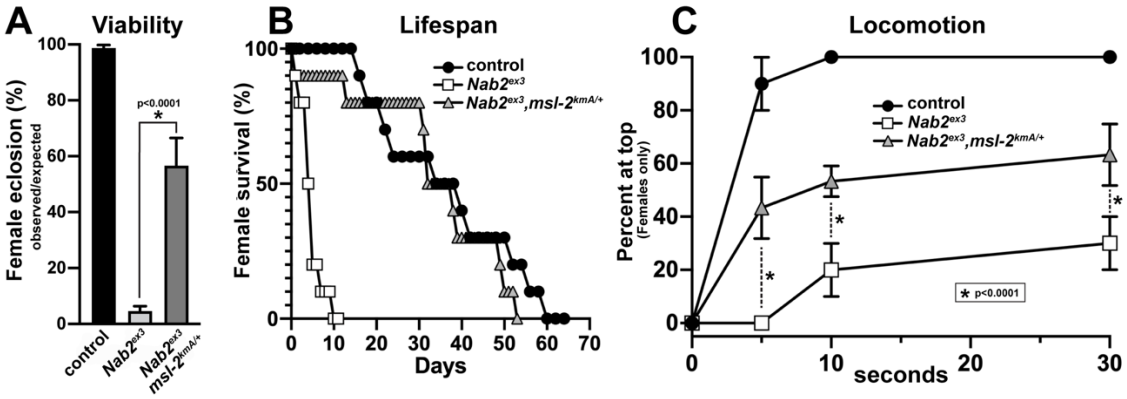

A
Viability
Female eclosion (%)
observed/expected
p<0.0001
*
control
Nab2ex3
Nab2ex3
msl-2kmA/+
B
Lifespan
Female survival (%)
control
Nab2ex3
Nab2ex3,msl-2kmA/+
Days
C
Locomotion
Percent at top
(Females only)
control
Nab2ex3
Nab2ex3,msl-2kmA/+
* p<0.0001
seconds

Figure 2—4: An allele of the DCC component *male-specific lethal-2* (*msl*-2) rescues *Nab2* phenotypes in females.

(A) Percent of control, $Nab2^{ex3}$, and $msl\text{-}2^{kmA}/+;;Nab2^{ex3}$ (*msl*-2 is on the X chromosome) females eclosing as viable adults (calculated as #observed/#expected). **(B)** Survival of age-matched adult female flies of the same genotypes. **(C)** Negative geotaxis of age-matched adult females of the same genotypes at 5sec, 10sec and 30 sec timepoints. Significance values are indicated (*p<0.0001).

Figure 2—5

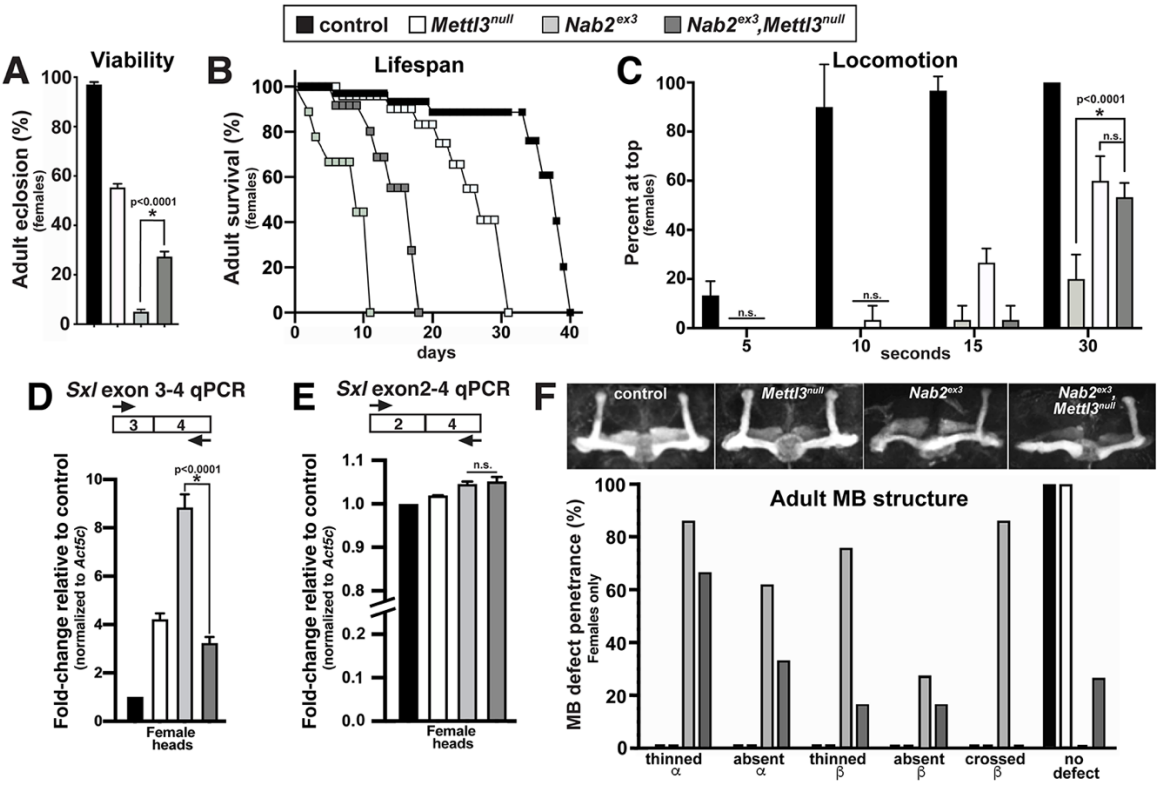

control
Mettl3null
Nab2ex3
Nab2ex3,Mettl3null
A
Viability
Adult eclosion (%)
(females)
p<0.0001
*
B
Lifespan
Adult survival (%)
(females)
days
C
Locomotion
Percent at top
(females)
p<0.0001
*
n.s.
seconds
D
Sxl exon 3-4 qPCR
Fold-change relative to control
(normalized to Act5c)
Female heads
E
Sxl exon2-4 qPCR
Fold-change relative to control
(normalized to Act5c)
Female heads
F
control
Mettl3null
Nab2ex3
Nab2ex3
Mettl3null
Adult MB structure
MB defect penetrance (%)
Females only
thinned α
absent α
thinned β
absent β
crossed β
no defect

Figure 2—5: Removing the Mettl3 m^{6}A transferase suppresses viability, behavioral, neuroanatomical and *Sxl* splicing defects in *Nab2* mutant females.

Color coding as indicated: control (**black** fill), *Mettl3*$^{null/null}$ (white fill), *Nab2*$^{ex3/3x3}$ (light grey fill), and *Nab2*$^{ex3/ex3}$*,Mettl3*$^{null/null}$ (**dark grey** fill).

(A) Percent of control, *Nab2*^{ex3} and *Nab2*^{ex3}*,Mettl3*^{null} females eclosing as viable adults (calculated as #observed**/**#expected). **(B)** Survival of age-matched adult female flies of the indicated genotypes. **(C)** Negative geotaxis of adult females of the indicated genotypes at indicated timepoints. Significance values are indicated at the 30sec timepoint (* p<0.0001; n.s.=not significant). **(D-E)** Quantitative real-time PCR analysis of the abundance of *Sxl* mRNA variants containing **(D)** the E3-E4 (exon3-exon4) or **(E)** the E2-E4 splice forms in female heads of the indicated genotypes. Primers denoted by arrows in the cartoon schematic. CT values are relative to control (*Nab2*^{pex41}) and normalized to *Act5c* control (*p<0.0001; n.s.=not significant). **(F) Top panel:** representative Z-stacked confocal images of anti-Fas2 staining to visualize MBs in female brains of indicated genotypes. **Bottom panel:** penetrance of MB phenotypes (thinned/absent a-lobes, thinned/absent b-lobes, and b-lobes mid-line crossing) in adult females of indicated genotypes (n=30 per genotype). Note that b-lobe crossing characteristic of *Nab2* nulls (>80% penetrance, light grey bar) is completely suppressed by loss of *Mettl3* (**dark grey** bar).

Figure 2—6

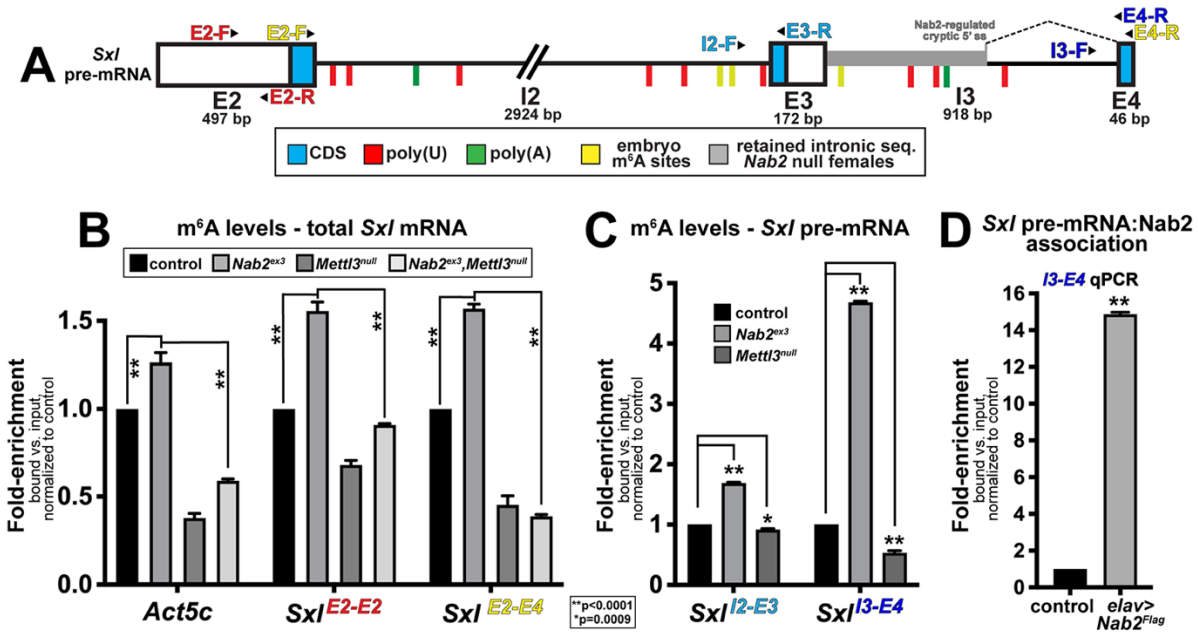

A
Sxl pre-mRNA
E2-F
E2-F
E2-R
E2
497 bp
I2
2924 bp
I2-F
E3-R
E3
172 bp
Nab2-regulated cryptic 5' ss
I3
918 bp
I3-F
E4-R
E4-R
E4
46 bp
CDS
poly(U)
poly(A)
embryo m6A sites
retained intronic seq. Nab2 null females
B
m6A levels - total Sxl mRNA
control
Nab2ex3
Mettl3null
Nab2ex3,Mettl3null
Fold-enrichment
bound vs. input, normalized to control
Act5c
Sxl E2-E2
Sxl E2-E4
**p<0.0001
*p=0.0009
C
m6A levels - Sxl pre-mRNA
control
Nab2ex3
Mettl3null
Sxl I2-E3
Sxl I3-E4
D
Sxl pre-mRNA:Nab2 association
I3-E4 qPCR
control
elav> Nab2Flag

Figure 2—6: Nab2 associates with the *Sxl* mRNA and inhibits its m^{6}A methylation. (A) Diagram of (E2, E3, E4) and introns (I2, I3) of the *Sxl* pre-mRNA annotated to show coding sequence (CDS; **blue**), the retained intronic region in *Nab2^{ex3}* females (**grey**), and locations of color-coded primer pairs (**E2-F** and **E2-R**, **E2-F** and **E4-R**, **I2-F** and **E3-R**, **I3-F** and **E4-R**), poly(U) sites **red lines**, poly(A) sites **green lines**, and mapped m^{6}A locations in *Drosophila* embryos **yellow lines** (98). **(B)** Quantitative real-time PCR analysis of *Act5c* and *Sxl* mRNAs present in anti-m^{6}A precipitates of control (*Nab2^{pex41}*; black), *Nab2^{ex3}* (**grey**), *Mettl3^{null}* (**dark grey**), or *Nab2^{ex3},Mettl3^{null}* (**light grey**) female heads. *Sxl* primer pairs are indicated (**E2-F**+**E2-R** and **E2-F**+**E4-R**). **(C)** Similar analysis as in **B** using **I2-F**+**E3-R** and **I3-F**+**E4-R** primer pairs to detect unspliced variants of *Sxl* mRNA in anti-m^{6}A precipitates of control (black), *Nab2^{ex3}* (**grey**) and *Mettl3^{null}* (**dark grey**) female heads. **(D)** Quantitative real-time PCR analysis with the **I3-F**+**E4-R** primer pair in anti-Flag precipitates from control (*Nab2^{pex41}*) and *elav-Gal4,UAS-Nab2:Flag* female heads. For all panels, 1-day old female heads were used in three biological replicates, and data represent bound vs. input ratios normalized to control (*Nab2^{pex41}*). Significance values are indicated (**p<0.0001, *p=0.0009).

2.5 Supplementary Figures

Figure 2—S1

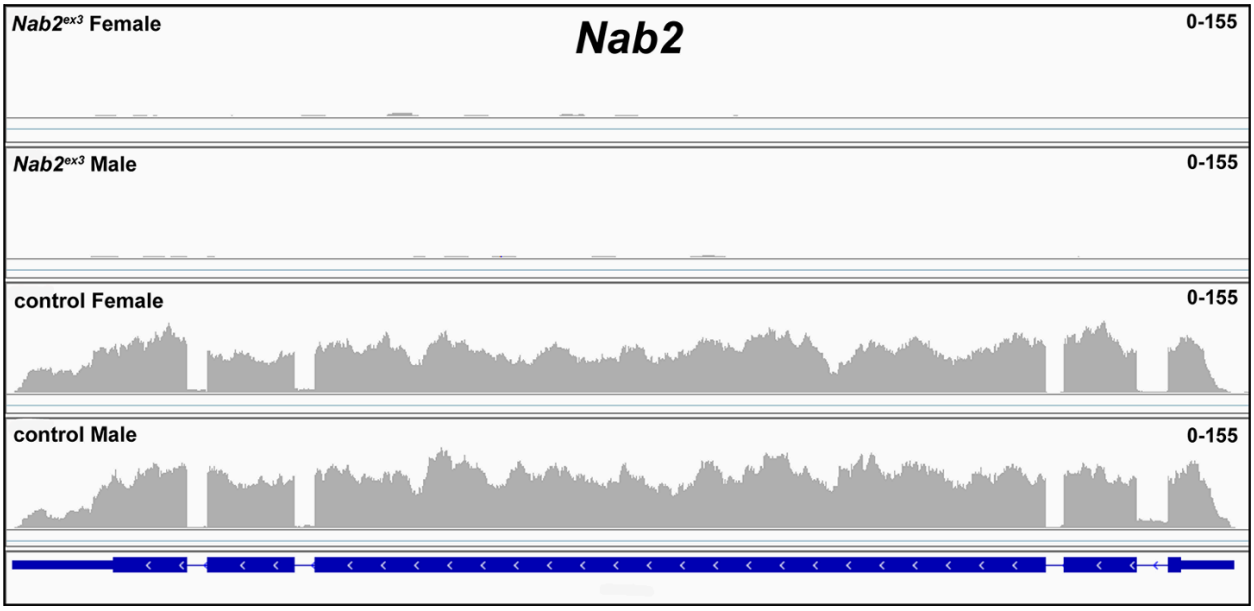


Figure 2—S1: RNA sequencing reads across the Nab2 locus.

Integrative Genomics Viewer (IGV) image of RNA sequencing reads across the Nab2 locus in $Nab2^{ex3}$ (top tracks) and control ($Nab2^{pex41}$) adult female and male heads. Intron-exon structure is indicated at bottom. Read depth scale is indicated (0-155).

Figure 2—S2

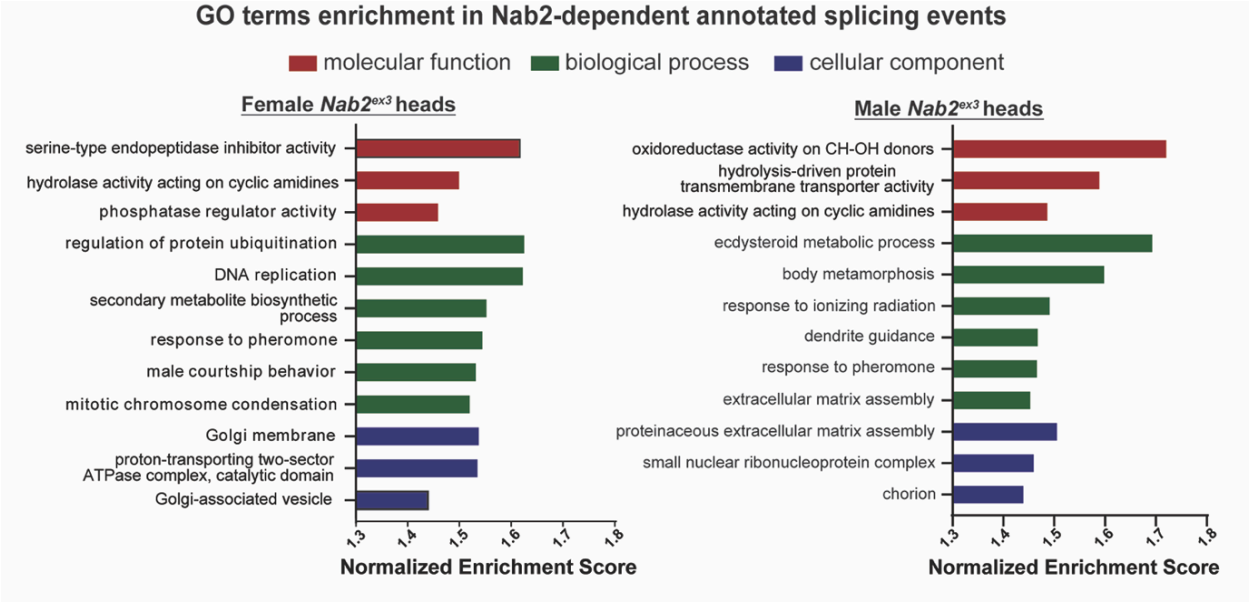

GO terms enrichment in Nab2-dependent annotated splicing events
molecular function
biological process
cellular component
Female Nab2ex3 heads
serine-type endopeptidase inhibitor activity
hydrolase activity acting on cyclic amidines
phosphatase regulator activity
regulation of protein ubiquitination
DNA replication
secondary metabolite biosynthetic process
response to pheromone
male courtship behavior
mitotic chromosome condensation
Golgi membrane
proton-transporting two-sector ATPase complex, catalytic domain
Golgi-associated vesicle
1.3
1.4
1.5
1.6
1.7
1.8
Normalized Enrichment Score
Male Nab2ex3 heads
oxidoreductase activity on CH-OH donors
hydrolysis-driven protein transmembrane transporter activity
hydrolase activity acting on cyclic amidines
ecdysteroid metabolic process
body metamorphosis
response to ionizing radiation
dendrite guidance
response to pheromone
extracellular matrix assembly
proteinaceous extracellular matrix assembly
small nuclear ribonucleoprotein complex
chorion
1.3
1.4
1.5
1.6
1.7
1.8
Normalized Enrichment Score

Figure 2—S2: GO term enrichment among Nab2-regulated alternative splicing events.

Chart illustrating gene ontology (GO) terms enriched in the 'molecular function', 'biological process' and 'cellular component' categories among altered alternative splicing events detected by MISO in female and male $Nab2^{ex3}$ head RNAs.

Figure 2—S3

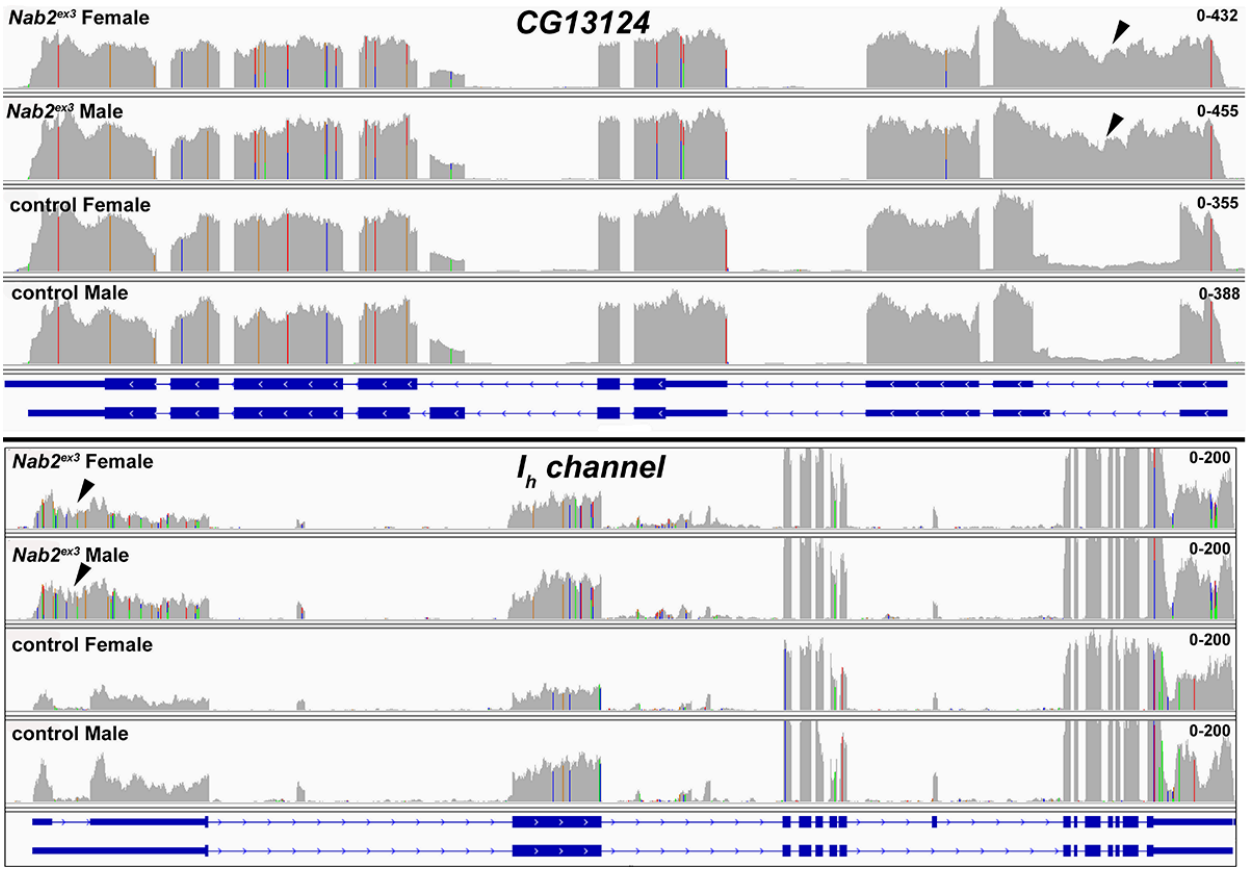

Nab2ex3 Female
CG13124
0-432
Nab2ex3 Male
0-455
control Female
0-355
control Male
0-388
Nab2ex3 Female
Ih channel
0-200
Nab2ex3 Male
0-200
control Female
0-200
control Male
0-200

Figure 2—S3: RNA sequencing reads across the CG13124 and Ih channel loci.

Integrative Genomics Viewer (IGV) images of RNA sequencing reads across CG13124 and Ih channel in Nab2ex3 (top tracks) and control (Nab2pex41) adult female and male heads. Intron-exon structure is indicated at bottom. Read depth scales are indicated. Arrowheads indicate reads across the first intron of each gene, consistent with intron-retention.

Figure 2—S4

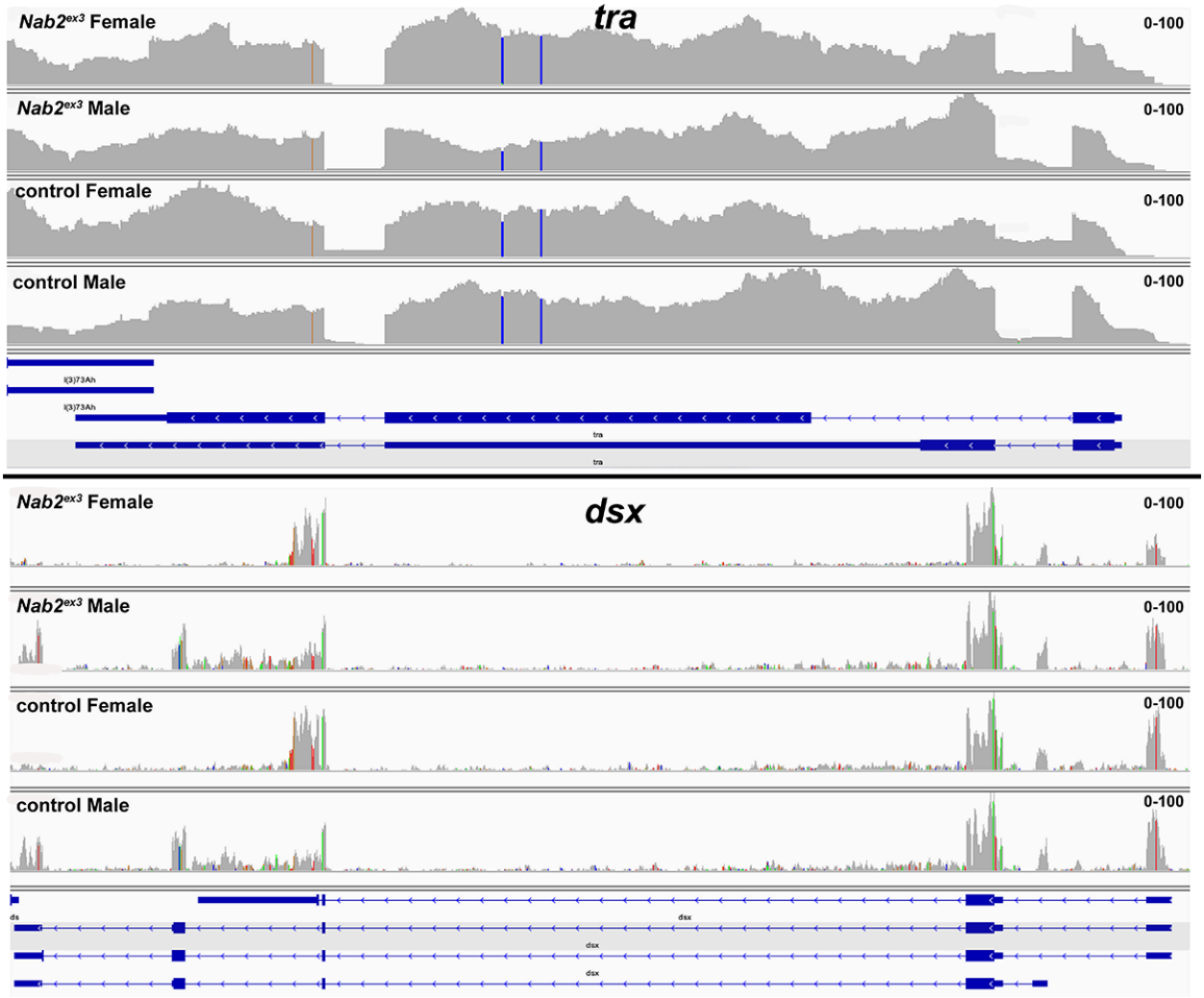

tra
Nab2ex3 Female
0-100
Nab2ex3 Male
0-100
control Female
0-100
control Male
0-100
l(3)73Ah
l(3)73Ah
tra
tra
dsx
Nab2ex3 Female
0-100
Nab2ex3 Male
0-100
control Female
0-100
control Male
0-100
dsx
dsx
dsx

Figure 2—S4: RNA sequencing reads across the tra and dsx loci.

Integrative Genomics Viewer (IGV) images of RNA sequencing reads across tra and dsx in Nab2ex3 (top tracks) and control (Nab2pex41) adult female and male heads. Intron-exon structure is indicated at bottom. Read depth scales are indicated.

Figure 2—S5

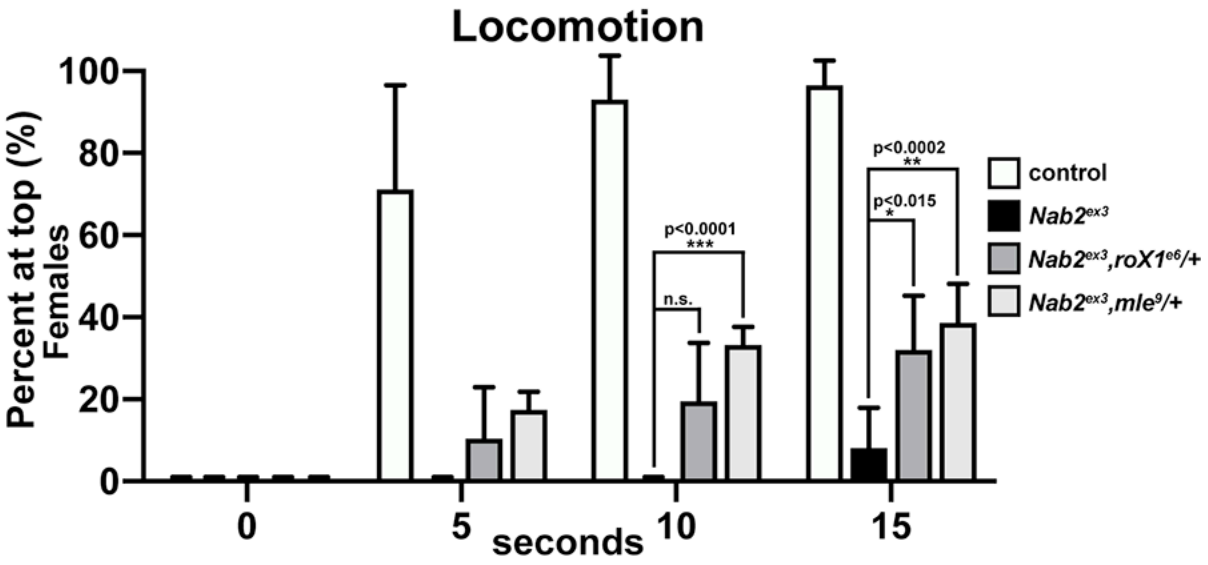

Locomotion
Percent at top (%)
Females
100
80
60
40
20
0
0
5
10
15
seconds
p<0.0001
***
n.s.
p<0.0002
**
p<0.015
*
control
Nab2ex3
Nab2ex3,roX1e6/+
Nab2ex3,mle9/+

Figure 2—S5: Modification of Nab2^{ex3} locomotor defect by roX1 and mle alleles.

Negative geotaxis of age-matched adult female controls (Nab2^{pex41}), Nab2^{ex3} mutants, or Nab2^{ex3} mutants carrying single copies of the roX1^{e6} or mle^{9} loss-of-function alleles at 5sec, 10sec, and15 sec timepoints. Significance values between indicated groups are indicated at the 30sec timepoint (p-values are indicated; n.s.=not significant).

Figure 2—S6

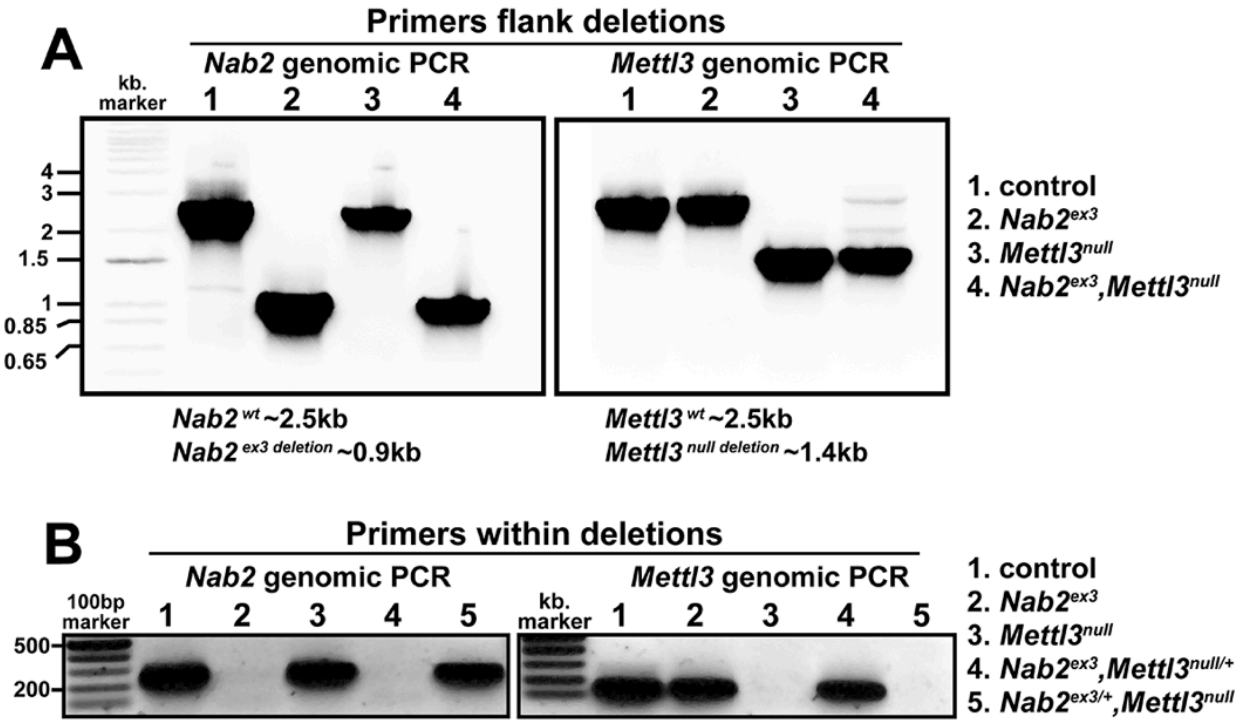

A
Primers flank deletions
Nab2 genomic PCR
Mettl3 genomic PCR
kb. marker
1 2 3 4
1 2 3 4
4
3
2
1.5
1
0.85
0.65
1. control
2. Nab2ex3
3. Mettl3null
4. Nab2ex3,Mettl3null
Nab2 wt ~2.5kb
Nab2 ex3 deletion ~0.9kb
Mettl3 wt ~2.5kb
Mettl3 null deletion ~1.4kb
B
Primers within deletions
Nab2 genomic PCR
Mettl3 genomic PCR
100bp marker
1 2 3 4 5
kb. marker
1 2 3 4 5
500
200
1. control
2. Nab2ex3
3. Mettl3null
4. Nab2ex3,Mettl3null/+
5. Nab2ex3/+,Mettl3null

Figure 2—S6: Genomic PCR confirms the Nab2^{ex3},Mettl3^{null} recombinant.

Genomic PCR on the indicated genotypes using primer pairs that either flank (**A**, top panel) or lie within (**B**, bottom panel) the ex3 deletion in Nab2 (left half of each gel) or the null crispr deletion in Mettl3 (right half of each gel). Approximate product sizes are indicated.

Figure 2—S7

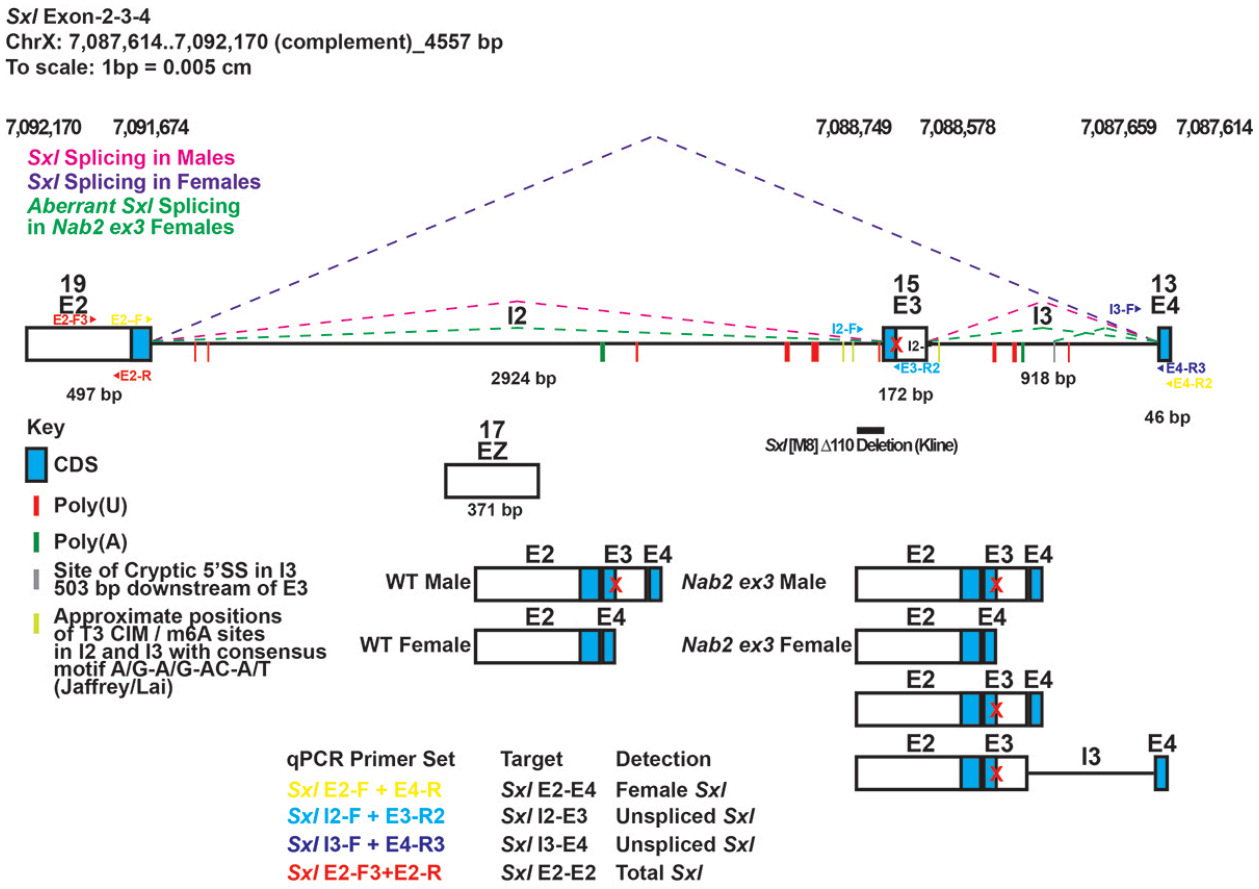

Sxl Exon-2-3-4
ChrX: 7,087,614..7,092,170 (complement)_4557 bp
To scale: 1bp = 0.005 cm
7,092,170
7,091,674
7,088,749
7,088,578
7,087,659
7,087,614
Sxl Splicing in Males
Sxl Splicing in Females
Aberrant Sxl Splicing in Nab2 ex3 Females
19
E2
15
E3
13
E4
I2
I3
E2-F3
E2-F
E2-R
I2-F
I2-
E3-R2
I3-F
E4-R3
E4-R2
497 bp
2924 bp
172 bp
918 bp
46 bp
Key
CDS
Poly(U)
Poly(A)
Site of Cryptic 5'SS in I3 503 bp downstream of E3
Approximate positions of T3 CIM / m6A sites in I2 and I3 with consensus motif A/G-A/G-AC-A/T (Jaffrey/Lai)
17
EZ
371 bp
Sxl [M8] Δ110 Deletion (Kline)
WT Male
WT Female
Nab2 ex3 Male
Nab2 ex3 Female
E2
E3
E4
I3
qPCR Primer Set
Target
Detection
Sxl E2-F + E4-R
Sxl E2-E4
Female Sxl
Sxl I2-F + E3-R2
Sxl I2-E3
Unspliced Sxl
Sxl I3-F + E4-R3
Sxl I3-E4
Unspliced Sxl
Sxl E2-F3+E2-R
Sxl E2-E2
Total Sxl

Figure 2—S7: Schematic of *Sex lethal* (*Sxl*) primers for qPCR.

Detailed schematic of the exon 2-3-4 *Sxl* locus with annotated locations of introns and exons annotated to show coding sequence (CDS; blue), the retained intronic region in *Nab2*^{ex3} females (grey), and locations of color-coded primer pairs (E2-F and E2-R, and , I2-F and E3-R, I3-F and E4-R), poly(U) sites red lines, poly(A) sites green lines, and mapped m^{6}A locations in *Drosophila* embryos (Kan et al., 2017). Colored dotted lines indicated sex-specific splicing in wildtype adults and the altered splicing documented in this study. Boxed areas below summarize exon-intron structure in wild type heads and *Nab2*^{ex3} heads. Base pair coordinates are indicated (Dm Release 6).

Figure 2—S8

Relative m^{6}A methylation (MeRIP-qPCR)

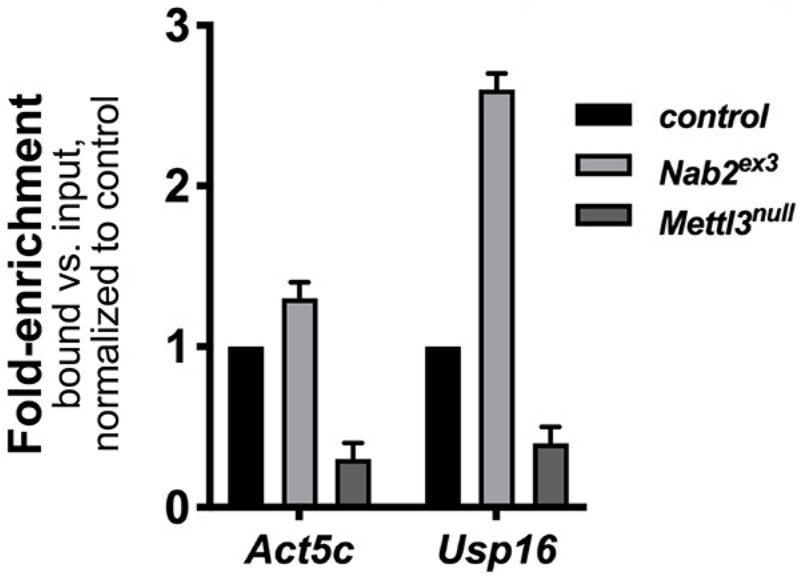


Figure 2—S8: Nab2 limits m^{6}A methylation of additional Mettl3 target RNAs.

Quantitative real-time PCR analysis of Act5c and Usp16 mRNAs present in anti-m^{6}A precipitates of control (Nab2^{pex41}; black), Nab2^{ex3} (grey), Mettl3null (dark grey) adult female heads. 1-day old female heads were used in three biological replicates, and data represent bound vs. input ratios normalized to control (Nab2^{pex41}). p-values are indicated (n.s.=not significant).

Tables

Table 2---S1

https://docs.google.com/spreadsheets/d/1NRF3FCA8rqsWBxmuCy1Um_Xe1nXV0J0rlVh0ZYN0tP0/edit?usp=sharing

Table 2---Supplemental Table 1: DESeq2 results for all genes

Table 2---S2

https://docs.google.com/spreadsheets/d/1_nSiQrMdN-Xsh6nzkG-Izhd1EZl2CQJY-ewi_dAdr7g/edit?usp=sharing

Table 2---Supplemental Table 2: MISO called AS events in Nab2 mutant heads

Table 2---S3

https://docs.google.com/spreadsheets/d/1aeBC7fAn1Ku0gpMxg8yt-fuPsEsHSpwFVgyg56116M8/edit?usp=sharing

Table 2---Supplemental Table 3: DEXSeq called differential exon usage for Nab2 mutant heads

Chapter 3

A *Drosophila* model of Pontocerebellar Hypoplasia reveals a critical role for the RNA exosome in neurons

***Please Note*:**

In the current chapter, Binta Jalloh contributed to the experiments and the data associated with these Figures (Figure 1B, 1D), showing the requirement for the *Drosophila* Rrp40 subunit of the RNA exosome for viability and brain development. Figure 2B, 2D, showing the viability and brain developmental phenotypes associated with a temperature sensitive knockdown of Rrp40RNAi. Figure 5, showing Rrp40 mutants with defects in the brain (mushroom body structures). Figures 8 and 9, showing RNA sequencing results from the RNA Binta Jalloh harvest and isolated from fly heads. The datasets from the sequenced RNAs were further processed and analyzed by others. All other *Drosophila* experiments were carried out and the associated text written by the remaining co-authors, primarily by Derrick J. Morton and Anita Corbett.

This chapter has been published in Plos Genetics:

Derrick J. Morton, **Binta Jalloh**, Lily Kim, Isaac Kremsky, Rishi J. Nair, Khuong B. Nguyen, J. Christopher Rounds, Maria C. Sterrett, Brianna Brown, Thalia Le, Maya C. Karkare, Kathryn D. McGaughey, Shaoyi Sheng, Sara W. Leung, Milo B. Fasken, Kenneth H. Moberg, Anita H. Corbett, "A *Drosophila* model of Pontocerebellar Hypoplasia reveals a critical role for the RNA exosome in neurons" PLoS genetics 16, no.7 (2020): e1008901
https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008901#

3.1 Abstract

The RNA exosome is an evolutionarily-conserved ribonuclease complex critically important for precise processing and/or complete degradation of a variety of cellular RNAs. The recent discovery that mutations in genes encoding structural RNA exosome subunits cause tissue-specific diseases makes defining the role of this complex within specific tissues critically important. Mutations in the RNA exosome component 3 (*EXOSC3*) gene cause Pontocerebellar Hypoplasia Type 1b (PCH1b), an autosomal recessive neurologic disorder. The majority of disease-linked mutations are missense mutations that alter evolutionarily-conserved regions of EXOSC3. The tissue-specific defects caused by these amino acid changes in EXOSC3 are challenging to understand based on current models of RNA exosome function with only limited analysis of the complex in any multicellular model in vivo. The goal of this study is to provide insight into how mutations in *EXOSC3* impact the function of the RNA exosome. To assess the tissue-specific roles and requirements for the *Drosophila* ortholog of EXOSC3 termed Rrp40, we utilized tissue-specific RNAi drivers. Depletion of Rrp40 in different tissues reveals a general requirement for Rrp40 in the development of many tissues including the brain, but also highlight an age-dependent requirement for Rrp40 in neurons. To assess the functional consequences of the specific amino acid substitutions in EXOSC3 that cause PCH1b, we used CRISPR/Cas9 gene editing technology to generate flies that model this RNA exosome-linked disease. These flies show reduced viability;

however, the surviving animals exhibit a spectrum of behavioral and morphological phenotypes. RNA-seq analysis of these *Drosophila Rrp40* mutants reveals increases in the steady-state levels of specific mRNAs and ncRNAs, some of which are central to neuronal function. In particular, *Arc1* mRNA, which encodes a key regulator of synaptic plasticity, is increased in the *Drosophila Rrp40* mutants. Taken together, this study defines a requirement for the RNA exosome in specific tissues/cell types and provides insight into how defects in RNA exosome function caused by specific amino acid substitutions that occur in PCH1b can contribute to neuronal dysfunction.

3.2 Author Summary

Pontocerebellar Hypoplasia Type 1b (PCH1b) is a devastating genetic neurological disorder that preferentially affects specific regions of the brain. Typically, children born with PCH1b have structural defects in regions of the brain including those associated with key autonomic functions. Currently, there is no cure or treatment for the disease. PCH1b is caused by mutations in the RNA exosome component 3 (*EXOSC3*) gene, which encodes a structural component of the ubiquitous and essential multi-subunit RNA exosome complex. The RNA exosome is critical for both precise processing and turnover of multiple classes of RNAs. To elucidate the functional consequences of amino acid changes in EXOSC3 that cause PCH1b, we exploited well-established genetic approaches in *Drosophila melanogaster* that model EXOSC3 mutations found in individuals with PCH1b. Using this system, we find that the *Drosophila* EXOSC3 homolog (termed Rrp40) is essential for normal development and has an important function in neurons. Furthermore, PCH1b missense mutations modeled in *Rrp40* cause reduced viability and produce neuronal-specific phenotypes that correlate with altered levels of target RNAs that encode factors with key roles in neurons. These results provide a basis for understanding how amino acid changes that occur in the RNA exosome contribute to neuronal dysfunction and disease.

3.3 Introduction

Precise production of both coding and non-coding RNAs requires regulation not only at the level of transcription but also at the post-transcriptional level to ensure precise processing and regulated turnover (147-149). The significance of post-transcriptional control of gene expression is underscored by the number of diseases associated with defects in RNA processing, surveillance and/or decay machinery (54). Specifically, mutations in genes encoding structural subunits of the ubiquitously expressed RNA exosome, a critical mediator of both RNA processing and decay, cause a variety of tissue-specific diseases (149-151).

The evolutionarily conserved RNA exosome complex, first identified in *Saccharomyces cerevisiae* (147, 148, 152), plays critical roles in RNA processing, decay and surveillance. The multi-subunit structure of the RNA exosome complex, which is conserved from archaea to humans, consists of ten subunits (147, 148, 153, 154). In humans, the RNA exosome subunits are termed EXOSome Component (EXOSC) proteins. In *S. cerevisiae* and *Drosophila*, most RNA exosome subunits are termed Rrp proteins after the original yeast genetic screen for ribosomal RNA processing (*rrp*) mutants (152). As illustrated in the Fig 1A, the complex is comprised of a catalytically inert core consisting of a hexameric 'ring' of catalytically inactive PH-like ribonuclease domain-containing core subunits 'capped' by a trimer of S1/KH putative RNA binding domain-containing cap subunits and a catalytically active 3' to 5' processive exo- and endoribonuclease, DIS3, which associates with the bottom of the 'ring-like' core. In *S. cerevisiae*, all core RNA exosome genes are essential (147, 148, 155). Furthermore, in *Drosophila* all RNA exosome genes studied thus far (*Dis3, Mtr3, Rrp6, Rrp41, Rrp42*) are essential and required for normal fly development (156-158). Studies in mice have been limited to assessing the role of *Exosc3* in B-cells *ex vivo* (159).

Given that the RNA exosome is essential for viability in model organisms examined thus far, it was surprising when mutations in genes encoding RNA exosome subunits were identified as the cause of several tissue-specific genetic diseases (160). Mutations in genes that encode structural subunits of the RNA exosome have now been linked to genetic diseases with distinct clinical presentations [reviewed in (29, 161)]. The first such example was the finding that recessive mutations in the *EXOSC3* gene, encoding a structural RNA exosome cap subunit (Fig 1A), cause a neurologic disease termed Pontocerebellar Hypoplasia type 1b (PCH1b) (149, 162). PCH1b is characterized by hypoplasia/atrophy of the cerebellum and pons that leads to severe motor and developmental delays with Purkinje cell abnormalities and degeneration of spinal motor neurons (162-164). Most individuals with PCH1b do not live beyond childhood. A number of missense mutations that cause single amino acid substitutions in conserved regions of EXOSC3 have been linked to PCH1b (151). Of note, clinical reports indicate that the severity of PCH1b pathology is influenced by *EXOSC3* allelic heterogeneity, suggesting a genotype-phenotype correlation (151, 165-167). These observations indicate that EXOSC3 plays a critical role in the brain and suggest specific functional consequences for amino acid changes linked to disease.

A previous study that modeled loss of *EXOSC3* in zebrafish by knockdown with antisense morpholinos reported embryonic maldevelopment, resulting in small brain size and reduced motility (151). While co-injection of wildtype zebrafish *exosc3* mRNA could rescue these phenotypes, *exosc3* containing mutations modeling PCH1b failed to rescue (151), suggesting that these mutations disrupt the normal function of EXOSC3. Other studies to assess the functional consequences of RNA exosome disease-linked amino acid changes in RNA exosome subunits have only been performed thus far in yeast (168, 169) providing little insight into how these specific changes could cause tissue-specific consequences. Thus, no studies have produced a genetic

model to explore the functional consequences of amino acid changes that cause disease in a multi-cellular model *in vivo*.

In this study, we exploit *Drosophila* as a model to understand the tissue-specific roles and requirements for Rrp40, the *Drosophila* ortholog of EXOSC3. This system also provides an *in vivo* platform to test the functional consequences of amino acid substitutions linked to disease by employing CRISPR/Cas9 gene editing technology to create an allelic series of mutations to model the amino acid substitutions that occur in PCH1b. This work reveals a critical role for *Rrp40* in neurons during development and demonstrates that mutations in Rrp40 in flies, which model genotypes in individuals with PCH1b, cause reduced viability, neuronal-specific behavioral phenotypes, and morphological defects in the brain. RNA-sequencing analysis reveals that *Rrp40* mutants modeling different patient genotypes show differences in levels of mRNAs and ncRNAs that are specific to each mutant providing insight into the molecular basis for the phenotypes observed. For example, we identify the neuronal gene *Arc1*, a critical regulator of neuronal function as increased in the RNA-seq dataset for both *Rrp40* mutants, underscoring the critical role the RNA exosome plays in regulating neuronal transcripts. Taken together, our results provide the first evidence for a critical function of the RNA exosome subunit Rrp40 in neurons. We show that the *Drosophila* model can be used to provide insight into tissue-specific function of the RNA exosome *in vivo* to explore the functional consequences of amino acid substitutions linked to distinct disease phenotypes.

3.4 Results

3.4.1 The RNA exosome subunit Rrp40 is essential for development.

To examine *in vivo* requirements for Rrp40, we employed RNAi to deplete Rrp40 (TRiP HMJ23923) from specific tissues/cell types. Successful knockdown of Rrp40 using this RNAi was confirmed by immunoblotting (**S1 Fig**). This RNAi transgene targeting the *Drosophila* EXOSC3 homolog Rrp40 was used in combination with a small panel of four *Gal4* lines with distinct expression patterns: *Actin5c>Gal4* (global); *elav>Gal4* (neurons); *repo>Gal4* (glia); and *Mhc>Gal4* (muscle) to define the requirement for EXOSC3. Expression of each *Gal4* driver to deplete *Rrp40* led to complete lethality, with no surviving adults (**Fig 1B**). This general developmental requirement for *Rrp40* in multiple tissues is consistent with previous studies of other RNA exosome subunits in *Drosophila* (156-158).

3.4.2 Rrp40 is required for proper development of the *Drosophila* mushroom body neurons.

To explore the requirement for Rrp40 in brain development, we utilized one structure in the fly brain that is not essential for viability (170), the mushroom body (MB). As illustrated in **Fig 1C**, each MB contains α/α' and β/β' composed of bundled axons that project from a dorsally located group of ~2000 Kenyon cells (78, 171). The β/β' branches project medially toward the ellipsoid and fan-shaped bodies of the midbrain, while the α/α' branches project dorsally (81, 172). Ellipsoid body neurons normally project axons medially towards the midline and eventually form a closed ring structure (81). The neuronal adhesion protein Fasciclin-II (FasII), which is enriched on α and β axon branches and regions of the Drosophila central complex (81), was used as a marker to assess the effect of loss of Rrp40 on MB structure. This analysis employed a *Gal4* driver specific for MBs (*OK107-Gal4*) in combination with the RNAi transgene targeting *Rrp40* (*UAS-*

Rrp40^{RNAi}; *OK107-Gal4*) to deplete Rrp40 specifically from MBs. Optical sectioning of FasII-stained fly brains reveals missing MB α- and β-lobes and outgrowth defects of β-lobes (arrows in **1D**) in the flies with Rrp40 depleted from MBs. We also noted defects in the pattern of FasII staining of the ellipsoid body (EB) in Rrp40-depleted flies. Thus, Rrp40 loss in MBs affects the morphology of the MB α- and β-lobes as well as ellipsoid bodies.

3.4.3 Rrp40 is required for age-dependent function in neurons.

To further investigate the requirement for Rrp40 in neurons, we employed an RNAi temperature shift strategy to bypass the embryonic requirement for Rrp40. As depicted in **Fig 2A**, a ubiquitously expressed, temperature-sensitive *tubulin(tub)-Gal80*^{ts} transgene was used to deplete *Rrp40* in combination with the panel of *Gal4* drivers. RNAi was induced by a shift from the permissive to restrictive temperature (18°C to 29°C) 24 hours after egg laying. Global *Rrp40* depletion (*Actin5c>Gal4*) using this temperature-shift still yielded no adults, indicating complete lethality; however, *Rrp40* depletion at this time-point in neurons and glia produced some viable adults (**Fig 2B**). In contrast, depletion of *Rrp40* from muscle caused only a mild survival defect. Intriguingly, adult survivors of neuronal *Rrp40* depletion had severely reduced lifespans (~1-2 days) relative to both *Gal4* driver alone and wildtype controls. By contrast, the lifespan of survivors of the glial or muscle depletion is indistinguishable from either control (**Fig 2C**).

To assess the requirement for Rrp40 in neurodevelopment, we exploited this temperature shift strategy to knockdown *Rrp40* expression in neurons in the developing brain (*elav>Rrp40*^{IR}*;tub>Gal80*^{ts}) and assess MB development by optical sectioning of FasII-stained brains. As shown in **Fig 2D**, pan-neuronal depletion of Rrp40 (*elav>Rrp40*^{IR}*;tub>Gal80*^{ts}) elicits MB defects, including thinned α-lobes and/or fused β-lobes (**Fig 2D**). In sum, these RNAi data argue

for a strict requirement for *Rrp40* and thus RNA exosome function in multiple tissues during embryonic development. Bypassing this embryonic requirement reveals an enhanced requirement for Rrp40 in neurons during later stages of development and adult homeostasis, including the branching and projection of axons into the α- and β-lobes.

3.4.4 *PCH1b* mutations engineered into the *Rrp40* locus impair viability and shorten lifespan.

A number of missense mutations encoding single amino acid substitutions in conserved regions of EXOSC3 have been linked to PCH1b (160). As shown in **Fig 3A**, The EXOSC3/Rrp40 protein consists of three domains: an N-terminal domain, an S1 putative RNA binding domain, and a K-homology (KH) putative RNA binding domain. Human EXOSC3 Glycine 31 [G31] and the analogous *Drosophila* G11 residue are located in a conserved region within the N-terminal domain, while EXOSC3 Aspartate 132 [D132] and the analogous *Drosophila* D87 residue are located in a conserved region within the S1 domain (**Fig 3A**).

We utilized CRISPR/Cas9-mediated genome editing technology to create alleles of Rrp40 corresponding to amino acid substitutions in EXOSC3 that cause PCH1b disease (**Fig 3B**). As illustrated in **Fig 3B**, *Drosophila* allows for true modeling of both homozygous and heterozygous alleles that cause disease. In parallel, a control *Rrp40* chromosome was also created using a wildtype Homology Directed Repair (HDR) donor template to create flies we refer to as $Rrp40^{wt}$ that had undergone the same editing process as flies with engineered base substitutions. As described in Materials and methods and illustrated in **S2 Fig**, all three *Rrp40* alleles ($Rrp40^{wt}$, $Rrp40^{G11A}$, and $Rrp40^{D87A}$) contain the *3xP3-DsRed* cassette inserted in the intergenic region downstream of the *Rrp40 3'UTR*, which allows for visual sorting by red fluorescence in eye tissue. The HDR donor templates containing *3xP3-DsRed* were injected into *nos-Cas9* embryos. To

confirm targeting events in DsRed-positive flies, genomic DNA was amplified and sequenced in the Rrp40 wildtype (*Rrp40^{wt}*) control flies and in flies containing missense mutations (independent flies, n=5) (*Rrp40^{G11A} and Rrp40^{D87A}*) (**S3 Fig**).

Initial analysis of these missense *Rrp40* alleles revealed reduced homozygote viability to adulthood within intercrosses of heterozygous parents (*Rrp40^{G11A}*: 14% of expected, n=126, and *Rrp40^{D87A}*: 38% of expected, n=182) (**Fig 3B**). In contrast, *Rrp40^{wt}* control flies eclose at a rate comparable to wildtype. The numbers of surviving adults were further reduced by placing *Rrp40^{G11A}* or *Rrp40^{D87A}* over a genomic deletion ((2L) Exel600) that removes the *Rrp40* locus, suggesting that both of these missense alleles are hypomorphs, rather than amorphs or hypermorphs (**Fig 3B**). The lower viability of the *Rrp40^{G11A}* allele, either in trans to itself or to (Df(2L) Exel6005), indicates that this is a stronger loss-of-function allele than *Rrp40^{D87A}*.

Significantly, flies with the engineered *Rrp40^{wt}* allele have a similar lifespan to control wildtype (*w^{1118}*) flies within the tested 30-day testing period, showing that the *3xP3-DsRed* cassette inserted into the intergenic region downstream of *Rrp40* does not appreciably impair survival (**Fig 3C**). Thus, *Rrp40^{wt}* was included as a control for subsequent studies. In contrast to *Rrp40^{wt}*, Kaplan-Meier plots of *Rrp40^{G11A}* homozygotes and *Rrp40^{D87A}* homozygotes reveal significantly impaired survival, with all *Rrp40^{G11A}* animals dead by day 14 and all *Rrp40^{D87A}* homozygotes dead by day 25 (**Fig 3C**). We are also able to assess the lifespan of flies with the *Rrp40^{D87A}* allele in trans to a deficiency (*Df(2L) Exel6005*), which mimics a reported patient genotype (**Fig 3C**) (151). These flies have the shortest lifespan of all genotypes analyzed. As no flies with *Rrp40^{G11A}* in trans to a deficiency (*Df(2L) Exel6005*) were produced, we could not assess lifespan. Indeed, no

patient with EXOSC3 Glycine 31 [G31] over a deletion has been reported (Not detected, ND in **Fig 3B**).

3.4.5 *Rrp40* mutant flies exhibit age-dependent morphological and behavioral defects.

Rrp40^{G11A} and *Rrp40*^{D87A} adult survivors exhibit wing posture and behavioral defects that are consistent with neurological deficits (173). *Rrp40*^{G11A} and *Rrp40*^{D87A} adults position their wings in an abnormal "held-up and held-out" posture rather than folding them together over the dorsal surface of the thorax and abdomen (**Fig 4A**) and this phenotype worsens with age (**Fig 4B**). *Rrp40*^{G11A} and *Rrp40*^{D87A} mutant animals also show reduced locomotor activity in a negative geotaxis assay that becomes more severe with age (**Fig 4C**). Both genotypes also display bouts of spasm interposed with periods of paralysis (**S1**, **S2**, and **S3 Video**) relative to age-matched control *Rrp40*^{wt} flies. The severity of the wing posture and locomotor defects among *Rrp40* alleles follow a similar pattern: *Rrp40*^{G11A} causes a more severe phenotype than *Rrp40*^{D87A}, which is in turn is enhanced by (*Df(2L) Exel6005*). Notably, this pattern of phenotypes mimics the pattern of disease severity observed among individuals with PCH1b (160).

3.4.6 *Rrp40*^{G11A} mutant flies have a mushroom body β-lobe midline-crossing defect.

We analyzed MB structure in wildtype and *Rrp40* mutants by immunostaining for FasII in whole adult brains. Confocal microscopy of *Rrp40*^{wt} flies (n=18), reveals no apparent MB defects in bifurcation or projection of α- and β-lobes neurons (***Rrp40***^{wt}, **Fig 5A**). In contrast, brains of *Rrp40*^{G11A} mutants show β-lobe fibers encroaching toward the midline, 56% (n=23) (**Fig 5A**), sometimes sufficient to cause apparent fusion of the left and right β-lobes

(**Fig 5A**). In contrast, *Rrp40*^{D87A} mutant flies (n=18) show no detectable defects in α- or β-lobes (***Rrp40***^{D87A}, **Fig 5**).

3.4.7 Pan-neuronal expression of human *EXOSC3* is sufficient to rescue behavioral phenotypes in *Rrp40* mutant flies.

To both ensure that phenotypes observed are due to the engineered *Rrp40* mutations and explore functional conservation between Rrp40 and human EXOSC3, we created transgenic flies that express human *EXOSC3-myc* from a UAS inducible promoter. The *UAS-EXOSC3-myc* transgene was expressed in the neurons (*elav>Gal4*) or muscle (*Mhc>Gal4*) of *Rrp40* mutant animals (**Fig 6A).** Notably, *UAS-EXOSC3-myc* produced a mild but consistent rescue of *Rrp40* mutant locomotor defects in the absence of the *Gal4* driver (rescue control, **Fig 6B**), suggesting leaky expression from the transgene. Addition of the *elav>Gal4* (neuronal) driver led to a robust rescue of climbing behavior: *Rrp40* mutant animals with neuronal expression of *EXOSC3* (*elav>EXOSC3*) perform normally in locomotor assays (**Fig 6B**); however, *Rrp40* mutant animals with expression of *EXOSC3* in muscle (*Mhc>EXOSC3*) did not show any statistically significant rescue (**Fig 6B**). Thus, human *EXOSC3* can restore functions that are lacking in adult *Rrp40* mutant neurons. This phenotypic rescue also confirms that the locomotor phenotype observed is due to the introduced CRISPR/Cas9 mutations.

3.4.8 Steady-state levels of RNA exosome subunit levels are affected in *Rrp40* mutant flies.

Amino acid substitutions in Rrp40 could alter protein steady-state levels or impact the levels of other RNA exosome complex components. To test this possibility in vivo, we compared wildtype and *Rrp40* variant levels in fly heads. As shown in **Figure 7A**, we performed immunoblotting of RNA exosome cap subunits (Rrp4 and Rrp40) and a core ring subunit (Mtr3) (see **Fig 1A**). We quantitated the results for Rrp40 and Rrp4, which reveals that levels of these cap subunits are modestly reduced in *Rrp40* mutant fly heads (**Fig 7B**) compared to control. However, the decrease in levels of Rrp40 and Rrp4 is not statistically significantly different between the two mutants, suggesting these decreases are not sufficient to explain the phenotypic differences detected.

3.4.9 RNA-sequencing analysis of *Rrp40* alleles reveals distinct gene expression profiles.

To identify relevant RNA targets of the RNA exosome, we performed RNA-Seq analysis on three biological replicates of adult heads for each of the following genotypes ($Rrp40^{wt}$, $Rrp40^{G11A}$ and $Rrp40^{D87A}$) from newly eclosed (Day 0) flies. Unbiased principal component analysis (PCA) of the resulting RNA expression data produced three distinct clusters (**Fig 8A**), indicating that each PCH1b-linked *Rrp40* allele alters the steady-state transcriptome relative to control $Rrp40^{wt}$. The transcriptomes of the two missense *Rrp40* mutants also differ from one another but not to same degree that either mutant differs from wildtype.

The high degree of reproducibility amongst the RNA-Seq replicates allowed us to identify transcripts that show statistically significant changes in each of the *Rrp40* mutant backgrounds relative to wildtype controls (**Fig 8B**,**C** and **S2** and **S3 Tables**). Consistent with the primary role of the RNA exosome in RNA decay (174), the majority of gene expression changes detected for $Rrp40^{G11A}$ (**Fig 8B**) or $Rrp40^{D87A}$ (**Fig 8C**) compared to $Rrp40^{wt}$ are increases in steady-state RNA

transcript levels. To illustrate this point, of the 385 transcripts that change 5-fold or more in the *Rrp40*^{G11A} mutant sample, 337 (88%) are increased (q<0.05), while of the 435 (84%) transcripts that change 5-fold or more in the *Rrp40*^{D87A} mutant sample, 366 (84%) (q<0.05) are increased. The total transcriptomic changes in the mutant *Rrp40* alleles show high correlation (R=0.85) to one another based on comparing each *Rrp40* mutant to wildtype (**Fig 8D**). This analysis suggests a correlation in both the direction and magnitude of transcriptomic changes.

To provide some insight into the transcripts that are likely to be direct targets of the RNA exosome in this first analysis in neuronal tissues, we analyzed transcripts that show a 5-fold or more increase in steady-state levels. Of the total number of altered transcripts that change 5-fold or more, a majority are mRNAs (61% in *Rrp40*^{G11A} and 66% in *Rrp40*^{D87A}) and the remainder include several classes of ncRNAs (**Fig 8E**). The different classes of effected ncRNAs include lncRNA, snoRNA, asRNA, tRNA, scaRNA and sisRNA (**Fig 8F**). We assessed whether the *Rrp40*^{G11A} and *Rrp40*^{D87A} alleles show an increase in overlapping or distinct sets of RNAs (q<0.05, >5-fold change). Consistent with the correlation coefficient comparing overall expression changes (**Fig 8D**), we found many coordinately increased coding and non-coding transcripts between the two groups, but also some RNAs that were only increased >5-fold in one of the mutants (**S4 Fig**).

FlyEnrichr analysis (FlyEnrichr; amp.pharm.mssm.edu/FlyEnrichr/) of coding transcripts significantly increased (q<0.05, >1.2-fold change) in *Rrp40*^{G11A} and *Rrp40*^{D87A} heads identified enriched pathways linked to human disease, including frontotemporal dementia, amyotrophic lateral sclerosis, and neurodegenerative disease (**Fig 8G**).

3.4.10 RNA-sequencing analysis reveals increased steady-state levels of key neuronal genes in *Rrp40* mutant flies.

RNA-sequencing analysis revealed increased steady-state levels of multiple functionally important neuronal genes including *Arc1* [*Rrp40*^{G11A} (30-fold increase); *Rrp40*^{D87A} (16-fold increase)], *Or82a* [*Rrp40*^{G11A} (35-fold increase); *Rrp40*^{D87A} (2-fold increase)], *PNUTS* [*Rrp40*^{G11A} (3-fold increase); *Rrp40*^{D87A} (3-fold increase)], *WFS1* [Rrp40^{G11A} (7-fold increase); *Rrp40*^{D87A} (6-fold increase)], and *knon* [*Rrp40*^{G11A} (24-fold increase); *Rrp40*^{D87A} (17-fold increase)] in *Rrp40* mutants (**Fig 9**). We show Integrated Viewer (IGV) tracks generated from the RNA-seq dataset for *Arc1* (**Fig 9A**) as well as *Or82a*, *PNUTS*, *WFS1* and *knon* (*CG7813*) (**S5A-D Fig**). To validate altered gene expression from RNA-seq analysis, qRT-PCR was performed. This analysis confirmed that steady-state levels of *Arc1* mRNA in *Rrp40* mutants (*Rrp40*^{G11A} and *Rrp40*^{D87A}) id significantly increased as compared to control (*Rrp40*^{wt}) (**Fig 9B**). As an additional measure of validation by RT-qPCR, we analyzed *CG8617* (**Fig 9B**), a neighboring gene of *Arc1* that showed no change in RNA-seq reads (**Fig 9A**). To extend the validation, we analyzed target transcripts increased in both *Rrp40* mutants compared to *Rrp40*^{wt}, including *Or82a*, *PNUTS*, *WFS1* and *knon* (**Fig 9C**). These results validate our RNA-seq analysis as well as highlight transcripts that may be potential regulators of neuronal function, providing candidates to help define the molecular mechanism for the phenotypes observed in both *Rrp40* mutants. Next, we tested whether expression of wildtype human *EXOSC3* (*UAS-EXOSC3-myc*) only in neurons could rescue increased steady-state levels of validated transcripts, *Arc1* and *WFS1*. RT-qPCR confirmed that steady-state levels of these transcripts are significantly decreased compared to control UAS alone (-), when *EXOSC3* is expressed in neurons using a *Gal4* driver (+*Gal4*) (**Fig 9D**), providing evidence for molecular rescue by human EXOSC3 to complement rescue of behavioral phenotypes (**Fig 6B**).

3.5 Discussion

In this study, we generated a *Drosophila* model to explore the requirement for RNA exosome function in specific tissues/cell types and define the functional consequences of amino acid substitutions in structural subunits of this complex that are linked to human disease. We provide evidence that the *Drosophila* Rrp40 subunit is essential for a critical step in early development. We also define an age-dependent requirement for Rrp40 in neurons critical for later stages of development and adult homeostasis. We extended this analysis to model genotypes that occur in PCH1b. These *Rrp40* missense mutations cause reduced viability and produce neuronal-specific behavioral phenotypes that align with PCH1b severity, revealing a genotype-phenotype correlation. The $Rrp40^{G11A}$ mutant flies show defects in mushroom body morphology, highlighting the neurodevelopmental requirement for Rrp40. RNA-sequencing of the brain-enriched transcriptome of *Rrp40* mutant flies identifies specific changes in both coding RNAs, including the synaptic regulator *Arc1*, and non-coding RNAs. Consistent with impaired RNA exosome activity the majority of transcripts that change show an increase in steady-state levels. These data establish a critical role for the RNA exosome in neurons and provide an initial characterization of the functional consequences of amino acid changes that underlie PCH1b in vivo using a multicellular organism.

Previous studies of the RNA exosome have provided evidence that individual subunits of this complex are essential in the organisms analyzed (147, 148, 155-157, 159). Thus, identification of mutations in genes encoding structural subunits of the RNA exosome that cause tissue-specific pathology was unexpected. In fact, only a few studies have explored the requirement for the RNA exosome in any specific tissues or cell types (157, 158, 175-177). Prior work using methods to deplete individual subunits of the complex in *Drosophila* revealed that these subunits are essential (157, 178). More recent work examining the role of the EXOSC3 subunit in human embryonic stem

cells built on studies that have implicated the RNA exosome in restraining differentiation (175-177). If there are specific requirements for the RNA exosome that are most critical for the development and/or homeostasis of certain cells or tissues, this could explain why mutations that impair the function of this complex preferentially affect certain tissues. Indeed, our studies exploiting *Drosophila* uncover an age-dependent requirement for Rrp40 in neurons. However, mutations in multiple genes that encode structural subunits of the RNA exosome termed exosomopathies, including two cap subunit genes, *EXOSC3* (151, 165, 179-181) and *EXOSC2* (150), and two core subunit genes, *EXOSC8* (149) and *EXOSC9* (182, 183), have now been linked to genetic diseases with a variety of clinical presentations reported. The finding that multiple mutations that likely impair the function of the RNA exosome cause different clinical consequences is not consistent with a simple model where a defined tissue or cell type is highly dependent on the function of this complex. Rather the missense mutations that have been identified in multiple RNA exosome subunit genes likely impair tissue-specific functions of this complex. Indeed, although the number of patients for any given exosomopathy is small, there are now multiple missense mutations that have been identified (160), raising the questions of whether different amino acid changes that occur within the same RNA exosome gene impair the function of the complex in the same manner or to the same extent. A combination of approaches can be employed to understand how the amino acid changes that occur in patients affect the function of the RNA exosome.

This study uncovered at least two requirements for Rrp40. As with other RNA exosome subunits (151, 157, 158), depletion of Rrp40 globally or in multiple tissues causes complete lethality. This reveals an essential role for the RNA exosome in early development. When we bypass this early requirement, depletion of Rrp40 in neurons causes a decreased lifespan, which is not observed for other cell types. Modeling PCH1b mutations in *Rrp40* allowed us to assess age-

dependent functional consequences of these amino acid changes. These PCH1b model flies show a number of behavioral and morphological changes linked to altered neuronal function (173, 184). Consistent with predicted neuronal dysfunction, depletion or mutation of *Rrp40* is sufficient to cause morphological defects in MBs including the loss of α-lobes and defective β-lobe axon extension. These defects argue that Rrp40 function is required for proper axon guidance and neurite extension within the brain.

We employed the *Gal4-UAS* system to demonstrate that neuronal expression of human *EXOSC3* is sufficient to rescue climbing deficits and molecular phenotypes in *Rrp40* mutants. In contrast to expression in neurons, expression of *EXOSC3* in muscle does not rescue climbing defects (**Fig 6B**). Neuronal expression of *EXOSC3* could also restore multiple transcripts that increase in *Rrp40* mutant flies to normal levels. Interestingly, while locomotor function is rescued in *Rrp40* mutants by neuronal-specific expression of *EXOSC3*, viability is not. These findings suggest that there is an early developmental requirement for *Rrp40* that *EXOSC3* cannot replace; however, *EXOSC3* is sufficient to replace *Rrp40* in adult homeostasis.

There are multiple mechanisms by which a single amino acid substitution could impair RNA exosome function (**Fig 10**). Ultimately, any of these mechanistic consequences would alter RNA exosome activity toward target RNAs. Changes in all or a subset of these RNAs could cause pathology.

The wildtype RNA exosome associates with cofactors to facilitate both decay and processing of target RNAs (**Fig 10A**) (174). In the mutant case, levels of the individual subunit or the overall assembled complex could decrease (a). In fact, a previous study using patient fibroblast and myoblasts to examine variants of EXOSC8, an RNase (PH)-like domain-containing core 'ring' subunit linked to another form of PCH (182, 183), not only showed reduced levels of the

EXOSC8 variants, but also EXOSC3, indicating that defects in core 'ring' subunits of the RNA exosome can lead to a reduction in other subunits (149). This finding is rather surprising given that the patients analyzed did not exhibit pathology related to the cell types analyzed. While our data show a modest decrease in levels of Rrp40 and an associated cap subunit Rrp4 (**Fig 7**), two lines of evidence suggest this decrease is not sufficient to explain the phenotypes reported in this study. First, both *Rrp40* alleles examined show a similar decrease in levels of Rrp40 as well as Rrp4, but $Rrp40^{G11A}$ mutant flies consistently show more severe phenotypes than $Rrp40^{D87A}$ flies (**Figs 3B**, **3C**, **4**, **5**). Furthermore, analysis of flies with a deficiency that removes the *Rrp40* gene, creating flies with a single copy of *Rrp40*, do not display any of the phenotypes detected in the engineered *Rrp40* mutant flies. These deficiency flies should have a 50% decrease in levels of *Rrp40*, which is a larger decrease than we detected in either *Rrp40* mutant fly (**Fig 7B**).

In addition to overall loss of the complex, amino acid changes in the RNA exosome could alter interactions with cofactors (b). Indeed, structural studies of the budding yeast RNA exosome show that the nuclear RNA exosome cofactor Mpp6 interacts with the Rrp40 cap subunit (185, 186), and that an amino acid change in Rrp40 that models a change found in PCH1b decreases the interaction of Mpp6 with the RNA exosome (186). The idea that interactions with specific cofactors could be perturbed by the amino acid changes that occur in disease is attractive because altered interactions with cofactors in different tissues or cell types could explain the phenotypic variation observed for mutations within the same *EXOSC* gene or the different *EXOSC* subunit genes that are linked to distinct diseases. RNA exosome cofactors have not been studied in *Drosophila*. The *Rrp40* mutant flies described here could be a valuable tool for future genetic studies to define such cofactors.

Given that EXOSC3/Rrp40 is a cap subunit, amino acid substitutions could also directly impair interactions with target RNAs (c). In budding yeast, the three cap subunits of the RNA

exosome make direct contact with target RNAs (154, 185), making this mechanistic consequence a formal possibility. Future studies will be required to understand how the different amino acid changes that occur in PCH1b impair the function of the RNA exosome.

As a first approach to consider how amino acid substitutions in EXOSC3 could impact the overall RNA exosome complex, we can take advantage of the availability of structures of this large macromolecular complex (153, 154, 185, 187, 188). While there is no structure available of the *Drosophila* RNA exosome, the considerable evolutionary conservation illustrated in **Fig 3B** provides some context to consider how distinct amino acid changes in EXOSC3/Rrp40 could impair RNA exosome function. As shown in **Fig 3A**, the amino acid changes modeled in this study lie within the N-terminal (G31A) and S1 (D132A) domains of EXOSC3/Rrp40. The G31 residue in EXOSC3 packs against the surface of EXOSC5 and, therefore, the EXOSC3-G31A substitution could impair the interaction of EXOSC3 with EXOSC5 (153). The D132 residue in EXOSC3 is located in a loop between strands in the S1 domain and the D132A substitution could thus impair the folding of the loop and, subsequently, disturb interactions between EXOSC5 and EXOSC9 (153). These observations suggest that PCH1b amino acid changes in EXOSC3 could impair interactions with other RNA exosome subunits, leading to compromised RNA exosome complex integrity; however, wholesale loss of the RNA exosome seems unlikely to explain the phenotypic variability reported in PCH1b (189).

Here, we focused on defining the spectrum of RNA targets that are affected when Rrp40 function is impaired. To address this aspect of the mechanistic model (**Fig 10B**), we performed RNA-Seq analysis, providing the first dataset of candidate RNA exosome targets in brain tissue. Our brain-enriched RNA-sequencing data on *Rrp40* mutant flies shows a global increase in transcripts and specific differences in both coding and non-coding transcripts in each *Rrp40* mutant compared to control (**Fig 8B, C**). The finding that more RNAs show an increase in steady-

state levels as compared to a decrease is consistent with impairing the function of a complex required for RNA turnover. Our results show significant correlation between the transcriptomic changes in *Rrp40*^{G11A} flies and *Rrp40*^{D87A} flies (**Fig 8D**), but also some clear differences both in the identity of the transcripts effected (**S4 Fig**) and the magnitude of the effects (**Fig 9**). The specific transcripts altered, the magnitude of the changes in the same transcripts or sets of transcripts, or a combination of these effects could underlie the phenotypic differences in the *Rrp40* mutant flies analyzed. Future studies, including additional models, will be required to define how changes in the transcriptome contribute to these phenotypic differences.

Levels of a critical synaptic regulator, *Arc1*, are markedly increased in both *Rrp40* mutants, suggesting that the RNA exosome regulates key neuronal transcripts. Notably, *Arc1* levels increase more in the *Rrp40*^{G11A} flies compared to *Rrp40*^{D87A}, consistent with a model where the magnitude of changes in key regulators could contribute to phenotypic variation. Critically, these data provide insight into how amino acid changes in Rrp40 could disrupt processing and/or decay of transcripts required for proper development/maintenance of the nervous system.

In this study, we modeled three patient genotypes associated with PCH1b (**Fig 3B**). The analysis reveals a range of phenotypes, suggesting a genotype-phenotype correlation. Furthermore, the spectrum of phenotypes observed in flies modeling alleles of PCH1b aligns with some aspects of genotype-phenotype correlations reported for individuals with PCH1b, albeit the clinical phenotype is highly variable (181, 189), suggesting that these amino acid changes differentially impact RNA exosome activity. For example, homozygous PCH1b *EXOSC3* p.Gly31Ala mutations (151, 164, 181), modeled as *Drosophila Rrp40*^{G11A}, cause severe phenotypes (**Fig 4**). In contrast, homozygous PCH1b *EXOSC3* p.Asp132Ala mutations, modeled as *Drosophila Rrp40*^{D87A}, cause more moderate phenotypes as compared to the *Rrp40*^{G11A} mutant flies. Moreover, PCH1b *EXOSC3* p.Asp132Ala in trans to a presumptive *EXOSC3* null mutation, modeled as *Drosophila*

Rrp40$^{D87A/Df(2L)}$ causes the most severe patient phenotypes (151), arguing that *D132A* is likely hypomorphic. Interestingly, *EXOSC3* p.Gly31Ala in trans to a presumptive *EXOSC3* null has not been reported, and the lethality of the corresponding genotype *Rrp40*$^{G11A/Df(2L)}$ in flies (**Fig 3B**) suggests that human G31A and *Drosophila* G11A are strong loss of function alleles. While the fly models can provide insight into genotype-phenotype correlation, it is important to note that pathology and/or disease severity could arise from a combination of effects, and these combined effects could be specific to the genetic background of the patients. Critically, the number of reported individuals with *EXOSC3*-related PCH1b is small (151, 164, 179-181, 190), thus limiting the ability to make genotype-phenotype correlations in patients and highlighting the value of the fly model.

In summary, this study identifies an essential role for Rrp40 in early development in *Drosophila* together with an age-dependent requirement for Rrp40 in neurons. Using CRISPR/Cas9 gene editing technology and *Drosophila* genetic approaches, we modeled PCH1b-linked *EXOSC3* amino acid changes in Rrp40, which reveals neuronal-specific phenotypes and defines RNA targets critical for neuronal function. Strikingly, both behavioral phenotypes of *Rrp40* mutant flies and levels of RNA targets are rescued by pan-neuronal transgenic expression of human *EXOSC3*. Taken together, these results demonstrate that this model can be utilized to explore the broader consequences of amino acid changes in the RNA exosome.

3.6 Figures

Figure 3—1

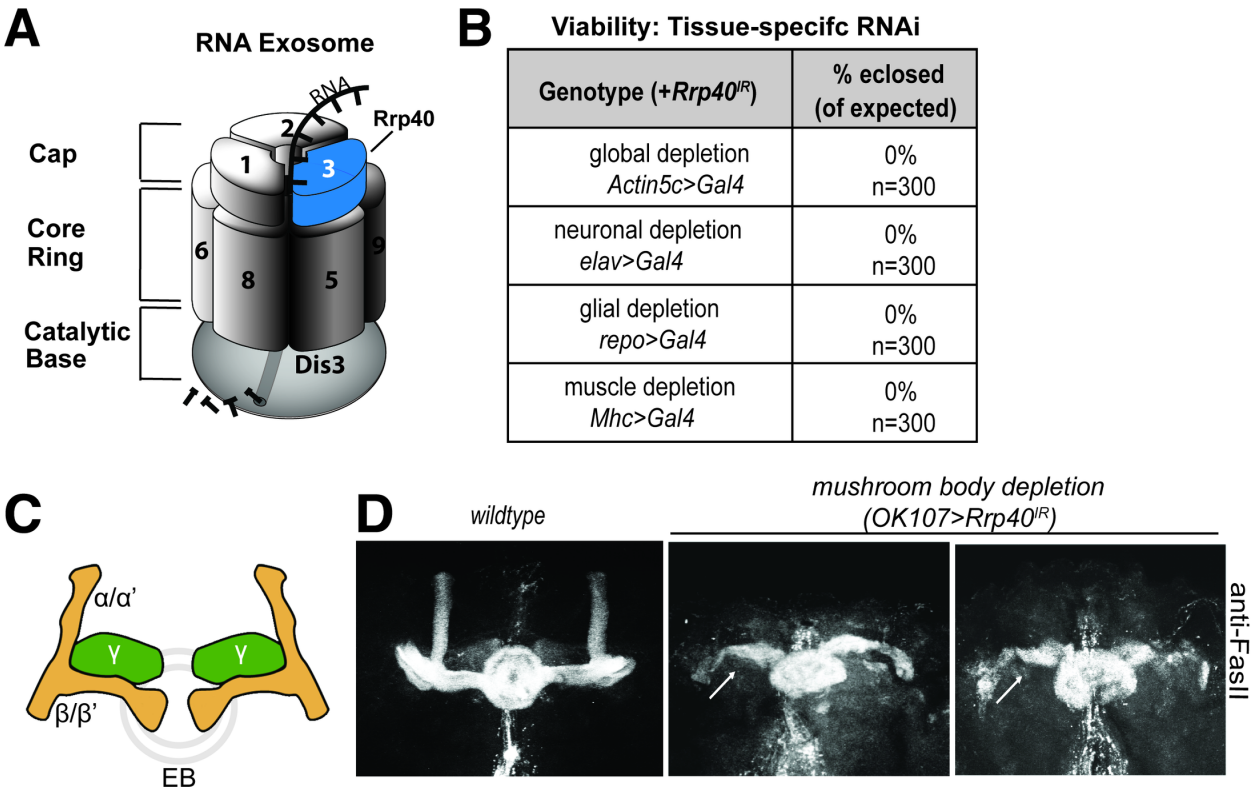
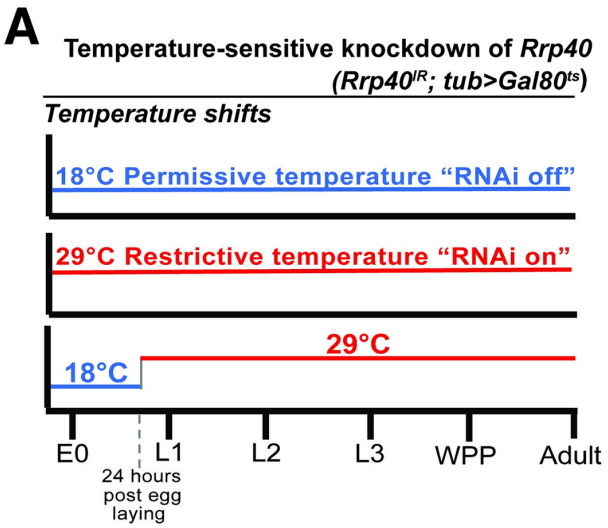


Figure 3—1: The Rrp40 subunit of the RNA exosome is required for viability and proper mushroom body development in Drosophila.

(A) The RNA exosome is an evolutionarily conserved ribonuclease complex composed of nine structural subunits termed exosome components (EXOSC1-9)—three Cap subunits: EXOSC1; EXOSC2; EXOSC3, six Core Ring subunits: EXOSC4 (not visible); EXOSC5; EXOSC6; EXOSC7 (not visible); EXOSC8; EXOSC9, and a Catalytic Base subunit: Dis3. Mutations in the EXOSC3 gene encoding a cap subunit [light blue, labeled with 3 (termed Rrp40 in *Drosophila*)] cause PCH1b (151). **(B)** Tissue-specific drivers as indicated were employed to knockdown Rrp40 in Drosophila. The percentage of flies eclosed (of expected) for the indicated genotypes is shown for a total of 300 flies for each genotype. **(C)** Diagram of the adult *Drosophila* mushroom body (MB) lobes depicting the vertical alpha (α) and alpha prime (α') neurons, the medially projecting (β) and beta-prime (β') neurons, and the γ neurons (green) as well as the ellipsoid body (EB) (gray ring). **(D)** Fasciclin II antibody (anti-FasII) was used to stain either control brains or brains with Rrp40 depleted from mushroom bodies (*OK107>Gal4*). Maximum intensity Z-stack projections of mushroom bodies are shown. β-lobes of control brains do not cross the mid-line and these brains have well-formed α-lobes. Rrp40-depleted mushroom bodies (*OK107-Gal4>Rrp40^{IR}*) (n=30) have β-lobes defects (white arrows) or have missing α- lobes.

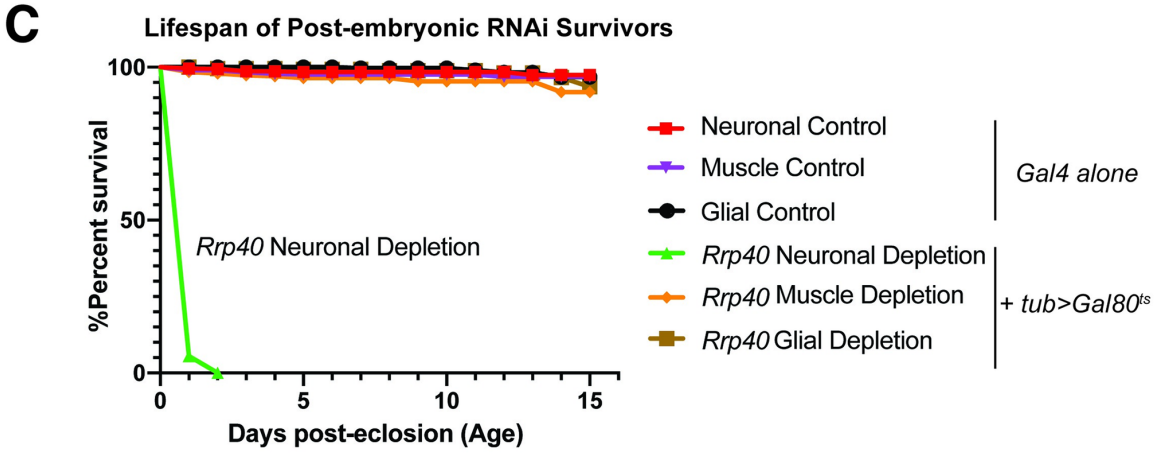
Figure 3—2



B

Frequency of Post-embryonic RNAi Survivors

Genotype (+*Rrp40*^{IR}; *tub>Gal80*^{ts})	% eclosed (of expected)
global depletion *Actin5c>Gal4*	0% n=300
neuronal depletion *elav>Gal4*	6% n=300
glial depletion *repo>Gal4*	1% n=300
muscle depletion *Mhc>Gal4*	63% n=300



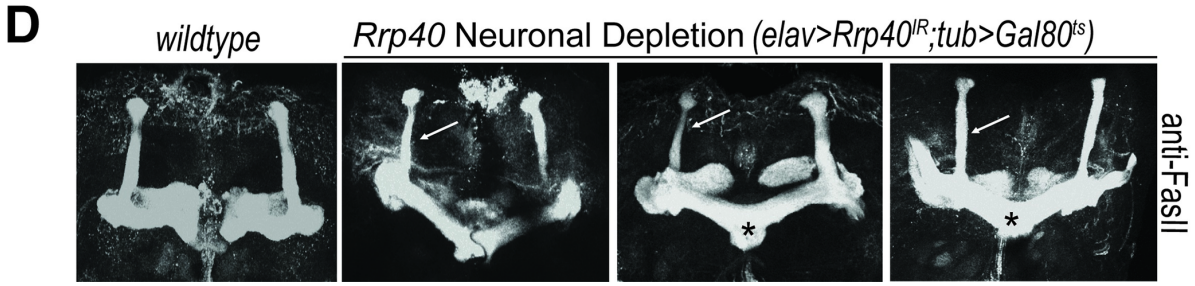
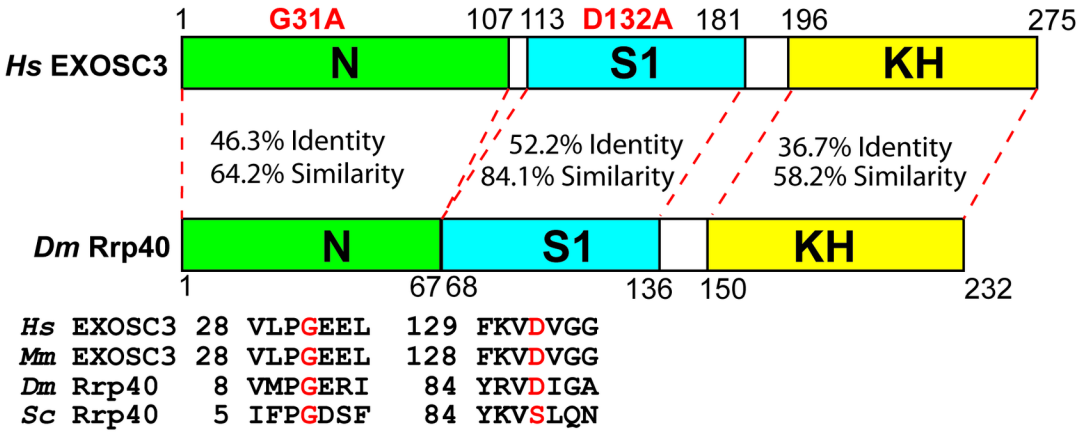


Figure 3—2: Rrp40 is required for age-dependent function in neurons.

(A) A scheme to illustrate the approach to achieve temperature-sensitive knockdown of Rrp40 is shown: The Gal80^{ts} allele shifts between 18°C (blue, RNAi off) and 29°C (red, RNAi on). E0=0hr embryo, L1/2/3=1st, 2nd or 3rd instar larvae; WPP=white prepupa. **(B)** The viability of flies in which temperature sensitive RNAi was employed to deplete Rrp40 24 hours post egg laying is shown. The viability of these post-embryonic RNAi survivors as percentage of flies eclosed (of expected) is shown for the indicated tissue-specific drivers. **(C)** Kaplan-Meier analysis of post-embryonic RNAi survivors for either control (*Gal4* alone) or temperature sensitive knockdown (*tub> Gal80^{ts}*) for the following tissue-specific drivers: Neuronal (elav>Gal4, n=50); Muscle (Mhc>Gal4, n=50); or Glial (repo>Gal4, n=50). **(D)** Fasciclin II antibody (anti-FasII) staining of both wildtype control brain and three representative brains with temperature-sensitive neuronal depletion of Rrp40 (*elav>Rrp40^{IR};tub>Gal80^{ts}*) are shown. Maximum intensity Z-stack projections of mushroom bodies are shown. β-lobes in control wildtype brains do not cross the mid-line and these brains have well-formed α and β-lobes (n=30). Mushroom bodies from flies with Rrp40 depleted from neurons (*elav>Rrp40^{IR};tub>Gal80^{ts}*) (n=5) have thinned α-lobes (white arrows) and β-lobes that often project to the contralateral hemisphere and appear to fuse (black asterisks).

Figure 3—3

A



B

EXOSC3	Rrp40	Patient Genotype	*Drosophila* Genotype	% eclosed (of expected)
G31A	*G11A*	*G31A* Homozygous	*G11A* Homozygous	14% n=126
		ND	*G11A/Df(2L)* Heterozygous	0% n=70
D132A	*D87A*	*D132A* Homozygous	*D87A* Homozygous	38% n=182
		D132A/null Heterozygous	*D87A/Df(2L)* Heterozygous	15% n=160

C

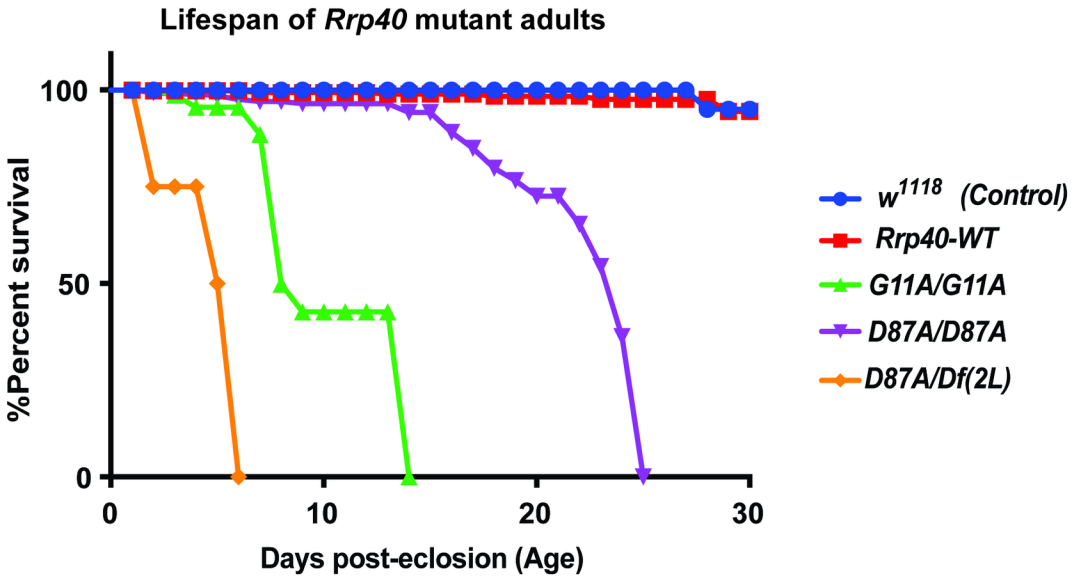


Figure 3—3: Generation of *Drosophila* models of PCH1b amino acid substitutions.

(A) The EXOSC3/Rrp40 protein consists of three domains: an N-terminal domain (N), a central S1 putative RNA binding domain, and a C-terminal putative RNA binding KH (K homology) domain. The disease-causing amino acid substitutions (G31A and D132A) modeled in *Drosophila* in this study are indicated in red above the domain structure. The position and flanking sequence of the amino acid substitutions in PCH1b-associated human EXOSC3 (shown in red) and the corresponding amino acids in Mus musculus (Mm), *Drosophila* melanogaster (Dm), and Saccharomyces cerevisiae (Sc) are shown. The percent identity/similarity between the human and *Drosophila* domains is indicated. **(B)** The amino acid substitutions that occur in EXOSC3 in patients that are modeled in *Drosophila* Rrp40, are indicated with Patient Genotype (Homozygous/Heterozygous) as modeled in *Drosophila* (*Drosophila* Genotype) shown. For each *Drosophila* model, the viability of that genotype engineered in flies by CRISPR/Cas9 gene editing is shown as % eclosed (of expected). Rrp40 mutants show decreased viability, indicated by skewed Mendelian ratios. n= the number of individual flies analyzed. The G11A over null genotype has not been reported in patients-indicated as not detected (ND) under patient genotype. **(C)** A Kaplan-Meier analysis shows that Rrp40 mutant flies with the indicated genotypes to model patient genotypes exhibit early adult lethality compared to wildtype Control flies (w^{1118}) or Rrp40-WT ($Rrp40^{wt}$) flies that have undergone CRISPR/Cas9 gene editing to produce a control wildtype Rrp40 allele (see Materials and methods).

Figure 3—4

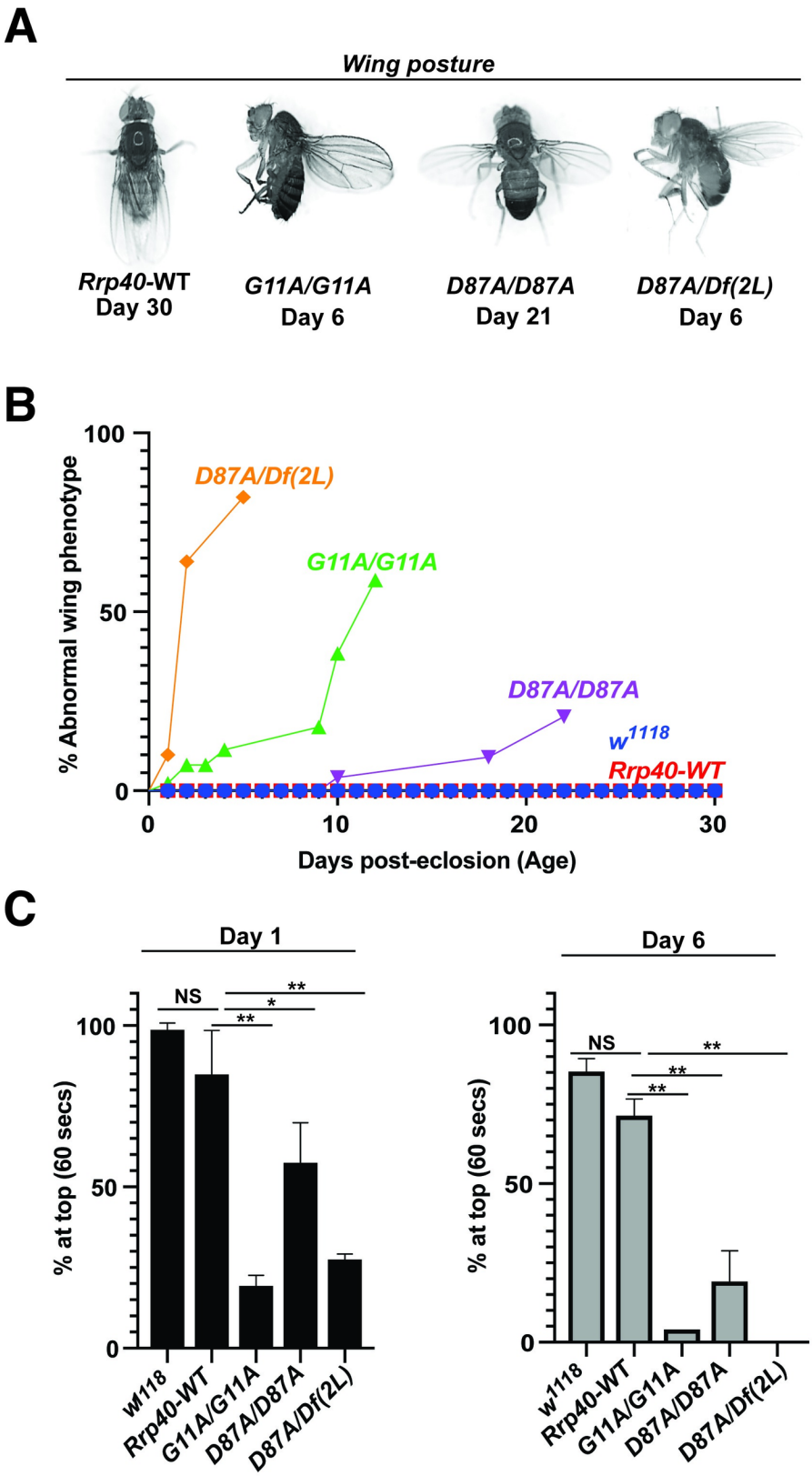

A
Wing posture
Rrp40-WT
Day 30
G11A/G11A
Day 6
D87A/D87A
Day 21
D87A/Df(2L)
Day 6
B
% Abnormal wing phenotype
100
50
0
0
10
20
30
Days post-eclosion (Age)
D87A/Df(2L)
G11A/G11A
D87A/D87A
w1118
Rrp40-WT
C
Day 1
Day 6
% at top (60 secs)
NS
*
**
w1118
Rrp40-WT
G11A/G11A
D87A/D87A
D87A/Df(2L)

Figure 3—4: Flies that model PCH1b amino acid substitutions in *Drosophila* show a range of morphological and behavioral phenotypes.

(A) *Rrp40* mutant flies show a progressive wing position phenotype in adults. Abnormal wing posture phenotypes (wings held-up or -out) are shown for *Rrp40* mutant flies compared to *Rrp40-WT* flies with Days post-eclosion indicated for each representative image. **(B)** Quantitative analysis of the abnormal wing position phenotype of *Rrp40* mutants raised at 25°C. The percentages of adults with abnormal wing position over time (Days post-eclosion) were calculated based on 60 flies per genotype. Neither w^{1118} nor Control *Rrp40-WT* flies showed any abnormal wing position (0% abnormal wing posture, n>100) over the time course analyzed. **(C)** Locomotor activity was assessed using a negative geotaxis assay as described in Materials and methods. Data are presented as the average percentage of flies of the indicated genotypes that reach the top of a cylinder after 60 seconds across all trials. Groups of 10 age-matched flies [Day (1) and Day (6)] were tested for at least three independent trials per genotype. Values represent the mean ± SEM for n= 3 independent experiments. Asterisks (*) indicate results that are statistically significance at *p-value < 0.05; **p<0.01. Results that show no statistical significance when compared are indicated by NS.

Figure 3—5

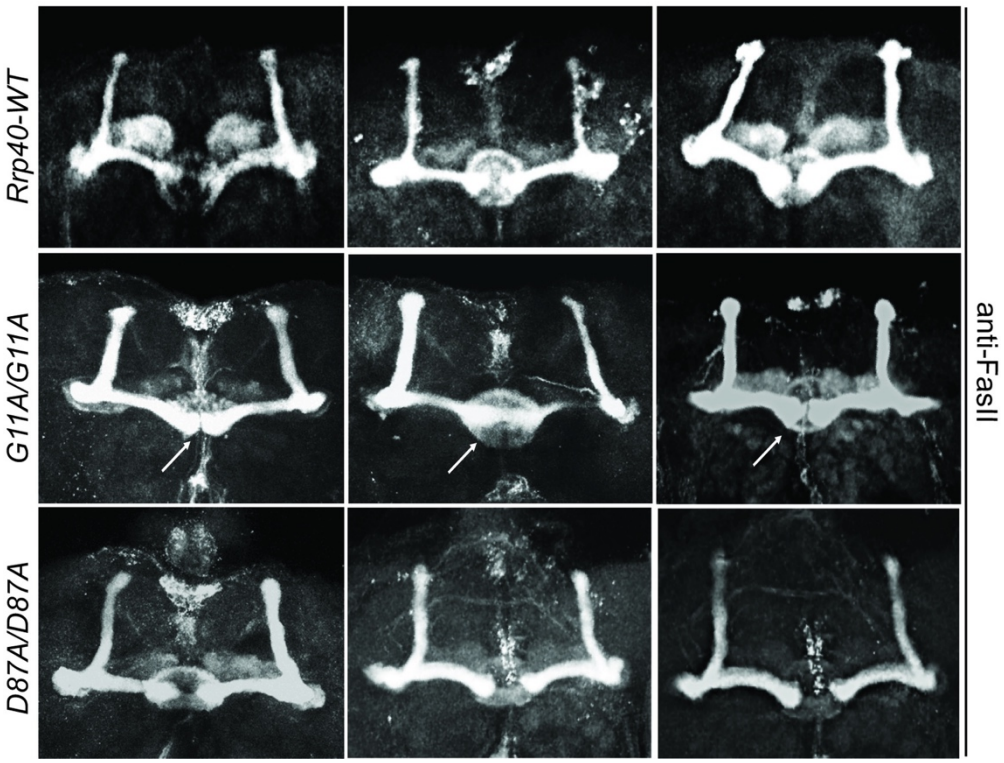


Figure 3—5: *Rrp40^{G11A}* mutant flies show defects in mushroom body morphology.

Three representative images of merged projections of FasII-stained mushroom bodies from *Rrp40-WT* control flies (n=18) (top row), *Rrp40^{G11A}* (*G11A/G11A*) mutant flies (n=23) (middle row), and *Rrp40^{D87A}* (*D87A/D87A*) mutant flies (n=18) (bottom row) are shown. In control flies, β-lobes do not cross the mid-line and well-formed α and β-lobes are readily apparent. In *Rrp40^{G11A}* mutant flies, 13/23 (57%) brains examined showed defects in β-lobe structure or β-lobe fusion (white arrows). Neither β-lobe defect was detected in the Rrp40^{D87A} mutant flies.

Figure 3—6

A

Model	Genotype	% eclosed (of expected)
WT	*Rrp40^{wt}/Rrp40^{wt}*	93% n=100
G11A	*Rrp40^{G11A}/Rrp40^{G11A} UAS-EXOSC3-myc* (rescue control)	22% n=100
	Rrp40^{G11A}/Rrp40^{G11A} elav-Gal4;UAS-EXOSC3-myc (pan-neuronal rescue)	20% n=100
	Rrp40^{G11A}/Rrp40^{G11A} Mhc-Gal4;UAS-EXOSC3-myc (muscle rescue)	16% n=100
D87A	*Rrp40^{D87A}/Rrp40^{D87A} UAS-EXOSC3-myc* (rescue control)	37% n=100
	Rrp40^{D87A}/Rrp40^{D87A} elav-Gal4;UAS-EXOSC3-myc (pan-neuronal rescue)	40% n=124
	Rrp40^{D87A}/Rrp40^{D87A} Mhc-Gal4;UAS-EXOSC3-myc (muscle rescue)	43% n=75

B

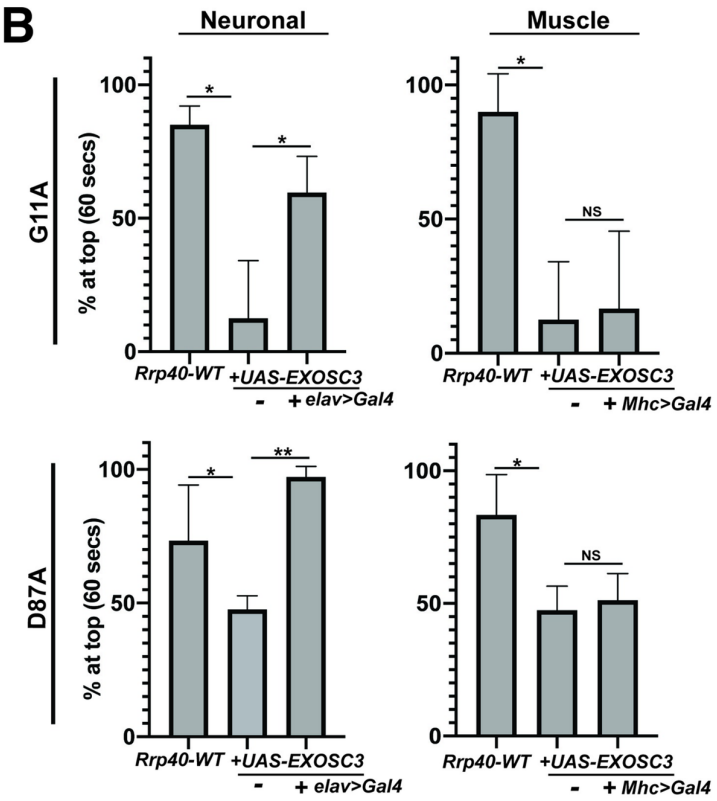


Figure 3—6: Locomotor defects in Rrp40 mutant flies are rescued by neuronal expression of human EXOSC3.

(A) Transgenic flies that express myc-tagged human *EXOSC3* under a UAS were generated as described in Materials and methods. The percentages of flies eclosed (of expected) for the indicated genotypes which include flies with transgenic *EXOSC3*, but no driver (rescue control), flies with the UAS and a pan-neuronal driver (*elav*) (pan-neuronal rescue), or the UAS and a muscle driver (*Mhc*) are shown for each *Rrp40* mutant. **(B)** Locomotor assays were employed to assess whether expression of human *EXOSC3* rescues behavioral phenotypes in *Rrp40* mutant flies (*G11A/G11A* or *D87A/D87A*) in either neurons (*elav*) or muscle (*Mhc*). Data are presented as the average percentage of flies that reach the top of a cylinder after 60 seconds across all trials. Groups of 10 age-matched flies (Day 6) were tested for at least three independent trials per genotype. Results are shown for control *Rrp40-WT* flies as compared to either $Rrp40^{G11A}$ (*G11A/G11A*, top) or $Rrp40^{D87A}$ (*D87A/D87A*, bottom) homozygous flies that either do not express human *EXOSC3* (-) or do express human *EXOSC3* (+*Gal4*) in either neurons (*elav*) (Neuronal, left) or muscle (*Mhc*) (Muscle, right). Values represent the mean ± SEM for n= 3 independent experiments. Asterisks (*) indicate results that show statistical significance at *p-value < 0.05; **p<0.01. Results that show no statistical significance when compared are indicated by NS.

Figure 3—7

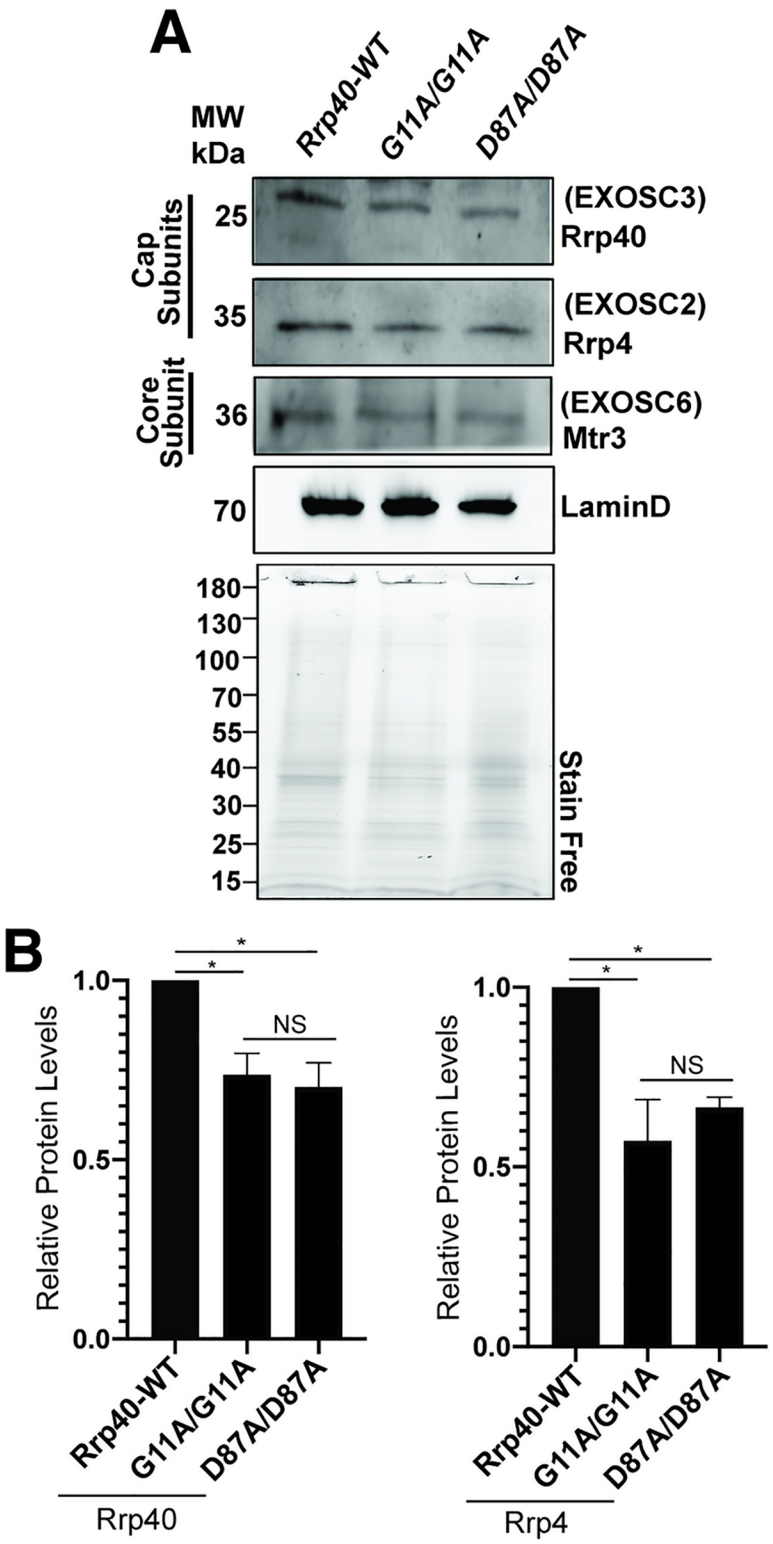

A
MW
kDa
Rrp40-WT
G11A/G11A
D87A/D87A
Cap Subunits
25
(EXOSC3)
Rrp40
35
(EXOSC2)
Rrp4
Core Subunit
36
(EXOSC6)
Mtr3
70
LaminD
180
130
100
70
55
40
30
25
15
Stain Free
B
Relative Protein Levels
1.0
0.5
0.0
*
*
NS
Rrp40-WT
G11A/G11A
D87A/D87A
Rrp40
Relative Protein Levels
1.0
0.5
0.0
*
*
NS
Rrp40-WT
G11A/G11A
D87A/D87A
Rrp4

Figure 3—7: Amino acid substitutions that model PCH1b in Rrp40 alter levels of RNA exosome subunits.

(A) Lysates prepared from heads of control *Rrp40*^{wt} or *Rrp40* mutant flies were resolved by SDS-PAGE and analyzed by immunoblotting with antibodies to detect RNA exosome Cap Subunits, Rrp40/EXOSC3 and Rrp4/EXOSC2, and Core Subunit Mtr3/EXOSC6. Both LaminD and Stain Free, as a measure of total protein, serve as loading controls. **(B)** Results from **(A)** were quantitated for RNA exosome cap subunits levels (Rrp40 and Rrp4- bands for Mtr3 had too much background to yield reproducible results) and are presented as Relative Protein Levels with the value from the control Rrp40-WT flies set to 1.0. Asterisks (*) indicate results that show statistical significance at *p-value< 0.05. Results that show no statistical significance when compared are indicated by NS.

Figure 3—8

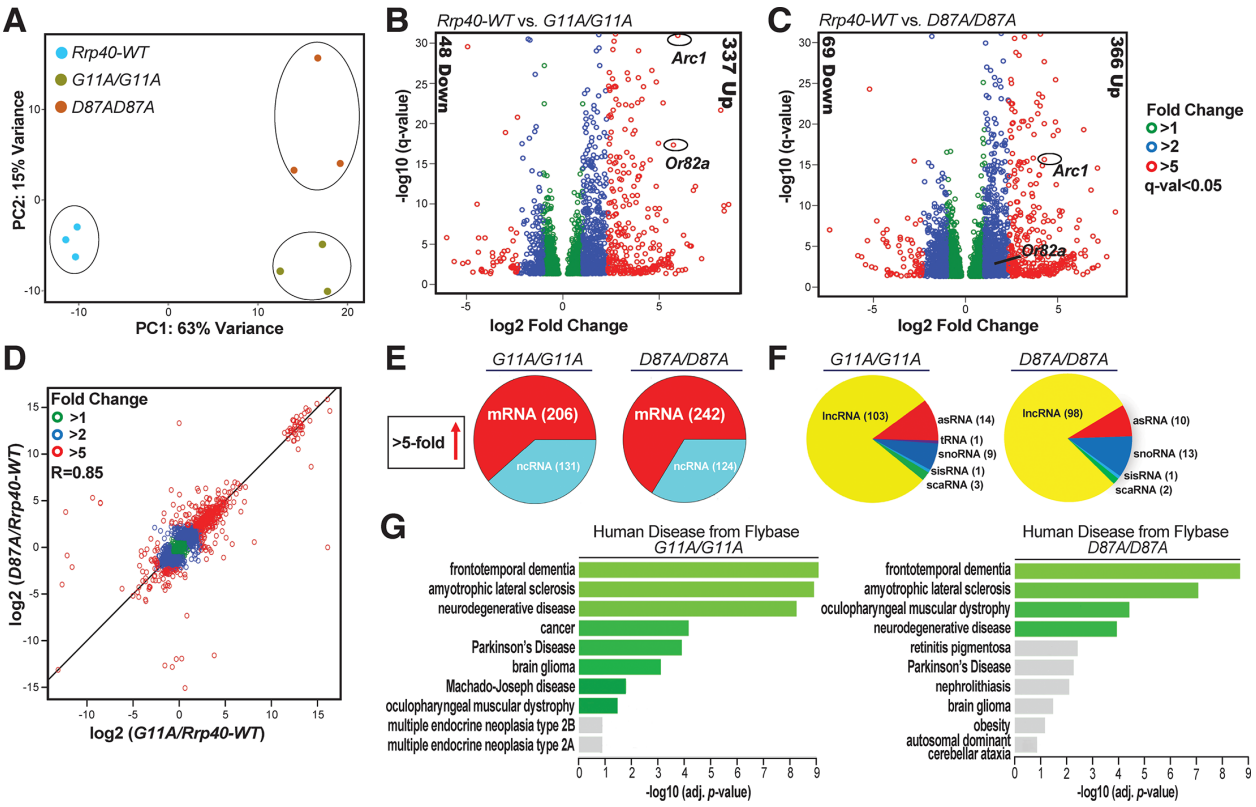

A
Rrp40-WT
G11A/G11A
D87AD87A
PC2: 15% Variance
PC1: 63% Variance
B
Rrp40-WT vs. G11A/G11A
48 Down
337 Up
Arc1
Or82a
-log10 (q-value)
log2 Fold Change
C
Rrp40-WT vs. D87A/D87A
69 Down
366 Up
Arc1
Or82a
-log10 (q-value)
log2 Fold Change
Fold Change
>1
>2
>5
q-val<0.05
D
Fold Change
>1
>2
>5
R=0.85
log2 (D87A/Rrp40-WT)
log2 (G11A/Rrp40-WT)
E
G11A/G11A
D87A/D87A
>5-fold
mRNA (206)
ncRNA (131)
mRNA (242)
ncRNA (124)
F
G11A/G11A
D87A/D87A
lncRNA (103)
asRNA (14)
tRNA (1)
snoRNA (9)
sisRNA (1)
scaRNA (3)
lncRNA (98)
asRNA (10)
snoRNA (13)
sisRNA (1)
scaRNA (2)
G
Human Disease from Flybase
G11A/G11A
frontotemporal dementia
amyotrophic lateral sclerosis
neurodegenerative disease
cancer
Parkinson's Disease
brain glioma
Machado-Joseph disease
oculopharyngeal muscular dystrophy
multiple endocrine neoplasia type 2B
multiple endocrine neoplasia type 2A
-log10 (adj. p-value)
Human Disease from Flybase
D87A/D87A
frontotemporal dementia
amyotrophic lateral sclerosis
oculopharyngeal muscular dystrophy
neurodegenerative disease
retinitis pigmentosa
Parkinson's Disease
nephrolithiasis
brain glioma
obesity
autosomal dominant cerebellar ataxia
-log10 (adj. p-value)

Figure 3—8: RNA-seq analysis of *Rrp40* mutant flies identifies key neuronal targets that are altered.

(A) Principal component analysis (PCA) analysis of RNA-seq data from adult *Drosophila* heads from *Rrp40-WT* and *Rrp40* mutant flies ($Rrp40^{G11A}$ and $Rrp40^{D87A}$) shows that results from the three independent experiments cluster based on the genotype and that mutants are more distinct from control flies than from one another. **(B,C)** Volcano plots show transcripts differentially expressed in each *Rrp40* mutant [**(B)** *G11A/G11A* (2,351) and **(C)** *D87A/D87A* (3,502)] compared to *Rrp40-WT* ($q<0.05$, DEseq). The number of transcripts that show ≥ 5-fold change in steady-state levels (Down or Up) are indicated to the side in each plot. Representative regulators of neuronal function in flies (*Arc1* and *Or82a*) are highlighted. **(D)** A correlation curve comparing the changes in gene expression relative to wildtype for each *Rrp40* allele (*G11A/G11A* and *D87A/D87A*) was produced by plotting on a logarithmic scale. This analysis shows that the changes in transcript levels are highly correlated ($R=0.85$) with respect to both magnitude and direction. **(E)** A pie chart illustrates the class of RNAs affected in each *Rrp40* mutant by showing total RNAs increased at least 5-fold in each *Rrp40* mutant corresponding to coding (mRNA) and non-coding RNA (ncRNA). **(F)** A pie chart illustrates the classes of ncRNAs increased at least 5-fold in each **Rrp40** mutant (lncRNAs, long noncoding RNAs; asRNAs, Antisense RNAs; tRNAs, transfer RNAs; snoRNAs, small nucleolar RNAs; sisRNAs, stable intronic sequence RNAs; scaRNAs, small Cajal body RNAs). **(G)** The enriched pathways from FlyEnrichr database for Human Disease from Flybase are shown for transcripts that are increased at least 1.2-fold in each *Rrp40* mutant (left, *G11A/G11A*; right, *D87A/D87A*) as compared to

Rrp40-WT. The bars shown in green correspond to significant enrichment (Adj. p-val<0.05).

Figure 3—9

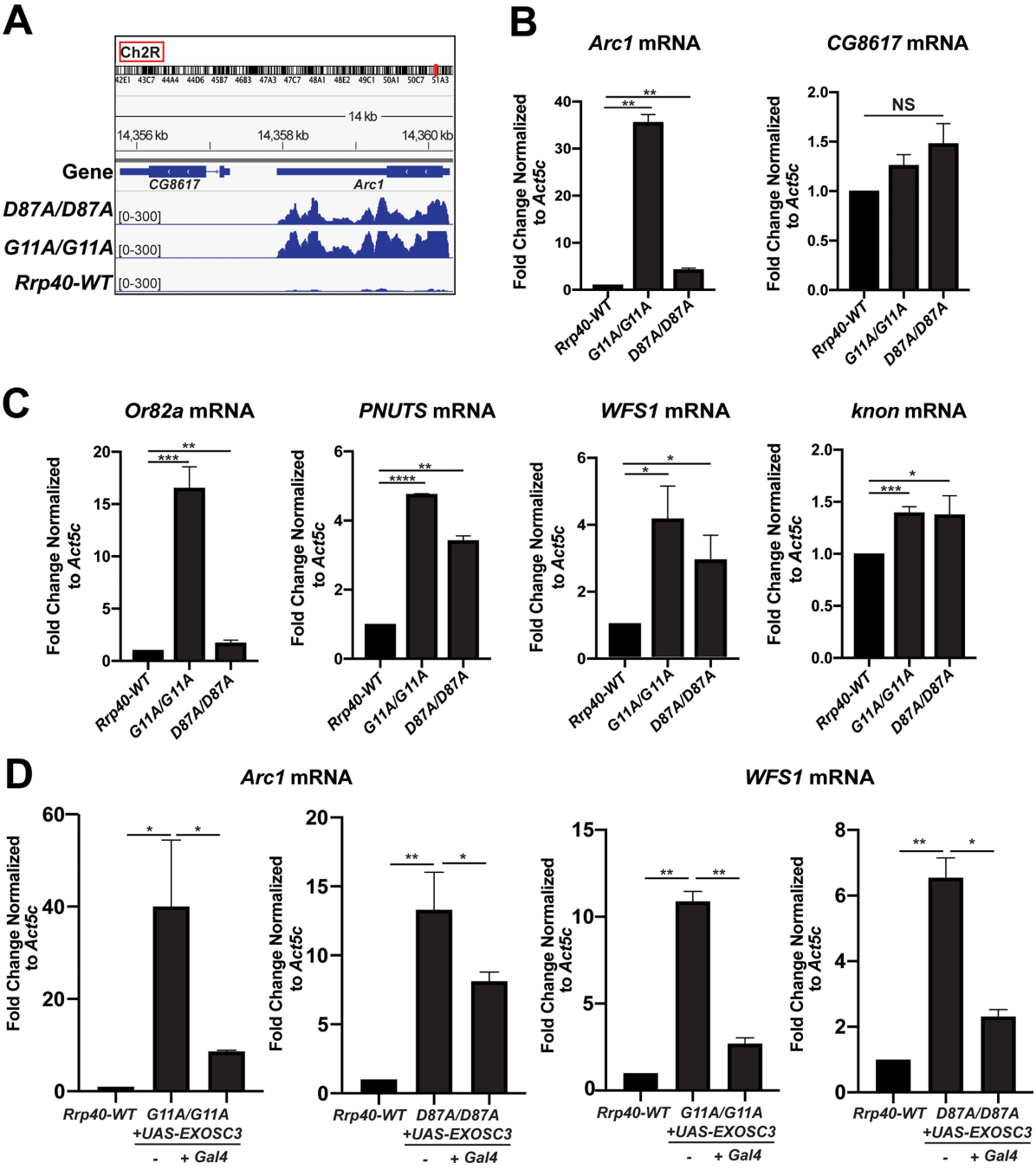


Figure 3—9: Key neuronal transcripts show an increase in steady-state levels in Rrp40 mutant flies.

(A) An Integrative Genomic View (IGV) screenshot

(http://software.broadinstitute.org/software/igv/) of *Arc1* track peaks across the *Drosophila* genome (dm6 assembly) in *Rrp40* mutants [*Rrp40 D87A/D87A* (Upper) and *Rrp40 G11A/G11A* (Middle)] as well as *Rrp40-WT* (*WT,* Lower). Chromosome number and location (red rectangle; top) are displayed at the top. (**B,C**) Changes in the level of representative transcripts were validated using RT-qPCR in *Rrp40-WT* and *Rrp40* mutant flies. Data are presented as the Fold Change Normalized to *Act5c* transcript (which was unchanged in the RNA-Seq datasets) where the value for the *Rrp40-WT* sample is set to 1.0. Results are presented as mean ± standard error of the mean (*p<0.05;**p<0.01;***p<0.001;****p<0.0001 vs. the *WT* group in a two-tailed t-test). **(D)** Expression of human *EXOSC3* (*UAS-EXOSC3-myc*) in neurons rescues the increased steady-state levels of the Arc1 (left) and *WFS1* (right) transcripts in both *Rrp40* mutants (*G11A/G11A* or *D87A/D87A*). The steady-state levels of the *Arc1* and *WFS1* transcripts were analyzed using RT-qPCR in control *Rrp40-WT* flies (set to 1.0) and mutant flies expressing human *EXOSC3* (+*Gal4*) or not (-). Data are presented as the means ± standard error of the mean (*p<0.05;**p<0.01 vs. the *WT* group in a two-tailed t-test).

Figure 3—10

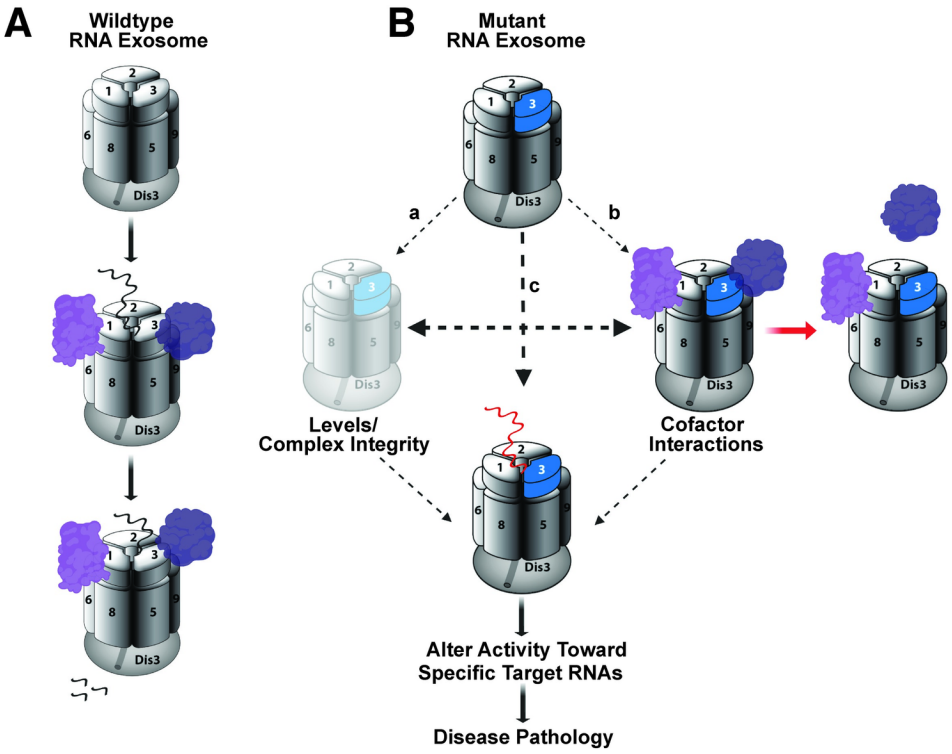

A
Wildtype
RNA Exosome
B
Mutant
RNA Exosome
a
b
c
Levels/
Complex Integrity
Cofactor
Interactions
Alter Activity Toward
Specific Target RNAs
Disease Pathology
Dis3

Figure 3—10: Mechanistic model for how amino acid substitutions could alter RNA exosome function.

(A) The Wildtype RNA exosome (top) associates with a variety of cofactors (purple) that can aid in target RNA recognition (middle) and decay/processing (bottom). **(B)** Amino acid substitutions that occur in the EXOSC3/Rrp40 cap subunit (shown in blue) could impact the function of the complex through multiple mechanisms, including: (a) decrease the steady-state level of the specific subunit or alter interaction of the subunit with the complex, leading to a decrease in overall Levels or altered Complex Integrity; (b) affect Cofactor Interactions, which could impact interactions with target RNAs; and/or (c) could directly alter interactions with target RNAs (red RNA indicates impaired interaction). Ultimately, any of these changes could Alter Activity Toward Specific Target RNAs. Changes in the steady state level or processing of RNAs could then cause downstream Disease Pathology.

3.7 Supplementary Figures

Figure 3—S1

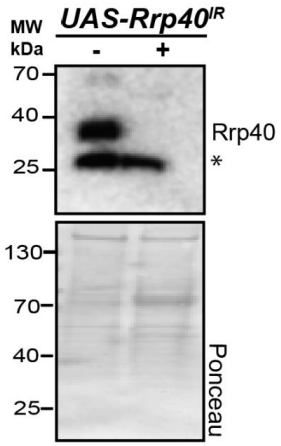


Figure 3—S1: Validation of Rrp40 RNAi.

Immunoblot of protein extracts from heads of adult flies raised at 18°C expressing *elav-Gal4* (neurons) in combination with a RNAi against *Rrp40* (*elav-Gal4>UAS-Rrp40^{IR}*) and control (*elav>Gal4*), probed with α-Rrp40 antibody. The α-Rrp40 antibody detects Rrp40 at the predicted size (25KDa) and this band is efficiently depleted in flies the express the UAS-RNA. A non-specific band (*) detected by the antibody is not affected. Ponceau staining provides total protein as a loading control. Flies were reared at lower temperatures (18°C) to recover viability.

Figure 3—S2

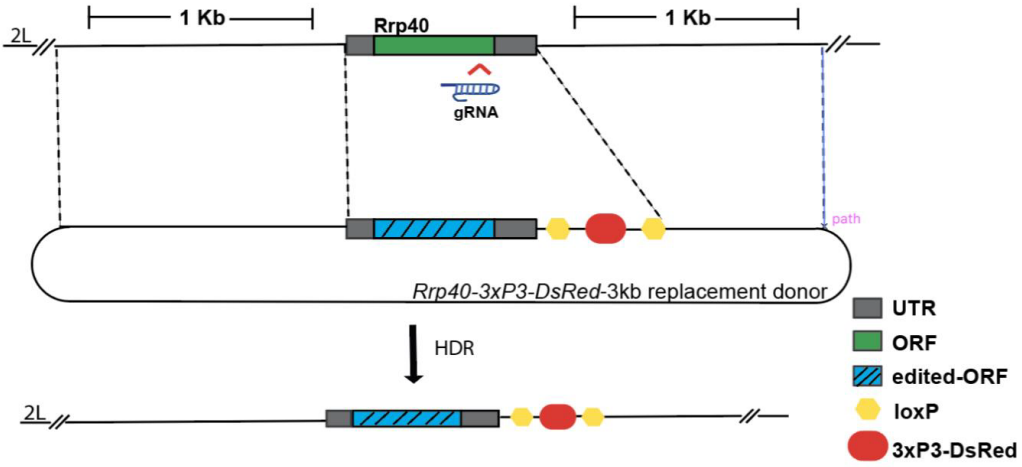

1 Kb
Rrp40
1 Kb
2L
gRNA
path
Rrp40-3xP3-DsRed-3kb replacement donor
HDR
2L
UTR
ORF
edited-ORF
loxP
3xP3-DsRed

Figure 3—S2: CRISPR/Cas9-induced Homology-directed recombination (HDR) with a double-stranded DNA donor vector. A Schematic of the HDR strategy used to replace the *Rrp40* locus with the *Rrp40-3xP3-DsRed* vector, which drives expression of DsRed in the eye. The DsRed is flanked by loxP recombination sites to remove DsRed if desired. The guide RNA (gRNA) target site is indicated (red arrow). Homology arms of 1kb immediately flanking *Rrp40* were cloned into the *Rrp40-3xP3-DsRed* vector.

Figure 3—S3

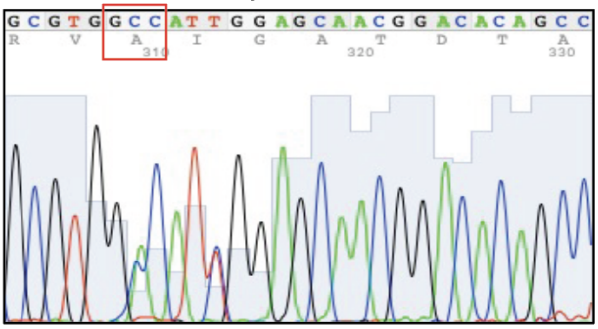
A

>D87Aseq_1_Rrp40F_A10
ATGAGCGCGACCAGCACAATAGTTATGCCCGGTGAGCGAATTGCGGCCATTGAGGAGCTGGCGAAGAGCA
AGCGGGTGATACTCGGACCGGGACTGCGGCGTCTGGATGACACGGTGGTGGCCAGCAAGGCAGGCCCACT
TCGACACAAGGAACCCGGCACCTTCTGGGTGGACAACTACCAGAGGAGGTACATCCCGGCACGCGGGGAT
CTCATTCTGGGCATTGTGCGAGCCAAGGCGGGCGATCTGTACCGCGTGGCCATTGGAGCAACGGACACAG
CCTCCATATCGTATCTCGCCTTTGAGGCGGCCACCAAGAAGAATCGCCCGGATCTGATCCCCGGTGATCT
GATTTATGCGAGAGTCCTGAATGCCAGCGCGGACATTGAACCGGAGCTGGTCTGCGTCAACTCGGTGGGC
AAAAGTGGCAAACTGGGCGTCCTCACCGATGGATTCTTCTTCAAGTGCAGCCTGAATCTGGGAAGGATGC
TGCTGCGGGAAAACTGCCCTGTTCTCGCCGCTCTTACCCGGGAACTGCCCTACGAGATCGCTGTGGGAGT
CAACGGCAGGATATGGTTGAAGGCCCATTCCCTGACCGTGGCCCTCGCCAATGCCATTTCAGCGCTGGAG
CAATCGGGATGTGCGGAAATCGACAAAATATGTGGCAATCTCGGAGACTTCCTGCAGGCCTAG

B

>G11Aseq_4_Rrp40-F_H02
GGAGTTTCAAATCGTAAAAGCAACATTAAAATGAGCGCGACCAGCACAATAGTTATGCCCGCTGAGCGAA
TTGCGGCCATTGAGGAGCTGGCGAAGAGCAAGCGGGTGATACTCGGACCGGGACTGCGGCGTCTGGATGA
CACGGTGGTGGCCAGCAAGGCAGGCCCACTTCGACACAAGGAACCCGGCACCTTCTGGGTGGACAACTAC
CAGAGGAGGTACATCCCGGCACGCGGGGATCTCATTCTGGGCATTGTGCGAGCCAAGGCGGGCGATCTGT
ACCGCGTGGACATCGGAGCAACGGACACAGCCTCCATATCGTATCTCGCCTTTGAGGCGGCCAGCAAGAA
GAATCGCCCGGATCTGATCCCCGGTGATCTGATTTACGCGAGAGTCCTGAATGCAAGCGCGGACATTGAA
CCGGAGCTGGTCTGCGTCAACTCGGTGGGCAAAAGTGGCAAACTGGGCGTCCTCACCGATGGATTCTTCT
TCAAGTGCAGCCTGAATCTGGGAAGGATGCTGCTGCGGGAAAACTGCCCTGTTCTCGCCGCTCTTACCCG
GGAACTGCCCTACGAGATCGCTGTGGGAGTCAACGGCAGGATATGGTTGAAGGCCCATTCCCTGA

C



D

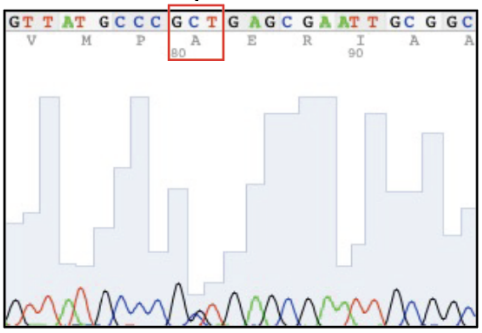


Figure 3—S3: *EXOSC3*-PCH1b mutations modeled in *Drosophila Rrp40*.

(A,B) Sequences confirming the creation of CRISPR/Cas9 genome-edited flies **(A)** *Rrp40*^{D87A} and **(B)** *Rrp40*^{G11A} isolates are shown. The engineered mutations in the *Rrp40* locus are highlighted in red and are noted along with sequencing traces of **(C)** *Rrp40*^{D87A} and **(D)** *Rrp40*^{G11A} confirming the successful generation of the edited alleles.

Figure 3—S4

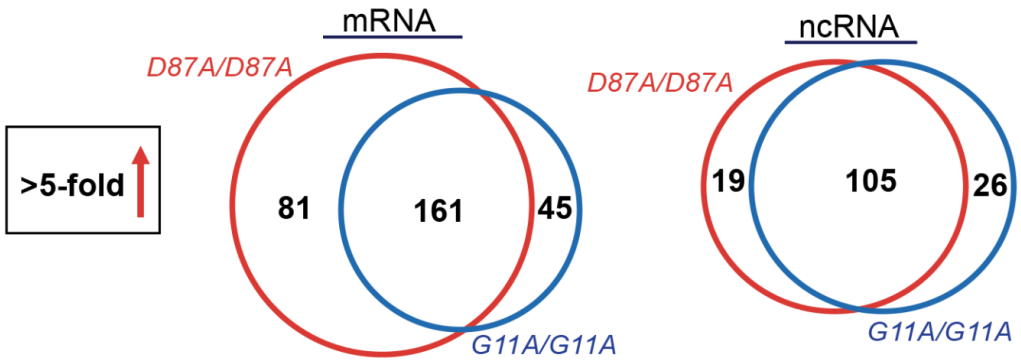

mRNA
D87A/D87A
>5-fold
81
161
45
G11A/G11A
ncRNA
D87A/D87A
19
105
26
G11A/G11A

Figure 3—S4: *Rrp40*^{G11A} and *Rrp40*^{D87A} mutant flies show overlapping or distinct sets of RNAs.

A Venn diagram illustrates the fraction of protein-coding (mRNA) and non-coding (ncRNA) transcripts changed 5-fold or more that are common to the *Rrp40* mutant flies.

Figure 3—S5

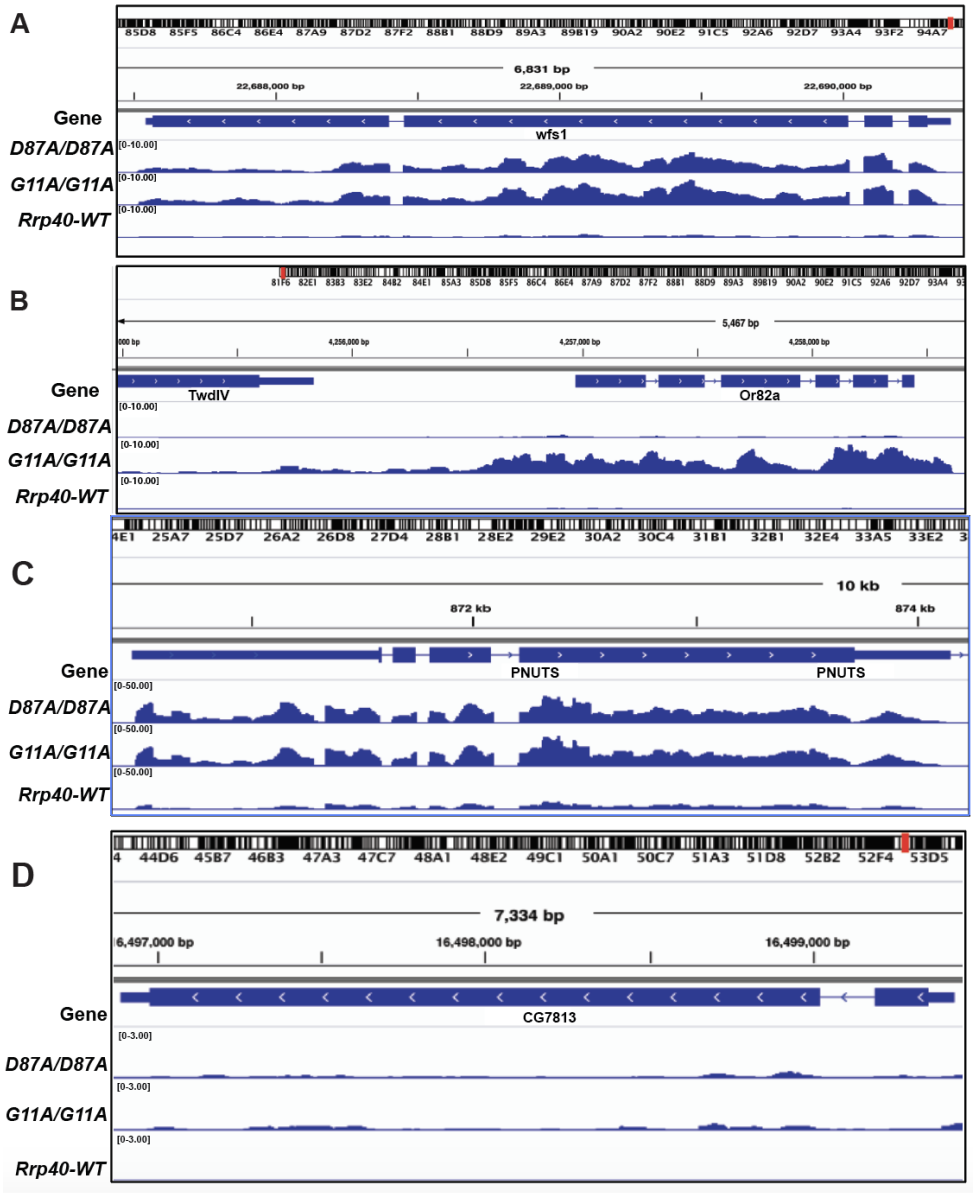

A
6,831 bp
22,688,000 bp
22,689,000 bp
22,690,000 bp
Gene
wfs1
D87A/D87A
G11A/G11A
Rrp40-WT
[0-10.00]
B
5,467 bp
4,256,000 bp
4,257,000 bp
4,258,000 bp
Gene
TwdlV
Or82a
D87A/D87A
G11A/G11A
Rrp40-WT
[0-10.00]
C
10 kb
872 kb
874 kb
Gene
PNUTS
PNUTS
D87A/D87A
G11A/G11A
Rrp40-WT
[0-50.00]
D
7,334 bp
6,497,000 bp
16,498,000 bp
16,499,000 bp
Gene
CG7813
D87A/D87A
G11A/G11A
Rrp40-WT
[0-3.00]

Figure 3—S5: Integrative Genomic View (IGV) screenshots of multiple functionally important neuronal to illustrate the RNA-Seq data obtained.

Integrative Genomics Viewer (IGV) screenshot (http://software.broadinstitute.org/software/igv/) of *WFS1* **(A)**, *Or82a* **(B)**, *PNUTS* (C) and *knon (CG7813)* track peaks across the *Drosophila* genome (dm6 assembly) in *Rrp40* mutants [$Rrp40^{G11A}$ (Middle) and $Rrp40^{D87A}$ (Upper)] as well as $Rrp40^{wt}$ (WT, lower).

Table 3—S1

Primer Sets used in RT-qPCR

Primer Name	Primer Sequence
Arc1	FWD: GCCACAAGCACAGCAATAG REV: ATTTTTTAGCAGTTCCACGCAAC
Or82a	FWD: GCCACAAGCACAGCAATAG REV: ATTTTTAGCAGTTCCACGCAAC
WFS1	FWD: GCTCTCACATTATCCGTCCC REV: ACACAATCAAGTTAAGCAGCC
PNUTS	FWD: GACGTACCAGTCCAAATTCAG REV: GCCATTAAAGCGGGTGAAG
knon	FWD: ACCGCGTTCGTAAGTATCC REV: TCACCACCTTCTTTTCTTCATC
Act5c	FWD: GAGCGCGGTTACTCTTTCAC REV: ACTTCTCCAACGAGGAGCTG

S1 Table: Oligonucleotide primer sequences employed for RT-qPCR.

Table 3—S2

https://doi.org/10.1371/journal.pgen.1008901.s007

S2 Table: RNA-Seq data analysis of *Rrp40*^{G11A} alleles.

Table 3—S3

https://doi.org/10.1371/journal.pgen.1008901.s008

S3 Table: RNA-Seq data analysis of *Rrp40*^{D87A} alleles.

Video 3—S1

S1 Video: Negative geotaxis assay for *Rrp40*^{wt} at Day 4.

https://doi.org/10.1371/journal.pgen.1008901.s009

S2 Video: Negative geotaxis assay for *Rrp40*^{G11A} at Day 4.

https://doi.org/10.1371/journal.pgen.1008901.s010

S2 Video: Negative geotaxis assay for *Rrp40*^{D87A} at Day 4.

https://doi.org/10.1371/journal.pgen.1008901.s011

Chapter 4

Discussion

4.1 New Insights Into The Central Dogma of Biology

The field of molecular biology has expanded exponentially since the initial conception of the Central Dogma in 1970 (1). Advances in research technology and extensive research has made it abundantly clear that the linear transfer of genetic information from DNA to RNA to protein is not as simple as originally thought. Biological processing of these macromolecules (DNA, RNA, and proteins) underlies the complexity we see in nature and in the development of more complex organisms (i.e., humans). New discoveries are constantly emerging as researchers and scientists tackle the fundamental questions associated with the Central Dogma.

Completion of the Human Genome Mapping Project in 2000 provided a wealth of knowledge from the sequenced genome, but this achievement also raised questions that challenged the central view of the Central Dogma (191). In 2004, J.C. Biro listed the seven fundamental, unsolved questions in molecular biology; each illuminating insufficiencies of the Central Dogma (192). Three of these 'gene-centric' alternatives to the Dogma formulated in Biro's work are the assertations that 1) the flow of genetic information between nucleic acids (DNA, RNA) and proteins is bi-directional, 2) that the complex genetic information is always carried and stored by nucleic acids and proteins, and 3) that introns are the sites of biological regulation and diversity (192). These three assertations are of particular significant to this dissertation, as these statements provide some insight into the diversity of cell types, due to alternative splicing of exons and potential retention of introns, as well as the dynamic interactions between RNAs and proteins in the processes regulating gene expression in a cell or tissue-specific manner.

4.2 Brief Overview of Main Findings

In experiments presented in this dissertation, the *Drosophila* model system was employed to explore two examples where key RNA regulatory factors that are ubiquitously expressed are linked to tissue-specific human disease. Similar to the cases described for these two ubiquitously expressed proteins, monogenic mutations in a number of genes that encode ubiquitously expressed proteins, including RNA-binding proteins (RBPs), have been linked to human disease (53, 54). Several of these RBPs have been implicated in neurological disorders, indicating a special requirement for these proteins in neuronal cells, likely due to the tightly regulated processes to fine tune gene expression in neurons (193). Chapters 2 and 3 of this dissertation describe example of mutations that cause tissue-specific neurological disease that affect different regions of the brain. This presents a challenge to define the tissue-specific requirements for these proteins and their role in neurological pathology. Thus, using well-developed and genetically tractable model system with a complex nervous system has provided some insight into role of *Drosophila* Nab2 RNA-binding protein (**Chapter 2**), and *Drosophila* Rrp40 RNA exosome subunit (**Chapter 3**) protein in neuronal tissues.

Previous work shows the human ZC3H14 orthologue Nab2 when deleted from the *Drosophila* genome can lead to brain specific and behavioral changes (49). These *Nab2* mutants display a wide range of phenotypes; from embryonic lethality, to inability of “escapers” to fly or climb (locomotion defects) with wings permanently held-out like an airplane (49). At the molecular level, loss of Nab2 results in extended poly(A) RNA tails (49). In addition, this ubiquitously expressed poly(A) RNA binding protein likely regulates specific RNAs in the brain, as indicated by the finding that fly and mammalian models of ZC3H14 affect specific neurons that

are essential for learning and memory, specifically the mushroom bodies (MBs) in flies and hippocampal neurons in mice brain (39, 42, 73, 82).

To build on previous studies of Nab2 (40-43, 49, 56, 59, 66, 73), an important question then became: **What specific brain mRNAs are regulated by *Drosophila* Nab2?** Using a high-throughput RNA-Sequencing approach, RNA was isolated from the heads of male and female *Nab2* mutant flies and isogenic controls. The sequence data revealed the following: (1) only a small fraction of total RNAs is affected by Nab2 loss (2.2% in males and 3.7% in females) and (2) affected RNAs are predominantly mRNAs (52% in males and 77% in females) out of the significantly affected RNAs. Taken together, the transcriptome-wide analysis of *ZC3H14* knock down in cells and *Drosophila* Nab2 in fly brains implicates ZC3H14/Nab2 in controlling critical RNAs.

4.3 Implications of ZC3H14/Nab2 In Regulating mRNAs In Brain Neurons

Depletion of mammalian ZC3H14 in cultured mouse hippocampal neurons resulted in abnormal neuronal development, specifically fewer and shorter neurites (82). These data strongly implicate ZC3H14 in regulating particular subsets of neuronal RNAs. Though *ZC3H14* patient brains cannot be accessed or analyzed as these individuals are still living, *ZC3H14* mutant mice brains are available for RNA-Sequencing (RNA-Seq), and it would be of great interest to perform RNA-Seq from *ZC3H14* mutant mice brain hippocampi and compare to the *Drosophila* (whole adult heads) RNA-Seq presented (see **Chapter 2**).

Although we have yet to discover the ZC3H14 mRNA targets in the mouse brain, a few possible candidates that emerge from the *Drosophila* RNA-Seq and are conserved in higher eukaryotes are the synaptic enzyme transcript *Acetylcholine esterase* (*Ace*), the voltage-gated ion channel transcript (I_h), the U2-snRNP (spliceosome) component *U2AF50* and the transcript

encoding the CASK protein. Mutations in the *CASK* gene are linked to an X-linked brain malformation syndrome where affected individuals present with congenital or postnatal microcephaly, cerebellar hypoplasia, and severe mental retardation (194). Moreover, further analysis of *Drosophila* Nab2-regulated RNAs show an enrichment for gene ontology (GO) term for "RNA splicing" and upregulation of alternative splicing events (see **Chapter 2**). These datasets link ZC3H14/Nab2 to regulating a small subset of target neuronal RNAs.

4.4 Exploring A Role For *Drosophila* Nab2 In Regulating mRNA Splicing

Analysis of the 534 RNAs affected in *Nab2* mutant heads showed a significant enrichment in genes involved in splicing events (**Figure 2** in **Chapter 2**). Though Nab2 has been implicated in splicing (specifically the accumulation of approximately 100 intron-containing pre-mRNAs) based on an array to examine intron-containing genes (39), this is the first *in vivo* evidence linking *Drosophila* Nab2 to splicing of brain specific neuronal mRNAs. The specific splicing defects observed in the *Nab2* mutants are exon skipping, intron retention, and alternative 5' and 3' splice site events. Differential exon sequence (DEXSeq) analysis of top mis-spliced transcripts revealed, the transcript *Sex lethal* (*Sxl*) as a top candidate with a ~3-fold-change in exon 3 inclusion in *Nab2* mutant females, while in males the transcript is properly spliced in a sex-specific manner. Exon 3 of the *Sxl* mRNA transcript is subject to sex-specific splicing regulation (97, 98). Thus, the sex-specific mis-spliced phenotype observed in the *Nab2* mutant females signifies two things: (1) under normal conditions, Nab2 does not affect all poly(A) mRNAs and (2) Nab2 promotes sex-specific splicing of a subset of target mRNAs (i.e., *Sxl*) in a tissue-specific manner (i.e., RNAs in fly head). Of all the transcripts affected by Nab2 loss, *Sxl* was the most statistically significant transcript detected in *Nab2* mutant female flies. This finding from the RNA-Seq analysis led to further molecular studies on Nab2-mediated regulation of *Sxl* splicing.

4.4.1 Impact of Nab2 Loss In A Subset of RNAs

The *Sxl* transcript encodes the Sxl protein, which binds to and controls *Sxl* pre-mRNAs splicing to promote production of more Sxl product, while simultaneously negatively controlling male-specific splicing, (i.e., no Sxl proteins are produced) (126). In other words, Sxl positively auto-regulates itself in a sex-specific manner (only produced in female flies), making this RBP a key regulator of sex determination in *Drosophila* (195).

The retention of *Sxl* exon 3, which contains a premature termination codon, and thus mean no functional Sxl protein is produced, results in male-specific characteristics and sexual development while exclusion of exon 3, which allows production of Sxl protein, promotes female-specific development (98, 127). The presence of the male-specific exon 3 *Sxl* transcript in *Nab2* mutant females could be explained by a model where under normal conditions, Nab2 binds *Sxl* pre-mRNA to promote splicing, however, upon Nab2 loss, the sites where Nab2 binds *Sxl* now become susceptible to RNA methylation (see **Figure 4—1**, Summary Model Figure). The RNA-Seq data revealed Nab2 regulates a small number of RNAs in the fly head, raising the possibility that affected Nab2-regulated RNAs and alternative splicing of sex-specific *Sxl* transcript could be modulated by RNA methylation within head neurons.

4.5 Implication of Nab2 In Regulating mRNAs Via Modulation of m^{6}A RNA Methylation

The gene encoding the m^{6}A methyltransferase, *Mettl3* (m^{6}A 'writer') is the core catalytic subunit that methylates target mRNAs. Like RBPs and their ability to control every aspect of RNA biology, m^{6}A RNA methylation is also involved in numerous steps that regulate gene expression (30, 53, 85, 86). Deletion of *Mettl3* in *Drosophila* (*Mettl3^{null}* mutant) is linked to aberrant splicing of *Sxl*, including retention of exon 3 in females, similar to loss of Nab2. *Nab2* mutants also show an aberrant 5' splice site which suggests that multiple regulatory sites for Nab2 exist on the *Sxl* RNA

transcript (i.e., the 5' and 3' splice sites). This then prompted the idea that perhaps, Nab2 could functionally interact with Mettl3 to modulate m^6A RNA levels.

Proteomic analysis from an unbiased mass spectrometry experiment on ZC3H14 immunoprecipitated lysates from mouse brains showed protein-protein interactions between ZC3H14 and the m^6A methyltransferase 'writer' KIAA1429 (144). The *Drosophila* orthologue of KIAA1429 is Virilizer (*vir*) which is known to interact with other m^6A 'writer' WTAP to regulate alternative splicing of pre-mRNAs involved in *Drosophila* sex determination (85, 145).

The genetic recombination of a *Nab2* mutant and a *Mettl3* (m^6A) mutant onto the same chromosome (1% chance of recombination) proved challenging, however, upon successful recombination, it became abundantly clear that there is a strong genetic interaction between *Nab2* and *Mettl3* genes. Interestingly, *Drosophila Nab2* interacts genetically with *Mettl3* and *vir*. Chapter 2 of this dissertation presents a model where *Nab2* and *Mettl3* work together to modulate splicing of a sex determination gene (*Sxl*), and effect MBs development in an m^6A dependent manner. In addition to genetic interaction, molecular evidence strongly supports a role for Nab2 in regulating brain mRNAs, specifically *Sxl* mRNAs in an m^6A-dependent manner. Using the m^6A anti-body, methylation status of the female- specific *Sxl* pre-mRNA and mRNA was determined in *Nab2* mutant female heads. The m^6A RNA-IP (Me-RIP) experiments followed by qPCR analysis revealed an increase in steady-state *Sxl* m^6A methylation in *Nab2* mutant females relative to controls. The hypermethylation and sex-specific mis-splicing of *Sxl* transcript is rescued in the *Nab2/Mettl3* recombinant double mutant females. Genetically reducing the relative m^6A levels by half in the background of the *Nab2* mutation results in improved splicing events, specifically a decrease in male-specific exon 3 retention in the *Sxl* transcript in female Nab2 mutants. The mechanism(s) by which Nab2 and Mettl3 could co-regulate mRNAs such as *Sxl* could result from direct or indirect blocking of the m^6A methylation

machinery by Nab2 bound to target RNAs. The association of Nab2 with *Sxl* pre-mRNAs as shown in the Me-RIP experiments (see **Figure 6D**, in **Chapter 2**) supports a model where Nab2 binding physically occludes the methylation machinery, resulting in a transcript-specific decrease in m^6A level. An important future experiment would be to measure the binding affinity of Nab2 for to RNAs identified to contain m^6A modifications and those that do not. Since Nab2 is a polyadenosine RNA binding protein, it is likely that Nab2 could bind and block m^6A modifications.

The aberrant splicing and hypermethylation of *Sxl* transcript have downstream functional consequences in controlling dosage compensation pathway genes. In the fruit fly, *Sxl* controls somatic sexual differentiation through the female-specific, alternative splicing of *transformer* (*tra*) pre-mRNA, and regulates the dosage compensation gene *male-specific-lethal 2* (*msl-2)*(196). Although *tra* and *msl-2* or other Sxl-regulated transcripts were not altered in the RNA-Seq data, the effect of Nab2 loss on these transcripts might be significant in other cell types/tissues. The particular splicing defect observed in the *Sxl* transcript in these *Nab2* mutant females is specific to brain neurons of interest, as most studies of Sxl targets have been focused on the germ cells. Thus, a more completed list of targets modulate by Sxl in neurons is required.

4.6 Exploring roles of Nab2 in controlling dosage compensation genes in neurons

The RBP, Sxl, controls *Drosophila* sex determination and dosage compensation (124). Dosage compensation only occurs in male insects, where transcription of genes on the X chromosome is increased 2-fold to compensate for having a single X chromosome instead of the XX present in females (133). The dosage compensation complex in *Drosophila* is composed of many male-specific mRNAs (*msl-2* and *mle*) and non-coding RNAs (*roX1* and *roX2*), but it is the translational repression of the male-specific lethal complex gene, *msl-2* in particular, that prevents dosage compensation in female flies (127). Although the steady-state mRNA levels of the genes involved

in dosage compensation are not altered in the *Nab2* mutants head RNA-Seq dataset, the effects of Nab2 loss on *Sxl* mRNA and the dosage compensation pathway genes could be at the level of translational control. Alternatively, there could be neuronal targets of Sxl that have not yet been defined.

To test this hypothesis, genetic interaction experiments between *Nab2* and the dosage compensation genes *mls-2*, *mle*, and *roX1* were carried out. As presented in **Chapter 2**, all these loss of function alleles rescue *Nab2* mutant phenotypes. This is the first genetic evidence implicating Nab2 in controlling dosage compensation genes. An interesting observation is that, both male and female *Nab2* mutants are rescued by expressing a mutant allele of the male-specific dosage compensation genes. The female *Nab2* mutants dosage compensation is consistent with the presence of a male-specific *Sxl* transcript only in neuronal tissues. This could imply these female brains are masculinized. Fly brains contain sexually dimorphic neurons that can be visualized using immunofluorescence microscopy, so a future experiment is to analyze the brains of Nab2 mutant flies.

4.7 The RNA Exosome Regulates Neuronal RNAs

Though the main focus of this dissertation was on the RBP ZC3H14/Nab2, similar experiments were carried out to explore the role of *Drosophila* EXOSC3 orthologue Rrp40 in tissue-specific pathology (see **Chapter 3**). The RNA-Seq experiments performed on the heads of the two distinct RNA exosome mutant fly models of disease, revealed each distinct mutation on *Rrp40* had a distinct effect on the overall transcriptome. RNA-Seq datasets from these exosome mutant flies show that loss of Rrp40 results in significant changes in mRNAs as well as an increase in steady-state RNA levels consistent with the function of a complex required for RNA turnover (28). A significant and surprising finding from this published work was the correlation between *Drosophila* amino acid substitution mutant disease phenotypes and disease severity observed in

EXOSC3 patients. In addition, the two different missense mutations in the *Rrp40* mutants show 1) Rrp40 is required for mushroom body neuron development, and 3) *Rrp40* mutants show clear transcriptomic differences which could underlie the phenotypic differences observed in the fly model of *EXOSC3*. These RNA-Seq results are similar to that of Nab2 in that loss of Rrp40 affect a target set of neuronal RNAs, coding and non-coding.

4.8 Future Direction and Questions

The experiments presented in this dissertation reveal a critical role for *Drosophila Nab2* and *Drosophila Rrp40* (EXOSC3 in humans) in neurons. However, establishing *Drosophila* as biologically relevant model for these human neurological diseases still leaves several open-ended questions in need of further investigation. The underlying mechanism(s) explaining how mutation of a single genes encoding a ubiquitously expressed protein can lead to tissue-specific diseases and abnormal neurological behaviors is a major question in the field. Why are some cells/tissue more sensitive to loss of these ubiquitously expressed proteins as compared to unaffected cells/tissues? Some clues into this requirement for certain proteins, particularly in neurons, are provided by the brain enriched RNA-Seq datasets for both *Nab2/ZC3H14* and *Rrp40/EXOCSC3*. The splicing defects of *Sxl* observed *Nab2* mutant females is specific to neuronal tissues, however, further experiments to determine whether the splicing defects happen in other tissues is required. Both these datasets provide a list of target mRNA candidates that could be examined in mice or human cells. The neuro-specific expression of human ZC3H14 or EXOSC3 in *Nab2* or *Rrp40* mutants respectively rescues developmental and behavioral phenotypes (see **Chapter 2** & **3**) (28, 49, 73). These findings indicate the human proteins are functional orthologues of the corresponding *Drosophila* protein, thus making the fly a biologically relevant model system.

In addition to exploring the rich RNA-Seq datasets generated here, it would be interesting to explore how brain enriched RNA-Seq datasets generated from mouse models of *ZC3H14* and *EXOSC3* compare to the *Drosophila* datasets. However, unlike *ZC3H14*, *EXOSC3* is essential, but single missense mutants could be utilized to generate an *EXOSC3* mouse model for analysis (82, 160). These models would enable investigators to determine what mRNAs are affected in the mammalian *ZC3H14/EXOSC3* mutants. Furthermore, performing long-read sequencing of nascent RNAs could help elucidate any mRNA processing defects in either the fly or the mouse model system, specifically in intron retention (197). In addition to long read sequencing, m^6A RNA-Seq on mammalian ZC3H14-bound mRNAs could provide insight into whether hypermethylation of these transcripts upon loss of ZC3H14 could be linked to working memory defects in mice and intellectual disability in humans. Most importantly, performing single-cell RNA-Seq on *Drosophila* mushroom body neurons or mice hippocampal neurons might address the cell-specific requirements for these genes. Finally, proteomics and RNA immunoprecipitation experiments would be valuable to distinguish if the effects of Nab2 loss on proteins and RNAs bound by Nab2 are direct or indirect.

4.9 Final Conclusions

The combined collaborative efforts of all the co-authors of the works presented here in this dissertation has led to the production of two high resolution RNA-Seq datasets on the *Drosophila* head transcriptome (see **Chapter 2** & **3**). These experiments were performed to explore the loss of ZC3H14/Nab2 in intellectual disability and the role of EXOSC3/Rrp40 in pontocerebellar hypoplasia. The datasets were compiled using both genetics and molecular biology approaches to investigate molecular pathways and mechanisms to elucidate how these genes may affect specific tissues/cell types. Although it is clear that neurons are the cell type most susceptible to loss of ZC3H14/Nab2 and EXOSC3/Rrp40, the outstanding questions of “why” still remains. Neuron are

highly specialized cells that requires spatiotemporal control of gene expression , so this need for spatial and temporal regulation could underlie the requirement not only for ZC3H14 and the RNA exosome, but also for other RNA-binding proteins linked to neurological disease (198).

In the case of ZC3H14/Nab2, the genetic and molecular biology data revealing a functional link between *Nab2* and the *Mettl3* gene in regulation of *Sxl* and presumably other transcripts supports a "goldilocks" model where excess m^6A RNA methylation in the *Nab2* mutants causes neuronal mRNA processing defects. This splicing effect of *Sxl* caused by Nab2 loss is similar to the splicing defect that results from m^6A hypomethylation in *Mettl3* mutants. Importantly, the *Nab2/Mettl3* double mutant restores m^6A methylation and *Sxl* splicing back to near normal (see **Chapter 2**). Although there are no known demethylases in *Drosophila*, Nab2 could play a role as a negative regulator of m^6A level by preventing m^6A methylation of A-rich sequences bound by Nab2 (see **Figure 4—1**, **Summary Model Figure**). Instead of recruiting the m^6A machinery directly or indirectly, Nab2 could act as a physical barrier on the adenosines it binds, protecting these sites from hypermethylation. Nab2 binding to *Sxl* mRNA modulated by m^6A methylation to control RNA metabolism are biologically significant findings, as the mammalian m^6A methylation enzyme, METTL3 is evolutionarily conserved and is essential for neurogenesis (199).

In conclusion, the conserved pathways presented in this dissertation open the doors to integrated analysis that exploit the strengths of different model systems and ultimately help to define molecular mechanisms underlying human disease. The *Drosophila* model has not been used extensively to explore these pathways and this dissertation highlights the strength and advantage of these approaches. Future studies, including other models, will be required to define and understand the functions of these conserved proteins in RNA biology and disease.

4.10 Summary Model Figure

Figure 4—1

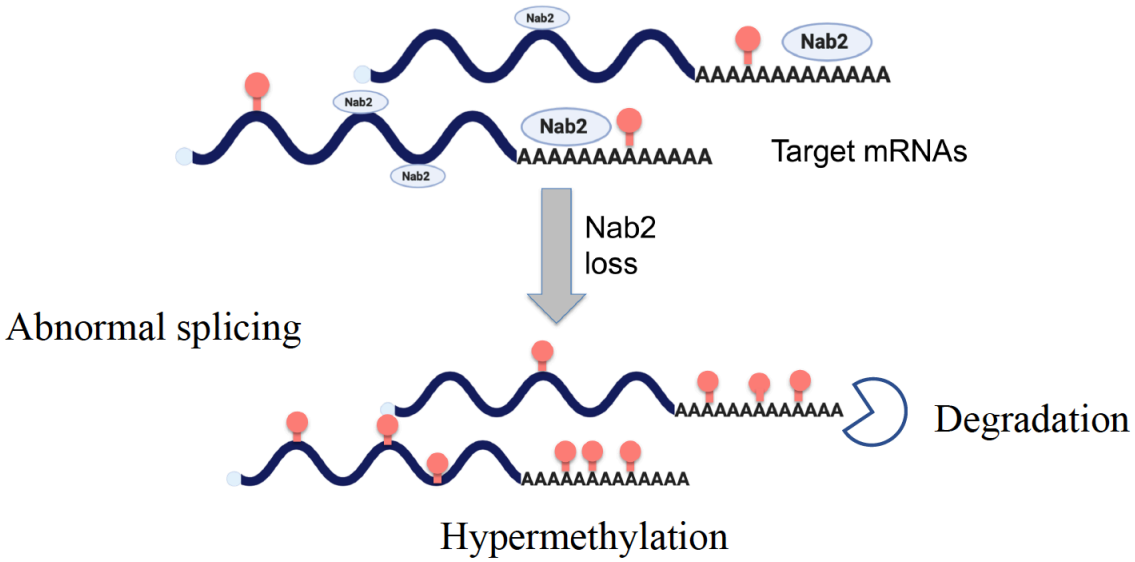

Nab2
Nab2
AAAAAAAAAAAAAA
Nab2
Nab2
Nab2
AAAAAAAAAAAAAA
Target mRNAs
Nab2
loss
Abnormal splicing
AAAAAAAAAAAAAA
Degradation
AAAAAAAAAAAAAA
Hypermethylation

Figure 4—1: Model Summary Figure for Roles of *Drosophila* Nab2.

Proposed model for functions of Nab2 in RNA metabolism based on data presented in chapter two of this dissertational work. Nab2 binds A-rich RNA sequences and m^6A methylation happens on adenosines (As). Intragenic regions could impact splicing (i.e., *Sxl*), while regulation of poly(A) tail could affect degradation of Nab2 bound RNAs. The data presented here show a role for Nab2 in modulating splicing of target neuronal mRNAs.

Chapter 5

Materials and Methods

5.1 Material and Methods: Chapter 2

Drosophila stocks and genetics

Drosophila melanogaster stocks and crosses were maintained in humidified incubators at 25°C with 12hr light-dark cycles. The alleles *Nab2*^{ex3} (null), *Nab2*^{pex41} (*precise excision 41*; control) and *UAS-Flag-Nab2* have been described previously (44, 49). Lines from Bloomington *Drosophila* Stock Center (BDSC): *GMR-Gal4 (*#1350), *elav*^{C155}*-Gal4* (#458), *msl-2*^{227} (#5871), *msl-2*^{kmA} (#25158), *mle*^{9} (#5873), *roX1*^{ex6} (#43647). The *Mettl3*^{null} allele was a kind gift of J-Y. Roignant. The *Nab2*^{ex3}, *Mettl3*^{null} chromosome was generated by meiotic recombination and confirmed by genomic PCR (**Supplemental Figure 6**).

RNA Sequencing (RNA-Seq) on *Drosophila* heads

RNA-Seq was performed on three biological replicates of 60 newly-eclosed adult female and male *Drosophila* heads genotype (control and *Nab2*^{ex3} mutants). Heads were collected on dry ice, lysed in TRIzol (ThermoFisher), phase-separated with chloroform, and ran through a RNeasy Mini Kit purification column (QIAGEN). Samples were treated with DNase I (QIAGEN) to remove DNA contamination and transported to the University of Georgia's Genomics and Bioinformatics Core for sequencing. rRNA was depleted using a Ribo-Zero Gold Kit (Illumina) and cDNA libraries were prepared using a KAPA Stranded RNA-Seq Kit (Roche). Quality control steps included initial Qubit quantification along with RNA fragment size assessment on an Agilent 2100 Bioanalyzer before and after rRNA depletion. The cDNA libraries were then sequenced for 150 cycles on a NextSeq 500 High Output Flow Cell (Illumina) set to generate paired-end, 75 base-pair (bp) reads. Total sequencing yield across all samples was 81.48 Gbp, equivalent to about 1.1 billion reads in total and 91 million reads per sample. Sequencing accuracy was high; 93.52% of reported bases have a sequencing quality (Q) score greater than or equal to 30.

Read mapping, differential expression, visualization of RNA-Seq dataset

Raw read FASTA files were analyzed on the Galaxy web platform (usegalaxy.org 200). The BDGP6 release *Drosophila melanogaster* genome (201) from release 92 of the Ensembl database (202) was used as input for subsequent read mapping, annotation, and visualization. Briefly, reads from all four NextSeq500 flow cell lanes were concatenated using the Galaxy *Concatenate datasets tail-to-head (cat)* tool and mapped using RNA STAR (114) with default parameters with some modifications (available upon request). Mapped reads were assigned to exons and tallied using featureCounts (115) default parameters with some modifications (available upon request). Differential expression analysis was conducted for all 12 samples using DESeq2 (116) (Galaxy Version 2.11.40.1) and default parameters with some modifications (available upon request). Differential exon usage was analyzed using Galaxy Version 1.20.1 of DEXSeq (122) and the associated Galaxy tool DEXSeq-Count in both "*prepare annotation*" and "*count reads*" modes. Both tools were run with the Ensembl GTF with default parameters with some modifications (available upon request). Unlike with DESeq2, female samples and male samples were compared in independent DEX-Seq analyses. Outputs from all of these tools were downloaded from Galaxy for local analysis, computation, and visualization.

For visualization, custom R scripts were written to generate volcano plots and heatmaps (available upon request). Additional R packages used include ggplot2 (203) and ggrepel (204). Scripts in the R programming language (205) were written and compiled in RStudio (206). Principal component analysis was conducted on Galaxy. Mapped reads were visualized in the Integrative Genomics Viewer (IGV) (129) and annotated based on data available on Flybase (207). DESeq2 was used with default settings for transcript assembly to generate pie charts and correlation scatter plot. The resulting assembled transcripts were compared using DESeq2 to identify genes that change significantly (p-value<0.05, >1-fold change) in expression. Significant fold change

values in either male or female from DESeq2 (adj. *p*-val<0.05 and $|\log_2 FC|>1$) were plotted, with the color indicating the fold change threshold reached in either males or females. Significantly DE genes (adj. *p*-val<0.05 and $|\log_2 FC|>1$) were classified by type, as indicated by their gene ID. Mixtures of isoforms (MISO) version 0.5.4 was run on aligned reads of length 76 bp, and then delta PSI values were computed for WT vs. mutant using miso for male and female samples separately with the flags --num-inc 10 --num-exc 10 --num-sum-inc-exc 50 --delta-psi 0.3 --bayes-factor 10.

Mixtures of Isoforms (MISO) Analysis

MISO (121) was used to determine percent spliced in (PSI) values for annotated alternative 3' splice sites, alternative 5' splice sites, and retained introns for each sample separately as follows. Alternative splicing annotations were generated using the rnaseqlib (a direct link to script is listed here)

(https://rnaseqlib.readthedocs.io/en/clip/) script, gff_make_annotation.py, with flags--flanking-rule commonshortest --genome-label dm6. Replicates for each sample were pooled, and only full-length, mapped reads (76 bp) were used for the MISO analysis since MISO requires all reads input to be of the same length. MISO was run with the flag –prefilter, and the output was then input into the script, summarize_miso, with the flag --summarize-samples. Next, differential, alternative 5' and 3' splice sites, and differential retained introns, were determined between *Nab2*^{ex3} and control for males and females, separately, using the script, compare_miso, with flag --compare-samples. The output of compare_miso was then input into the script, filter_events, with the flags --filter --num-inc 10 --num-exc 10 --num-sum-inc-exc 50 --delta-psi 0.3 --bayes-factor 10, to obtain the final differential PSI values.

Gene ontology (GO) analysis

Gene Set Enrichment Analysis (GSEA) software (118) was employed for Gene Ontology (GO) analysis. For clarity, analyses were conducted separately for each of the three top-level GO domains: *molecular function*, *biological process*, and *cellular component*. GSEA-compatible GO term gene sets for *Drosophila melanogaster* were acquired using the GO2MSIG web interface (208). GSEA Desktop for Windows, v4.0.3 (Broad Institute) was then used to identify two distinct classes of GO terms, independently for females and for males: (1) terms enriched among up- and downregulated transcripts in $Nab2^{ex3}$ compared to controls, and (2) terms enriched among transcripts alternatively spliced in $Nab2^{ex3}$ compared to controls. For the first class, inputs consisted of all genes whose expression could be compared by DESeq2 (i.e., adjusted *p*-value was not NA). For the second class, inputs consisted of all genes with known, annotated alternative splicing events according to MISO. To identify the first class of ontology terms, genes were ranked by $\log_2$ (fold change) calculated by DESeq2 and analyzed by the GSEA-Pre-ranked tool. To identify the second class of ontology terms, genes with were ranked by the absolute value of the difference in PSI (percent spliced in) between $Nab2^{ex3}$ and control calculated by MISO. This second ranking was analyzed by the GSEA-Preranked tool. Enriched GO terms (nominal p-value<0.05) identified for the first class were evaluated manually, surfacing multiple terms directly related to splicing. Enriched GO terms (nominal p-value<0.05) for the second class were ordered by normalized enrichment score (NES) and evaluated to identify the top “independent” GO terms. Terms were defined as “independent” by reference to their position in the GO hierarchy as reported on each term’s “Inferred Tree View” page of the AmiGO2 GO database web tool (209). “Independent” terms had no parent, child, or sibling terms in the GO hierarchy associated with a higher NES than their own.

RBPs Motif Enrichment Analysis using Mixture of Isoforms (MISO) Analysis

RNA sequences were taken at differentially retained introns and alternative 3' and 5' splice sites obtained from the MISO analysis on males and females separately (*Nab2*^{ex3} mutants vs. control). The sequence for each of these went 25 bp into the exon(s) of interest and 1 kb into the intron of interest. In the case of alternative 3' and 5' splice sites, the sequences went 25 bp into the exon starting from the alternative spice site that is closest to the center of the exon (i.e., the inner-most splice site), and 1 kb into the intron starting from that inner-most spice site. To convert these to RNA sequences, DNA sequences were first obtained using fastaFromBed (210), and then all T's were converted to U's with a custom script. To obtain putative binding sites for RBPs at these sequences, they were then input into fimo using the flags --text --max-strand and the "Ray2013_rbp_Drosophila_melanogaster.meme" file (211).

RNA isolation for reverse transcription (RT) PCR and real-time qPCR

Total RNA was isolated from adult tissues with TRIzol (Invitrogen) and treated with DNase I (Qiagen). For RT-PCR, cDNA was generated using SuperScript III First Strand cDNA Synthesis (Invitrogen) from 2μg of total RNA, and PCR products were resolved and imaged on 2% agarose gels (BioRad Image). Quantitative real-time PCR (qPCR) reactions were carried out in biological triplicate with QuantiTect SYBR Green Master Mix using an Applied Biosystems StepOne Plus real-time machine (ABI). Results were analyzed using the ΔΔCT method, normalized as indicated (e.g., to *Act5C*), and plotted as fold-change relative to control.

Supplemental Table 1: Primers used for RT and qPCR analysis

Name	Sequence	Detects
Sex-lethal pre-mRNA	Fwd: AGAACCAAAACTCCCTTACAGC	Intron2-Exon3
	Rev: GTGAGTGTCTTTCGCTTTTCG	
Sex-lethal pre-mRNA	Fwd: ACCAATAACCGACAACACAATC	Intron3-Exon4
	Rev: ACATCCCAAATCCACGCCCACC	
Sex-lethal mRNA	Fwd: GCTGAGCGCCAAAACAATTG	Exon2-Exon2
	Rev: AGGTGAGTTTCGGTTTTACAGG	
Sex-lethal RT-PCR	Fwd: ACACAAGAAAGTTGAACAGAGG	Exon 2-3-4
	Rev: CATTCCGGATGGCAGAGAATGG	
Sex-lethal mRNA	Fwd: GATTGAATCTCGATCATCGTTC	Exon2-Exon4
Female transcript	Rev: CATTCCGGATGGCAGAGAATGG	
Sex-lethal mRNA	Fwd: CGAAAAGCGAAAGACACTCACTG	Exon3-Exon4
Male transcript	Rev: CATTCCGGATGGCAGAGAATGG	
Act5C	Fwd: GAGCGCGGTTACTCTTTCAC	Actin
	Rev: ACTTCTCCAACGAGGAGCTG	
USP-16-45-RF	Fwd: ACACTTGGTCACGTCGTTCA	USP-16
	Rev: GGGCGCGCTCTTGAATTTAC	

Viability and lifespan analysis

Viability at 25°C was measured by assessing eclosion rates of among 100 wandering L3 larvae collected for each genotype and sex, and then reared in a single vial. Hatching was recorded for 5-6 days. At least 3 independents biological replicates per sex/genotype were tested and significance was calculated using grouped analysis on GraphPad (Prism). Lifespan was assessed at 25°C as described previously (28). In brief, newly eclosed animals were collected, separated by sex, placed in vials (10 per vial), and transferred to fresh vials weekly. Survivorship was scored

daily. At least 3 independents biological replicates per vials of each genotype was tested and significance was calculated using grouped analysis on GraphPad (Prism).

Locomotion assays

Negative geotaxis was tested as previously described (28). Briefly, newly eclosed flies (day 0) were collected, divided into groups of 10 male or females, and kept in separate vials for 2-5days. Cohorts of age-matched flies were then transferred to a 25-ml graduated cylinder for analysis. At least three biological replicates per sex were analyzed per genotype on GraphPad (Prism).

Brain dissection and mushroom body (MB) imaging

Brain dissections were performed essentially as previously described in (73). Briefly, adult brains were dissected in PBT (1xPBS, 0.1% TritonX-100) and collected in PBS at 4^{o}C. Brains were then fixed in 4% paraformaldehyde at RT, washed 3x in PBS, and incubated in 0.3% PBS-T (1xPBS, 0.3% TritonX-100). Following blocking for 1hr (0.1% PBS-T, 5% normal goat serum), brains were stained overnight in block+primary antibodies. After 5x washes in PBT, brains were incubated in block for 1hr and transferred into in block+secondary antibody for 3hrs. Brains were then washed 5x in PBT and mounted in Vectashield (Vector Labs). The FasII antibody clone 1D4 (Developmental Studies Hybridoma Bank) was used to label MBs at a 1:50 dilution. Whole brain images were captured with a 20x objective. Maximum intensity projections were obtained by combining serial optical sections (Z-stacks) with Nikon A1R HD25 software using Fiji. The number of a-lobe and b-lobe defects (e.g., thin, missing or fused) were scored analyzed using GraphPad (Prism).

Flag and m^{6}A RNA immunoprecipitation (Flag-RIP and MeRIP)

The FLAG-RIP and MeRIP protocols were performed using previously described protocols in (66) and (113) with some modification. Briefly, three replicates of 30 newly eclosed female flies were collected in 1.5 ml Eppendorf tubes and frozen in dry ice. Heads were removed with a 5.5 Dumont tweezer and homogenized with a mortar/pestle in Isolation buffer (50 mM Tris-HCl pH 8.1, 10 mM EDTA, 150 mM NaCl, and 1% SDS, 50 mM NaCl). This was diluted into IP buffer (50 mM HEPES, 150 mM NaCl, 5 mM EDTA, 0.5 mM DTT, 0.1% NP-40) supplemented with protease inhibitors (Roche) and RNasin Plus Inhibitor (Promega). Lysates were incubated with anti-Flag (M2 clone; Sigma) or anti-m^{6}A (Synaptic Systems) antibodies and recovered on magnetic Protein G Dynabeads (Invitrogen). After overnight incubation at 4^{o}C with rocking. Beads were washed 5x in IP buffer and RNA was isolated from antibody-bead precipitates, or controls (input samples) using TRIzol (ThermoFisher). Samples were treated with DNase-I and RNA was purified using RNeasy Kit (Qiagen).

Statistical Analysis

Group analysis on biological triplicate experiments was done using Two-way ANOVA (Turkey's multiple comparison test) on GraphPad (Prism) Version 8.4.2(464). Sample sizes (n) and *p*-values are denoted in the text or figures and noted by asterisks (for example, *p<0.05).

5.2 Materials and Methods: Chapter 3

Drosophila Stocks

Crosses were maintained in standard conditions at 25°C unless otherwise noted. Stocks used in this study: w^{1118}, *UAS-Rrp40*^{RNAi} (TRiP HMJ23923, BDSC #63834), *Df(2L)Exel6005* (BDSC #7492), *Tub>Gal80*^{ts} (BDSC #7017), and the following Gal4 stocks: *Actin5c>Gal4* (BDSC#25374), *elav>Gal4* (BDSC#458), repo>Gal4 (BDSC#7415) and *Mhc>Gal4* (BDSC#55133) were obtained from the Bloomington *Drosophila* Stock Center. *Rrp40*^{G11A} and *Rrp40*^{D87A} were generated by CRISPR/Cas9 editing (detail below) at Best gene, Inc. (CA).

Brain Dissections and Immunohistochemistry

Brain dissections and staining were performed as described previously (212). Briefly, brains of anesthetized animals were dissected in PTN buffer (0.1M $NaPO_4$, 0.1% Triton X-100), fixed in 4% paraformaldehyde (Electron Microscopy Sciences), and then stained overnight with primary antibody (ID4) diluted in PTN. Following several washes, brains were incubated with the appropriate fluorescently conjugated secondary antibody (1:250) in PTN for 3 hours at room temperature, washed in PTN, and then mounted in Vectashield (Vector Labs). The 1D4 anti-FasII hybridoma (1:20) developed by C. Goodman (213) was obtained from the Developmental Studies Hybridoma Bank (DSHB).

Generation of CRISPR/Cas9 Flies

Molecular Reagents.

pU6-gRNAs: Target-specific sequences were identified sequences using DRSC flyCRISPR optimal target finder (https://www.flyrnai.org/crispr/) for Rrp40 and gRNAs were synthesized as 5'-phosphorylated oligonucleotides, annealed, and ligated into the BbsI sites of pU6-BbsI chiRNA (214).

Homology-directed repair (HDR) templates: HDR donor vectors (Emory Integrated Genomic Core) were constructed by cloning a 3kb fragment of the *Rrp40* locus, including a loxP-flanked *3xP3-DsRed* cassette inserted downstream of the *Rrp40 3'UTR*, into KpnI/SmaI sites of the pBlueScript-II vector. Base changes corresponding to G11A and D87A were engineered into this backbone. The 3xP3-DsRed cassette allows positive selection based on red-fluorescence in the adult eye (215).

CRISPR-transformants: Injection of DNA mixtures (500 ng/μl HDR and 250 ng/μl U6-gRNA plasmid) into nos-Cas9 embryos and subsequent screening for dsRed+ transformants was performed by Bestgene, Inc. CRISPR. Edits were confirmed by direct sequencing (S3 Fig).

Lifespan Analysis

Lifespan was assessed at 25°C as described previously (216). In brief, newly eclosed animals were collected, separated by sex, placed in vials (up to 20 per vial), and transferred to fresh vials weekly. Survivorship was scored daily. At least 8 vials of each genotype were tested and averaged to calculate mean and standard deviation. Log-rank tests of survivorship curves were performed using PRISM (GraphPad, San Diego).

Behavioral Assays

Negative-geotaxis assay was performed as previously described (49) with some modification. Newly eclosed flies (*Rrp40*^{wt}, *Rrp40*^{G11A}, *Rrp40*^{D87A}) (day 0) were collected, divided into groups of 10, and kept in separate vials. Cohorts of age-matched flies were then transferred to a 25-ml graduated for analysis of climbing response. At least 10 groups (i.e., ~100 flies) were analyzed per genotype.

Immunoblotting

Protein lysates of whole flies or heads were resolved on 4–20% Criterion TGX polyacrylamide gels (Bio Rad), transferred to nitrocellulose membranes, incubated for ≥1hr in blocking buffer (5% non-fat dry milk in 0.1% TBS-Tween), followed by an overnight incubation at 4°C in primary antibody diluted in blocking buffer. Primary antibodies were detected using species-specific horse radish peroxidase (HRP) conjugated secondary antibodies (Jackson ImmunoResearch) and enhanced chemiluminescence (ECL, Sigma). Primary antibodies include: guinea pig anti-Rrp4 (1:1000), rabbit anti-Rrp40 (1:1000), guinea pig anti-Csl4 (1:000), and rat anti-Mtr3 (1:1000) gift of the Lis lab (178, 217).

RNA Sequencing Analysis

RNA-seq was performed on three replicates of 60 newly-eclosed adult heads per genotype (*Rrp40*^{G11A}, *Rrp40*^{D87A} and *Rrp40*^{wt}). Heads were collected and lysed in TRIzol (ThermoFisher). rRNA was depleted and cDNA libraries were prepared using Truseq RNA sample prep kit (Illumina). The cDNA paired-end reads were sequenced at the Georgia Genomics and Bioinformatics Core at UGA using a NextSeq (150 cycles) PE75 High Output Flowcell. The reads were mapped to the BDGP D. melanogaster (r6.25) genome. Several sequence datasets were downloaded from GenBank and

FlyBase for initial assessment of the data including: 1) All *Drosophila* ribosomal sequences found in Genbank: Drosophila_rRNA.fa; 2) All FlyBase D. melanogaster transcripts from assembly r6.25: dmel_transcripts_r6.25.fa. Adapters were trimmed from raw reads using Trimmomatic (ver. 0.36) (218), and trimmed reads were aligned to the ribosomal sequences using Bowtie2 (219), and only reads which did not map to the ribosome were kept for downstream analysis.

Non-ribosomal reads were then mapped to D. melanogaster r6.25 using Tophat2 (220). The resulting aligned reads were compared using DESeq2 (221) to identify genes that change significantly (p-value<0.05, >1.5-fold change) in expression. Only genes that were significantly changed compared to the control CRISPR line, *Rrp40*^{wt}, were used for further analysis.

RNA Isolation and qPCR

Total RNA was isolated from heads with TRIzol (Invitrogen) and treated with Turbo DNase (Invitrogen) to degrade contaminating DNA. For qRT-PCR analyses, cDNA was generated using MMLV Reverse transcriptase (Invitrogen) from 1 µg of total RNA, and then used as template for quantitative real-time PCR (qPCR) from duplicate samples of 10 ng cDNA with QuantiTect SYBR Green Master Mix using an Applied Biosystems StepOne Plus real time machine (ABI). Results were analyzed using the ΔΔCT method and normalized to Act5c as fold change relative to control. Oligonucleotide sequences are provided in (**S1 Table**).

Statistical analysis

All statistical analyses were performed in GraphPad (San Diego, CA). Comparisons between experimental groups were made using one-way analysis of variance (ANOVA), unless noted otherwise. All data are presented as means and standard error of the mean

(SEM) (error bars) for at least three independent experiments. Asterisks (*) indicate statistical significance at p-value < 0.05.

Chapter 6

References

1. F. Crick, Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).
2. J. D. Watson, F. H. Crick, The structure of DNA. *Cold Spring Harb Symp Quant Biol* **18**, 123-131 (1953).
3. F. J. Calero-Cuenca, C. S. Janota, E. R. Gomes, Dealing with the nucleus during cell migration. *Curr Opin Cell Biol* **50**, 35-41 (2018).
4. T. Kouzarides, Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
5. M. Lawrence, S. Daujat, R. Schneider, Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet* **32**, 42-56 (2016).
6. D. B. Dogini *et al.*, The new world of RNAs. *Genet Mol Biol* **37**, 285-293 (2014).
7. J. S. Mattick, The central role of RNA in human development and cognition. *FEBS Lett* **585**, 1600-1616 (2011).
8. H. G. Martinson, An active role for splicing in 3'-end formation. *Wiley Interdiscip Rev RNA* **2**, 459-470 (2011).
9. B. R. Graveley, Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* **17**, 100-107 (2001).
10. J. S. Mattick, RNA regulation: a new genetics? *Nat Rev Genet* **5**, 316-323 (2004).
11. M. D. Sacar Demirci, A. Adan, Computational analysis of microRNA-mediated interactions in SARS-CoV-2 infection. *PeerJ* **8**, e9369 (2020).
12. X. M. Chen, MicroRNA signatures in liver diseases. *World J Gastroenterol* **15**, 1665-1672 (2009).
13. B. Khraiwesh *et al.*, Transcriptional control of gene expression by microRNAs. *Cell* **140**, 111-122 (2010).
14. S. Kojima, D. L. Shingle, C. B. Green, Post-transcriptional control of circadian rhythms. *J Cell Sci* **124**, 311-320 (2011).
15. P. Klaff, D. Riesner, G. Steger, RNA structure and the regulation of gene expression. *Plant Mol Biol* **32**, 89-106 (1996).
16. K. Chen, X. Dai, J. Wu, Alternative splicing: An important mechanism in stem cell biology. *World J Stem Cells* **7**, 1-10 (2015).
17. N. Sonenberg, Cap-binding proteins of eukaryotic messenger RNA: functions in initiation and control of translation. *Prog Nucleic Acid Res Mol Biol* **35**, 173-207 (1988).
18. A. Ramanathan, G. B. Robb, S. H. Chan, mRNA capping: biological functions and applications. *Nucleic Acids Res* **44**, 7511-7526 (2016).
19. I. Topisirovic, Y. V. Svitkin, N. Sonenberg, A. J. Shatkin, Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip Rev RNA* **2**, 277-298 (2011).
20. S. Millevoi, S. Vagner, Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* **38**, 2757-2774 (2010).
21. R. Elkon, A. P. Ugalde, R. Agami, Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* **14**, 496-506 (2013).

22. C. R. Eckmann, C. Rammelt, E. Wahle, Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* **2**, 348-361 (2011).
23. D. L. Black, Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**, 291-336 (2003).
24. A. G. Matera, Z. Wang, A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15**, 108-121 (2014).
25. C. Kilchert, S. Wittmann, L. Vasiljeva, The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol* **17**, 227-239 (2016).
26. M. Schmid, T. H. Jensen, The exosome: a multipurpose RNA-decay machine. *Trends Biochem Sci* **33**, 501-510 (2008).
27. C. Schneider, D. Tollervey, Threading the barrel of the RNA exosome. *Trends Biochem Sci* **38**, 485-493 (2013).
28. D. J. Morton *et al.*, A Drosophila model of Pontocerebellar Hypoplasia reveals a critical role for the RNA exosome in neurons. *PLoS Genet* **16**, e1008901 (2020).
29. D. J. Morton *et al.*, The RNA exosome and RNA exosome-linked disease. *RNA* **24**, 127-142 (2018).
30. M. W. Hentze, A. Castello, T. Schwarzl, T. Preiss, A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**, 327-341 (2018).
31. F. Gebauer, T. Schwarzl, J. Valcarcel, M. W. Hentze, RNA-binding proteins in human genetic disease. *Nat Rev Genet* 10.1038/s41576-020-00302-y (2020).
32. K. E. Lukong, K. W. Chang, E. W. Khandjian, S. Richard, RNA-binding proteins in human genetic disease. *Trends Genet* **24**, 416-425 (2008).
33. A. H. Corbett, Post-transcriptional regulation of gene expression and human disease. *Curr Opin Cell Biol* **52**, 96-104 (2018).
34. S. A. Lima *et al.*, Short poly(A) tails are a conserved feature of highly expressed genes. *Nat Struct Mol Biol* **24**, 1057-1063 (2017).
35. Z. Wang, N. Day, P. Trifillis, M. Kiledjian, An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro. *Mol Cell Biol* **19**, 4552-4560 (1999).
36. C. P. Wigington, K. R. Williams, M. P. Meers, G. J. Bassell, A. H. Corbett, Poly(A) RNA-binding proteins and polyadenosine RNA: new members and novel functions. *Wiley Interdiscip Rev RNA* **5**, 601-622 (2014).
37. R. W. Smith, N. K. Gray, Poly(A)-binding protein (PABP): a common viral target. *Biochem J* **426**, 1-12 (2010).
38. D. A. Mangus, M. C. Evans, A. Jacobson, Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol* **4**, 223 (2003).
39. S. Soucek *et al.*, The Evolutionarily-conserved Polyadenosine RNA Binding Protein, Nab2, Cooperates with Splicing Machinery to Regulate the Fate of pre-mRNA. *Mol Cell Biol* **36**, 2697-2714 (2016).
40. J. T. Anderson, S. M. Wilson, K. V. Datar, M. S. Swanson, NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. *Mol Cell Biol* **13**, 2730-2741 (1993).
41. R. E. Hector *et al.*, Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *EMBO J* **21**, 1800-1810 (2002).
42. S. Soucek, A. H. Corbett, M. B. Fasken, The long and the short of it: the role of the zinc finger polyadenosine RNA binding protein, Nab2, in control of poly(A) tail length. *Biochim Biophys Acta* **1819**, 546-554 (2012).

43. M. B. Fasken, A. H. Corbett, M. Stewart, Structure-function relationships in the Nab2 polyadenosine-RNA binding Zn finger protein family. *Protein Sci* **28**, 513-523 (2019).
44. S. M. Kelly *et al.*, A conserved role for the zinc finger polyadenosine RNA binding protein, ZC3H14, in control of poly(A) tail length. *RNA* **20**, 681-688 (2014).
45. S. M. Kelly *et al.*, Recognition of polyadenosine RNA by zinc finger proteins. *Proc Natl Acad Sci U S A* **104**, 12306-12311 (2007).
46. A. Banerjee, L. H. Apponi, G. K. Pavlath, A. H. Corbett, PABPN1: molecular function and muscle disease. *FEBS J* **280**, 4230-4250 (2013).
47. M. Luigetti *et al.*, Oculopharyngeal muscular dystrophy: Clinical and neurophysiological features. *Clin Neurophysiol* **126**, 2406-2408 (2015).
48. B. L. Phillips *et al.*, Post-transcriptional regulation of Pabpn1 by the RNA binding protein HuR. *Nucleic Acids Res* **46**, 7643-7661 (2018).
49. C. Pak *et al.*, Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in Drosophila and humans. *Proc Natl Acad Sci U S A* **108**, 12390-12395 (2011).
50. A. W. Kuss *et al.*, Autosomal recessive mental retardation: homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots. *Hum Genet* **129**, 141-148 (2011).
51. R. L. Schalock, R. Luckasson, A Systematic Approach to Subgroup Classification in Intellectual Disability. *Intellect Dev Disabil* **53**, 358-366 (2015).
52. L. E. Vissers, C. Gilissen, J. A. Veltman, Genetic studies in intellectual disability and related disorders. *Nat Rev Genet* **17**, 9-18 (2016).
53. S. Gerstberger, M. Hafner, T. Tuschl, A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829-845 (2014).
54. S. Gerstberger, M. Hafner, M. Ascano, T. Tuschl, Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. *Adv Exp Med Biol* **825**, 1-55 (2014).
55. K. A. Marfatia, E. B. Crafton, D. M. Green, A. H. Corbett, Domain analysis of the Saccharomyces cerevisiae heterogeneous nuclear ribonucleoprotein, Nab2p. Dissecting the requirements for Nab2p-facilitated poly(A) RNA export. *J Biol Chem* **278**, 6731-6740 (2003).
56. S. M. Kelly *et al.*, Recognition of polyadenosine RNA by the zinc finger domain of nuclear poly(A) RNA-binding protein 2 (Nab2) is required for correct mRNA 3'-end formation. *J Biol Chem* **285**, 26022-26032 (2010).
57. C. R. Guthrie, G. D. Schellenberg, B. C. Kraemer, SUT-2 potentiates tau-induced neurotoxicity in Caenorhabditis elegans. *Hum Mol Genet* **18**, 1825-1838 (2009).
58. A. I. Lamond, D. L. Spector, Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol* **4**, 605-612 (2003).
59. L. M. Reuter, D. M. Meinel, K. Strasser, The poly(A)-binding protein Nab2 functions in RNA polymerase III transcription. *Genes Dev* **29**, 1565-1575 (2015).
60. T. Zars, R. Wolf, R. Davis, M. Heisenberg, Tissue-specific expression of a type I adenylyl cyclase rescues the rutabaga mutant memory defect: in search of the engram. *Learn Mem* **7**, 18-31 (2000).
61. M. Leyssen, B. A. Hassan, A fruitfly's guide to keeping the brain wired. *EMBO Rep* **8**, 46-50 (2007).
62. J. Roote, A. Prokop, How to design a genetic mating scheme: a basic training package for Drosophila genetics. *G3 (Bethesda)* **3**, 353-358 (2013).

63. M. Drozd, B. Bardoni, M. Capovilla, Modeling Fragile X Syndrome in Drosophila. *Front Mol Neurosci* **11**, 124 (2018).
64. C. L. Gatto, K. Broadie, Drosophila modeling of heritable neurodevelopmental disorders. *Curr Opin Neurobiol* **21**, 834-841 (2011).
65. E. Chen, M. R. Sharma, X. Shi, R. K. Agrawal, S. Joseph, Fragile X mental retardation protein regulates translation by binding directly to the ribosome. *Mol Cell* **54**, 407-417 (2014).
66. R. S. Bienkowski *et al.*, The Conserved, Disease-Associated RNA Binding Protein dNab2 Interacts with the Fragile X Protein Ortholog in Drosophila Neurons. *Cell Rep* **20**, 1372-1384 (2017).
67. G. J. Bassell, S. T. Warren, Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron* **60**, 201-214 (2008).
68. L. S. Wijetunge, S. Chattarji, D. J. Wyllie, P. C. Kind, Fragile X syndrome: from targets to treatments. *Neuropharmacology* **68**, 83-96 (2013).
69. L. Kaufman, M. Ayub, J. B. Vincent, The genetic basis of non-syndromic intellectual disability: a review. *J Neurodev Disord* **2**, 182-209 (2010).
70. H. H. Ropers, Genetics of intellectual disability. *Curr Opin Genet Dev* **18**, 241-250 (2008).
71. H. H. Ropers, Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet* **11**, 161-187 (2010).
72. L. Musante, H. H. Ropers, Genetics of recessive cognitive disorders. *Trends Genet* **30**, 32-39 (2014).
73. S. M. Kelly *et al.*, The Drosophila ortholog of the Zc3h14 RNA binding protein acts within neurons to pattern axon projection in the developing brain. *Dev Neurobiol* **76**, 93-106 (2016).
74. S. Benzer, BEHAVIORAL MUTANTS OF Drosophila ISOLATED BY COUNTERCURRENT DISTRIBUTION. *Proc Natl Acad Sci U S A* **58**, 1112-1119 (1967).
75. S. Benzer, Genetic dissection of behavior. *Sci Am* **229**, 24-37 (1973).
76. Y. Dudai, Neurogenetic dissection of learning and short-term memory in Drosophila. *Annu Rev Neurosci* **11**, 537-563 (1988).
77. M. Heisenberg, Mushroom body memoir: from maps to models. *Nat Rev Neurosci* **4**, 266-275 (2003).
78. J. D. Armstrong, J. S. de Belle, Z. Wang, K. Kaiser, Metamorphosis of the mushroom bodies; large-scale rearrangements of the neural substrates for associative learning and memory in Drosophila. *Learn Mem* **5**, 102-114 (1998).
79. C. I. Michel, R. Kraft, L. L. Restifo, Defective neuronal development in the mushroom bodies of Drosophila fragile X mental retardation 1 mutants. *J Neurosci* **24**, 5798-5809 (2004).
80. L. Pan, Y. Q. Zhang, E. Woodruff, K. Broadie, The Drosophila fragile X gene negatively regulates neuronal elaboration and synaptic differentiation. *Curr Biol* **14**, 1863-1870 (2004).
81. K. Fushima, H. Tsujimura, Precise control of fasciclin II expression is required for adult mushroom body development in Drosophila. *Dev Growth Differ* **49**, 215-227 (2007).
82. J. Rha *et al.*, The RNA-binding protein, ZC3H14, is required for proper poly(A) tail length control, expression of synaptic proteins, and brain function in mice. *Hum Mol Genet* **26**, 3663-3681 (2017).
83. C. P. Wigington, K. J. Morris, L. E. Newman, A. H. Corbett, The Polyadenosine RNA-binding Protein, Zinc Finger Cys3His Protein 14 (ZC3H14), Regulates the Pre-mRNA Processing of a Key ATP Synthase Subunit mRNA. *J Biol Chem* **291**, 22442-22459 (2016).

84. J. S. Satterlee *et al.*, Novel RNA modifications in the nervous system: form and function. *J Neurosci* **34**, 15170-15177 (2014).
85. Y. Yang, P. J. Hsu, Y. S. Chen, Y. G. Yang, Dynamic transcriptomic m(6)A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Res* **28**, 616-624 (2018).
86. D. Dominissini *et al.*, Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-206 (2012).
87. Y. Niu *et al.*, N6-methyl-adenosine (m6A) in RNA: an old modification with a novel epigenetic function. *Genomics Proteomics Bioinformatics* **11**, 8-17 (2013).
88. T. Lence, M. Soller, J. Y. Roignant, A fly view on the roles and mechanisms of the m(6)A mRNA modification and its players. *RNA Biol* **14**, 1232-1240 (2017).
89. E. Scholler *et al.*, Interactions, localization, and phosphorylation of the m(6)A generating METTL3-METTL14-WTAP complex. *RNA* **24**, 499-512 (2018).
90. A. Soldano *et al.*, The m6A reader Ythdf restricts axonal growth in Drosophila through target selection modulation of the Fragile X mental retardation protein. *bioRxiv* (2020).
91. U. Dahal, K. Le, M. Gupta, RNA m6A methyltransferase METTL3 regulates invasiveness of melanoma cells by matrix metallopeptidase 2. *Melanoma Res* **29**, 382-389 (2019).
92. B. Linder *et al.*, Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**, 767-772 (2015).
93. T. Sibbritt, H. R. Patel, T. Preiss, Mapping and significance of the mRNA methylome. *Wiley Interdiscip Rev RNA* **4**, 397-422 (2013).
94. K. Du, L. Zhang, T. Lee, T. Sun, m(6)A RNA Methylation Controls Neural Development and Is Involved in Human Diseases. *Mol Neurobiol* **56**, 1596-1606 (2019).
95. J. Li *et al.*, The role of mRNA m(6)A methylation in the nervous system. *Cell Biosci* **9**, 66 (2019).
96. P. J. Batista *et al.*, m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707-719 (2014).
97. T. Lence *et al.*, m(6)A modulates neuronal functions and sex determination in Drosophila. *Nature* 10.1038/nature20568 (2016).
98. L. Kan *et al.*, The m(6)A pathway facilitates sex determination in Drosophila. *Nat Commun* **8**, 15737 (2017).
99. C. X. Wang *et al.*, METTL3-mediated m6A modification is required for cerebellar development. *PLoS Biol* **16**, e2004880 (2018).
100. H. Huang, J. Camats-Perna, R. Medeiros, V. Anggono, J. Widagdo, Altered Expression of the m6A Methyltransferase METTL3 in Alzheimer's Disease. *eNeuro* **7** (2020).
101. K. D. Meyer, S. R. Jaffrey, The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat Rev Mol Cell Biol* **15**, 313-326 (2014).
102. K. D. Meyer *et al.*, 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **163**, 999-1010 (2015).
103. I. A. Roundtree, C. He, Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Trends Genet* 10.1016/j.tig.2016.03.006 (2016).
104. Z. Zhang *et al.*, METTL3-mediated N(6)-methyladenosine mRNA modification enhances long-term memory consolidation. *Cell Res* **28**, 1050-1061 (2018).
105. R. W. Smith, T. K. Blee, N. K. Gray, Poly(A)-binding proteins are required for diverse biological processes in metazoans. *Biochem Soc Trans* **42**, 1229-1237 (2014).
106. S. Gerstberger, M. Hafner, T. Tuschl, A census of human RNA-binding proteins. *Nature Reviews Genetics* **15**, 829-845 (2014).
107. E. G. Conlon, J. L. Manley, RNA-binding proteins in neurodegeneration: mechanisms in aggregate. *Genes Dev* **31**, 1509-1528 (2017).

108. A. E. Brinegar, T. A. Cooper, Roles for RNA-binding proteins in development and disease. *Brain research* **1647**, 1-8 (2016).
109. J. C. Darnell, J. D. Richter, Cytoplasmic RNA-binding proteins and the control of complex brain function. *Cold Spring Harb Perspect Biol* **4**, a012344 (2012).
110. S. W. Leung *et al.*, Splice variants of the human ZC3H14 gene generate multiple isoforms of a zinc finger polyadenosine RNA binding protein. *Gene* **439**, 71-78 (2009).
111. J. Rha *et al.*, The RNA-binding Protein, ZC3H14, is Required for Proper Poly(A) Tail Length Control, Expression of Synaptic Proteins, and Brain Function in Mice. *Hum Mol Genet* 10.1093/hmg/ddx248 (2017).
112. L. Wan, T. C. Dockendorff, T. A. Jongens, G. Dreyfuss, Characterization of dFMR1, a Drosophila melanogaster homolog of the fragile X mental retardation protein. *Mol Cell Biol* **20**, 8536-8547 (2000).
113. T. Lence *et al.*, m(6)A modulates neuronal functions and sex determination in Drosophila. *Nature* **540**, 242-247 (2016).
114. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
115. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
116. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550-550 (2014).
117. V. K. Mootha *et al.*, PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**, 267-273 (2003).
118. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
119. M. Ashburner *et al.* (2000) Gene ontology: Tool for the unification of biology. (NIH Public Access), pp 25-29.
120. C. The Gene Ontology, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* **47**, D330-D338 (2019).
121. Y. Katz, E. T. Wang, E. M. Airoldi, C. B. Burge, Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**, 1009-1015 (2010).
122. S. Anders, A. Reyes, W. Huber, Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008-2017 (2012).
123. D. A. Harrison, Sex determination: controlling the master. *Curr Biol* **17**, R328-330 (2007).
124. L. O. Penalva, L. Sanchez, RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation. *Microbiol Mol Biol Rev* **67**, 343-359, table of contents (2003).
125. B. Gawande, M. D. Robida, A. Rahn, R. Singh, Drosophila Sex-lethal protein mediates polyadenylation switching in the female germline. *EMBO J* **25**, 1263-1272 (2006).
126. H. Sakamoto, K. Inoue, I. Higuchi, Y. Ono, Y. Shimura, Control of Drosophila Sex-lethal pre-mRNA splicing by its own female-specific product. *Nucleic Acids Res* **20**, 5533-5540 (1992).
127. I. U. Haussmann *et al.*, m(6)A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* 10.1038/nature20577 (2016).
128. L. R. Bell, E. M. Maine, P. Schedl, T. W. Cline, Sex-lethal, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55**, 1037-1046 (1988).

129. J. T. Robinson, H. Thorvaldsdottir, A. M. Wenger, A. Zehir, J. P. Mesirov, Variant Review with the Integrative Genomics Viewer. *Cancer Res* **77**, e31-e34 (2017).
130. D. A. Barbash, T. W. Cline, Genetic and molecular analysis of the autosomal component of the primary sex determination signal of Drosophila melanogaster. *Genetics* **141**, 1451-1471 (1995).
131. L. Sanchez, N. Gorfinkiel, I. Guerrero, Sex determination genes control the development of the Drosophila genital disc, modulating the response to Hedgehog, Wingless and Decapentaplegic signals. *Development* **128**, 1033-1043 (2001).
132. J. I. Horabin, P. Schedl, Regulated splicing of the Drosophila sex-lethal male exon involves a blockage mechanism. *Mol Cell Biol* **13**, 1408-1414 (1993).
133. J. C. Lucchesi, M. I. Kuroda, Dosage compensation in Drosophila. *Cold Spring Harb Perspect Biol* **7** (2015).
134. C. I. Keller, A. Akhtar, The MSL complex: juggling RNA-protein interactions for dosage compensation and beyond. *Curr Opin Genet Dev* **31**, 1-11 (2015).
135. H. Amrein, R. Axel, Genes expressed in neurons of adult male Drosophila. *Cell* **88**, 459-469 (1997).
136. C. Bevan, Roote, J., Russell, S., Ashburner, M., On the allelism of killer-of-male and male-specific lethal mutations. *Drosophila Information Service* **72**, 125 (1993).
137. V. H. Meller, K. H. Wu, G. Roman, M. I. Kuroda, R. L. Davis, roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. *Cell* **88**, 445-457 (1997).
138. R. L. Kelley *et al.*, Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila. *Cell* **81**, 867-877 (1995).
139. R. Moschall, M. Gaik, J. Medenbach, Promiscuity in post-transcriptional control of gene expression: Drosophila sex-lethal and its regulatory partnerships. *FEBS Lett* **591**, 1471-1488 (2017).
140. G. J. Bashaw, B. S. Baker, The msl-2 dosage compensation gene of Drosophila encodes a putative DNA-binding protein whose expression is sex specifically regulated by Sex-lethal. *Development* **121**, 3245-3258 (1995).
141. G. J. Bashaw, B. S. Baker, The regulation of the Drosophila msl-2 gene reveals a function for Sex-lethal in translational control. *Cell* **89**, 789-798 (1997).
142. S. Primus, C. Pozmanter, K. Baxter, M. Van Doren, Tudor-domain containing protein 5-like promotes male sexual identity in the Drosophila germline and is repressed in females by Sex lethal. *PLoS Genet* **15**, e1007617 (2019).
143. R. Ota *et al.*, Transcripts immunoprecipitated with Sxl protein in primordial germ cells of Drosophila embryos. *Dev Growth Differ* **59**, 713-723 (2017).
144. K. J. Morris, A. H. Corbett, The polyadenosine RNA-binding protein ZC3H14 interacts with the THO complex and coordinately regulates the processing of neuronal transcripts. *Nucleic Acids Res* **46**, 6561-6575 (2018).
145. A. Hilfiker, H. Amrein, A. Dubendorfer, R. Schneiter, R. Nothiger, The gene virilizer is required for female-specific splicing controlled by Sxl, the master gene for sexual development in Drosophila. *Development* **121**, 4017-4026 (1995).
146. M. Niessen, R. Schneiter, R. Nothiger, Molecular identification of virilizer, a gene required for the expression of the sex-determining gene Sex-lethal in Drosophila melanogaster. *Genetics* **157**, 679-688 (2001).

147. P. Mitchell, E. Petfalski, A. Shevchenko, M. Mann, D. Tollervey, The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell* **91**, 457-466 (1997).
148. C. Allmang *et al.*, The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. *Genes Dev* **13**, 2148-2158 (1999).
149. V. Boczonadi *et al.*, EXOSC8 mutations alter mRNA metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. *Nat Commun* **5**, 4287 (2014).
150. N. Di Donato *et al.*, Mutations in EXOSC2 are associated with a novel syndrome characterised by retinitis pigmentosa, progressive hearing loss, premature ageing, short stature, mild intellectual disability and distinctive gestalt. *J Med Genet* **53**, 419-425 (2016).
151. J. Wan *et al.*, Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. *Nat Genet* **44**, 704-708 (2012).
152. P. Mitchell, E. Petfalski, D. Tollervey, The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev* **10**, 502-513 (1996).
153. Q. Liu, J. C. Greimann, C. D. Lima, Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**, 1223-1237 (2006).
154. D. L. Makino, E. Conti, Structure determination of an 11-subunit exosome in complex with RNA by molecular replacement. *Acta Crystallogr D Biol Crystallogr* **69**, 2226-2235 (2013).
155. M. W. Briggs, K. T. Burkard, J. S. Butler, Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. *J Biol Chem* **273**, 13255-13263 (1998).
156. D. Hou, M. Ruiz, E. D. Andrulis, The ribonuclease Dis3 is an essential regulator of the developmental transcriptome. *BMC Genomics* **13**, 359 (2012).
157. S. J. Lim, P. J. Boyle, M. Chinen, R. K. Dale, E. P. Lei, Genome-wide localization of exosome components to active promoters and chromatin insulators in Drosophila. *Nucleic acids research* **41**, 2963-2980 (2013).
158. X. Yang *et al.*, Genetic and genomic studies of pathogenic EXOSC2 mutations in the newly described disease SHRF implicate the autophagy pathway in disease pathogenesis. *Hum Mol Genet* 10.1093/hmg/ddz251 (2019).
159. E. Pefanis *et al.*, Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* **514**, 389-393 (2014).
160. M. B. Fasken *et al.*, The RNA Exosome and Human Disease. *Methods Mol Biol* **2062**, 3-33 (2020).
161. J. S. Muller, M. Giunta, R. Horvath, Exosomal Protein Deficiencies: How Abnormal RNA Metabolism Results in Childhood-Onset Neurological Diseases. *J Neuromuscul Dis* **2**, S31-S37 (2015).
162. J. Wan *et al.*, Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. *Nat Genet* **44**, 704-708 (2012).
163. V. R. Eggens *et al.*, EXOSC3 mutations in pontocerebellar hypoplasia type 1: novel mutations and genotype-phenotype correlations. *Orphanet J Rare Dis* **9**, 23 (2014).
164. S. Rudnik-Schoneborn *et al.*, Pontocerebellar hypoplasia type 1: clinical spectrum and relevance of EXOSC3 mutations. *Neurology* **80**, 438-446 (2013).

165. A. Halevy *et al.*, Novel EXOSC3 mutation causes complicated hereditary spastic paraplegia. *J Neurol* **261**, 2165-2169 (2014).
166. G. Zanni *et al.*, Exome sequencing in a family with intellectual disability, early onset spasticity, and cerebellar atrophy detects a novel mutation in EXOSC3. *Neurogenetics* **14**, 247-250 (2013).
167. T. van Dijk, F. Baas, P. G. Barth, B. T. Poll-The, What's new in pontocerebellar hypoplasia? An update on genes and subtypes. *Orphanet J Rare Dis* **13**, 92 (2018).
168. M. B. Fasken *et al.*, Insight into the RNA Exosome Complex Through Modeling Pontocerebellar Hypoplasia Type 1b Disease Mutations in Yeast. *Genetics* **205**, 221-237 (2017).
169. A. Gillespie, J. Gabunilas, J. C. Jen, G. F. Chanfreau, Mutations of EXOSC3/Rrp40p associated with neurological diseases impact ribosomal RNA processing functions of the exosome in S. cerevisiae. *RNA* **23**, 466-472 (2017).
170. J. S. de Belle, M. Heisenberg, Associative odor learning in Drosophila abolished by chemical ablation of mushroom bodies. *Science* **263**, 692-695 (1994).
171. T. Lee, A. Lee, L. Luo, Development of the Drosophila mushroom bodies: sequential generation of three distinct types of neurons from a neuroblast. *Development* **126**, 4065-4076 (1999).
172. J. R. Crittenden, E. M. Skoulakis, K. A. Han, D. Kalderon, R. L. Davis, Tripartite mushroom body architecture revealed by antigenic markers. *Learn Mem* **5**, 38-51 (1998).
173. Y. Q. Zhang *et al.*, Drosophila fragile X-related gene regulates the MAP1B homolog Futsch to control synaptic structure and function. *Cell* **107**, 591-603 (2001).
174. J. C. Zinder, C. D. Lima, Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* **31**, 88-100 (2017).
175. C. Belair *et al.*, The RNA exosome nuclease complex regulates human embryonic stem cell differentiation. *J Cell Biol* 10.1083/jcb.201811148 (2019).
176. D. S. Mistry, Y. Chen, G. L. Sen, Progenitor function in self-renewing human epidermis is maintained by the exosome. *Cell Stem Cell* **11**, 127-135 (2012).
177. S. C. McIver *et al.*, The exosome complex establishes a barricade to erythroid maturation. *Blood* **124**, 2285-2297 (2014).
178. D. L. Kiss, E. D. Andrulis, Genome-wide analysis reveals distinct substrate specificities of Rrp6, Dis3, and core exosome subunits. *RNA* **16**, 781-791 (2010).
179. R. Biancheri *et al.*, EXOSC3 mutations in isolated cerebellar hypoplasia and spinal anterior horn involvement. *J Neurol* **260**, 1866-1870 (2013).
180. A. P. Di Giovambattista, I. Jacome Querejeta, P. Ventura Faci, G. Rodriguez Martinez, F. Ramos Fuentes, [Familial EXOSC3-related pontocerebellar hypoplasia]. *An Pediatr (Barc)* **86**, 284-286 (2017).
181. V. R. Eggens *et al.*, EXOSC3 mutations in pontocerebellar hypoplasia type 1: novel mutations and genotype-phenotype correlations. *Orphanet J Rare Dis* **9**, 23 (2014).
182. S. Bizzari, A. R. Hamzeh, M. Mohamed, M. T. Al-Ali, F. Bastaki, Expanded PCH1D phenotype linked to EXOSC9 mutation. *Eur J Med Genet* 10.1016/j.ejmg.2019.01.012 (2019).
183. D. T. Burns *et al.*, Variants in EXOSC9 Disrupt the RNA Exosome and Result in Cerebellar Atrophy with Spinal Motor Neuronopathy. *Am J Hum Genet* **102**, 858-873 (2018).
184. P. E. Howse, Brain structure and behavior in insects. *Annu Rev Entomol* **20**, 359-379 (1975).

185. E. V. Wasmuth, J. C. Zinder, D. Zattas, M. Das, C. D. Lima, Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that Mpp6 stimulates RNA decay and recruits the Mtr4 helicase. *Elife* **6** (2017).
186. S. Falk, F. Bonneau, J. Ebert, A. Kogel, E. Conti, Mpp6 Incorporation in the Nuclear Exosome Contributes to RNA Channeling through the Mtr4 Helicase. *Cell Rep* **20**, 2279-2286 (2017).
187. E. Lorentzen, A. Dziembowski, D. Lindner, B. Seraphin, E. Conti, RNA channelling by the archaeal exosome. *EMBO Rep* **8**, 470-476 (2007).
188. J. C. Zinder, E. V. Wasmuth, C. D. Lima, Nuclear RNA Exosome at 3.1 A Reveals Substrate Specificities, RNA Paths, and Allosteric Inhibition of Rrp44/Dis3. *Mol Cell* **64**, 734-745 (2016).
189. I. Ivanov *et al.*, Pontocerebellar hypoplasia type 1 for the neuropediatrician: Genotype-phenotype correlations and diagnostic guidelines based on new cases and overview of the literature. *Eur J Paediatr Neurol* **22**, 674-681 (2018).
190. J. Schwabova *et al.*, Homozygous EXOSC3 mutation c.92G-->C, p.G31A is a founder mutation causing severe pontocerebellar hypoplasia type 1 among the Czech Roma. *J Neurogenet* **27**, 163-169 (2013).
191. F. S. Collins, M. Morgan, A. Patrinos, The Human Genome Project: lessons from large-scale biology. *Science* **300**, 286-290 (2003).
192. J. C. Biro, Seven fundamental, unsolved questions in molecular biology. Cooperative storage and bi-directional transfer of biological information by nucleic acids and proteins: an alternative to "central dogma". *Med Hypotheses* **63**, 951-962 (2004).
193. S. Bian, T. L. Xu, T. Sun, Tuning the cell fate of neurons and glia by microRNAs. *Curr Opin Neurobiol* **23**, 928-934 (2013).
194. J. Najm *et al.*, Mutations of CASK cause an X-linked brain malformation phenotype with microcephaly and hypoplasia of the brainstem and cerebellum. *Nat Genet* **40**, 1065-1067 (2008).
195. W. Traut, T. Niimi, K. Ikeo, K. Sahara, Phylogeny of the sex-determining gene Sex-lethal in insects. *Genome* **49**, 254-262 (2006).
196. B. S. Baker, Sex in flies: the splice of life. *Nature* **340**, 521-524 (1989).
197. T. Alpert, K. Straube, F. Carrillo Oesterreich, K. M. Neugebauer, Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns. *Cell Rep* **33**, 108324 (2020).
198. S. Fusi, E. K. Miller, M. Rigotti, Why neurons mix: high dimensionality for higher cognition. *Curr Opin Neurobiol* **37**, 66-74 (2016).
199. A. L. Rockwell, C. F. Hongay, The m(6)A Dynamics of Profilin in Neurogenesis. *Front Genet* **10**, 987 (2019).
200. E. Afgan *et al.*, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* **46**, W537-W544 (2018).
201. G. dos Santos *et al.*, FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res* **43**, D690-697 (2015).
202. A. D. Yates *et al.*, Ensembl 2020. *Nucleic Acids Res* **48**, D682-D688 (2020).
203. H. Wickham (2016) ggplot2 : Elegant Graphics for Data Analysis. in *Use R!,* (Springer International Publishing : Imprint: Springer,, Cham), pp 1 online resource (XVI, 260 pages 232 illustrations, 140 illustrations in color.
204. K. Slowikowski (2019) ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'.

205. R. C. Team (2019) R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria).
206. R. S. Team (2018) RStudio: Integrated Development Environment for R. (RStudio, Inc., Boston, MA).
207. J. Thurmond *et al.*, FlyBase 2.0: the next generation. *Nucleic Acids Res* **47**, D759-D765 (2019).
208. J. A. Powell, GO2MSIG, an automated GO based multi-species gene set generator for gene set enrichment analysis. *BMC Bioinformatics* **15**, 146 (2014).
209. S. Carbon *et al.*, AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288-289 (2009).
210. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
211. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018 (2011).
212. J. S. Wu, L. Luo, A protocol for dissecting Drosophila melanogaster brains for live imaging or immunostaining. *Nat Protoc* **1**, 2110-2115 (2006).
213. G. Grenningloh, E. J. Rehm, C. S. Goodman, Genetic analysis of growth cone guidance in Drosophila: fasciclin II functions as a neuronal recognition molecule. *Cell* **67**, 45-57 (1991).
214. S. J. Gratz, J. Wildonger, M. M. Harrison, K. M. O'Connor-Giles, CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. *Fly (Austin)* **7**, 249-255 (2013).
215. S. J. Gratz *et al.*, Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* **196**, 961-971 (2014).
216. J. H. Bauer, S. Goupil, G. B. Garber, S. L. Helfand, An accelerated assay for the identification of lifespan-extending interventions in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **101**, 12980-12985 (2004).
217. E. D. Andrulis *et al.*, The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. *Nature* **420**, 837-841 (2002).
218. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
219. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
220. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
221. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).