University of Nevada, Reno

Regulation and Function of the *Dscam1* Extended 3' UTR mRNA

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Neuroscience

by

Ryan M. Peterson

Dr. Pedro Miura/Thesis Advisor

December, 2017

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

RYAN M. PETERSON

entitled

Regulation and Function of the *Dscam1* Extended 3' UTR mRNA

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Pedro Miura, Ph.D., Advisor

Thomas Kidd, Ph.D., Committee Member

Ian Wallace, Ph.D., Graduate School Representative

David W. Zeh, Ph.D., Dean, Graduate School

December, 2017

Abstract

More than 50% of genes in species from *Drosophila* to human undergo alternative polyadenylation (APA). It has been shown that neuronal tissues are highly enriched for usage of transcripts with extended alternative 3' untranslated regions, which are products of APA. Specific function of these longer 3' UTR transcripts are largely unknown; however, because the 3' UTR can regulate transcript stability, translational efficiency, and localization, it is plausible that 3' UTR switching in nervous tissues may serve important biological functions. We chose to study the extended 3' UTR of *Dscam1* as it is one of several axon guidance genes that express an extended 3' UTR mRNA isoform. We created a *Dscam1* extended 3' UTR mutant using the relatively new CRISPR/CAS9 system. We show that the *Dscam1* extended 3' UTR loss of function mutants have severe deficiencies in locomotor activity and display an increased mortality rate. We have identified that in S2 cells the RNA binding protein ELAV promotes use of the extended 3' UTR of *Dscam1* and also show that ELAV is capable of binding at least two putative binding sites located on the 3' UTR of *Dscam1.* Elucidating the important function of these extended 3' UTRs will benefit not only our knowledge of axon guidance, but may reveal the function of a pervasive neural phenomenon.

Acknowledgements

I would first like to thank my advisor Dr. Miura for the opportunity to be a part of his research group. The knowledge and skillset I have developed in his lab is an invaluable resource that will continue to benefit me as I move forward with my scientific career. Apart from the science, Dr. Miura has helped develop my ability to ask the right questions and taught me how to seek the answers.

I would like to thank my committee for their tremendous help and insight throughout my graduate career. Dr. Thomas Kidd's expertise in axon guidance and genetics was a great asset to me throughout my project. I would like to give a special thanks to Dr. Ian Wallace and Dr. Kelly Phelps for helping express and purify the ELAV protein used in the EMSA experiments.

I would also like to thank everyone in the Miura lab group, with a special thanks going out to Henry Ng. Henry was a huge help in many of the assays performed in this thesis. I would also like to give a special thanks to Zhiping Zhang for her hard work and advancing this project forward. I have made great friends here in the biology department, and without their support and aid, I wouldn't have made it this far.

Finally, I would like to give tremendous thanks my parents and my girlfriend for their support in all my endeavors. My family has always driven me to do my best in life and without them I would not be the person I am today.

This work would have not been possible without funding from the NSF (Award number 1656463 and the COBRE (grant P20GM103650).

Table of Contents

List of Figures

Chapter One: Introduction

1. 1 Regulation of Gene Expression

A gene is a discrete segment of DNA that encodes the sequence of a polypeptide or functional RNA [1]. Given that nearly all mitotic cells of an organism are genetically identical, it is primarily through differential regulation of gene expression that functional and phenotypic differences arise between diverse tissue types [2, 3]. At the most fundamental level, gene expression can be described as transcription of DNA to RNA, followed by translation of RNA to protein [4]. Regulation of gene expression is essential for proper cellular function and collectively refers to any mechanism that influences expression levels or spatio-temporal expression patterns of gene products. Gene expression can be regulated by a wide variety of processes including: chromatin remodeling, histone modification, DNA methylation status, transcription factors, co/post-transcriptional regulation of RNA, translation, and post-translational modifications. Together, these processes account for the majority of all eukaryotic gene expression.

A typical eukaryotic protein coding gene consists of a 5' untranslated region (5' UTR), one or more exons, one or more introns, and a 3' untranslated region (3' UTR) [5]. Once the precursor messenger RNA (pre-mRNA) is transcribed by RNA polymerase II (RNA Pol II), it must undergo co/post transcriptional modifications (5' capping, splicing, and 3' end processing) to generate the final mature transcript [6]. These modifications not only influence overall gene expression, but can also generate gene products (usually protein or non-coding RNA) with differing function [7].

1.2 Tissue-Specific Transcriptional Regulation Through Transcription Factors

In human, there are more than 1,400 different sequence-specific transcription factors and many are expressed in a signal- or tissue-specific manner [8, 9]. This allows

for tightly controlled expression of genes in different tissues despite every cell type containing the same DNA cis-regulatory sequences.

Transcription factors promote transcriptional activation upon binding their corresponding cis-regulatory sequences, which are located within promoter regions and enhancer elements of genes [10]. Unlike the promoter, which is often located just upstream of the genes coding region, enhancers can be located tens to hundreds of kilobases away from the gene they regulate; despite this, enhancers heavily influence transcriptional activation. This is because DNA is able to loop out extra sequence in order to bring the enhancer elements spatially close to the promoter. In contrast to enhancers, silencers are enhancer –like motifs capable of repressing transcription.

1.3 The 3' UTR: A Platform for Post-transcriptional Regulation

Aside from transcription factors, which are the primary determinates of gene activity, post transcriptional regulation allows attenuation of expression. Post-transcriptional gene regulation refers to a collection of regulatory mechanisms that take place at the RNA level. In particular, the binding of microRNAs (miRNAs), and RNA binding proteins (RBPs) account for a large portion of post-transcriptional regulation. While both miRNAs and RBPs are able to bind many regions of a transcript, binding of these elements frequently occurs at the 3' UTR.

The 3' UTR of a transcript is heavily enriched for regulatory sequence motifs and secondary structure that miRNA and RBPs recognize and act on [11, 12]. This enrichment of regulatory motifs is what makes the 3' UTR a powerful regulatory element, able to influence transcript stability, localization, and translational efficiency [13-16]. The 3' UTR is found immediately downstream of the gene's coding region. Its length, and therefore the motifs it contains, is determined after cleavage and polyadenylation occurs.

1.4 Post-transcriptional Regulatory Elements

MicroRNAs are a pervasive class of short non-coding regulatory RNAs with more than 200 in fly and 1,000 validated in human [17, 18]. miRNAs are able to repress their target genes by binding to a specific target site complementary to the miRNAs seed region. While some miRNAs are ubiquitously expressed, many are able to silence genes in a tissue-specific manner. This can happen through transcript deadenylation, destabilization, and translational repression [19]. miRNA target sites are most often located within the 3' UTR region of a mRNA transcript and a given transcript may be under the control of multiple miRNAs. Collectively, miRNAs are predicted to regulate expression of at least 50% genes in the human genome, if not more [20]. Several studies have shown some miRNAs to be critical for proper development in *C. elegans*, bovine, and zebrafish [21-23]. In contrast to this, another study deleted whole families of miRNAs in *C. elegans* with little effect, suggesting some miRNAs may be redundant or only important in the right biological context [24, 25]. Regardless, miRNAs fine tune gene expression in many organisms from fly to human [26, 27].

RNA binding proteins (RBPs) are a large class of proteins capable of interacting with RNA to influence many aspects of a transcripts fate. Collectively, RBPs exhibit a plethora of functions by serving as localization signals, acting as scaffolds, tethers, nucleases, RNA editors, and binding competitors. RBPs are composed of modular recognition motifs and enzymatic domains separated by linker sequences. RBPs utilize these domains in a combinatorial fashion, giving rise to a plethora of RBPs with different target specificity and function [28].

1.5 Cleavage and Polyadenylation

In eukaryotic organisms, most nascent mRNA must first undergo 5' capping, splicing, and 3' end processing before it is considered a mature mRNA [29, 30]. 3' end processing is a concerted co-transcriptional event referred to as cleavage and polyadenylation, which consists of endonucleolytic cleavage of the transcript followed by the addition of a homopolymeric sequence of adenosines via poly(A) polymerase (PAP). The poly(A) tail, which in mammals has been shown to be roughly ~250 nucleotides in length, serves as a nuclear export signal, influences transcript stability, and affects translational efficiency [29, 31, 32].

Cleavage and polyadenylation in mammals is a relatively well understood process that is carried out by the cleavage and polyadenylation complex which is comprised of protein cleavage/polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), symplekin, cleavage factor I (CFI), cleavage factor II (CFII), PAP, poly(A) binding protein (PABP) in addition to a myriad of other accessory proteins, some of which have yet to be identified (Figure 1) [33, 34]. Out of these, CPSF and CstF constitute the two most important proteins of the 3' cleavage machinery, consisting of 6 and 3 subunits, respectively [35]. Subunits of both CPSF and CstF are able to interact with the C terminal Domain (CTD) of the large subunit of RNA polymerase II (RNAP II) [36]. Together, CPSF, CstF, and the CTD of RNAP II act as the scaffold for the rest of the 3' processing factors [37]. It is in this fashion that the cleavage machinery complex is able to processively scan the nascent mRNA in a co-transcriptional fashion for its target sequences.

Specific constituents of the cleavage machinery complex recognize conserved sequence motifs located at the 3' terminus of the pre-mRNA. The polyadenylation signal (PAS) is the most notable motif, consisting of a highly conserved hexameric sequence

A[A/U]UAAA, and although multiple sequence variants of the PAS exist, this is by far the most common PAS in mice and human [38, 39]. It was long thought that CPSF-160 (the largest subunit of CPSF) was the primary component in PAS recognition; however, new evidence has indicated that Wdr33 and CPSF-30 (two subunits of CPSF) may in fact have the highest affinity for the A(A/U)UAAA element according to recent PAR-CLIP assays [40, 41]. The PAS is generally located 11-23 nucleotides upstream of the cleavage site, often a CA dinucleotide, where the β-CASP domain of CPSF-73 will cleave the transcript [42-44].

In addition to the PAS, mammalian pre-mRNAs have been shown to often, but not always, contain a downstream sequence element (DSE) as well as an upstream sequence element (USE) at the 3' end of the transcript. These elements greatly enhance the activity of the cleavage and polyadenylation machinery [42]. The DSE consists of a U/GU-rich element within 30 nucleotides 3' of the cleavage site that is recognized by CstF-64 (one of the three subunits that make up CstF) [45-47]. A typical USE consists of a UGUA or U-rich motif that is recognized and bound by CFI, which helps stabilize CPSF to the PAS [48]. The USE is especially important in cases where a canonical PAS is absent and recruitment of CPSF is solely dependent on CFI [49]. Collectively, the PAS with the USEs and DSEs define the poly(A) site of nascent mRNA and allow for specificity by the cleavage and polyadenylation complex.

Upon CPSF, CstF, and CFI binding their appropriate sequence elements, the remainder of cleavage factors such as CFII and PAP that compose the cleavage machinery are recruited. The completed complex induces the CPSF-73 subunit to cleave the transcript [50]. At this point PAP begins polyadenylating the 3' end and PABP binds the elongating poly(A) tail which promotes faster elongation by PAP [51]. When the poly(A) tail reaches between 250 and 300 nucleotides long elongation is terminated and

the cleavage and polyadenylation machinery becomes completely dissociated from the transcript, which is then exported to the cytoplasm [51].

1.6 Alternative cleavage and Polyadenylation

The most recent studies estimate that the human genome contains roughly 19,000 protein coding genes [52]. Comparatively, the *Mus musculus* genome is estimated to encode slightly over 20,000 protein coding genes, while *drosophila melanogaster* encodes approximately 14,000, and *C. elegans* encode more than 19,000 protein coding genes [53-55]. The differences in phenotypic complexity arisen between species cannot be explained by number of protein coding genes alone. It is now thought that differences in complexity arise from regulation via non-coding DNA as well as mechanisms including DNA methylation, histone modification, and co/post-transcriptional modification [56-58].

One form of co-transcriptional modification is alternative cleavage and polyadenylation (APA), which allows formation of gene transcripts with varying 3' UTR lengths (Figure 2). 3' UTR APA occurs when more than one polyadenylation site exists within the 3' end of a particular gene. 3' UTR APA is the most prevalent type of APA and generates transcripts that code for the same protein, while only changing the length of the 3' UTR used [59]. Cleavage machinery can cleave at the more proximal polyadenylation site, generating a short 3' UTR isoform of that gene, or at the more distal site, creating a long 3' UTR isoform of that gene. As previously discussed, the 3' UTR harbors many regulatory elements, and by altering the 3' UTR being used, the repertoire of trans-acting elements changes, which can cause dramatic changes in regulation.

Messenger RNA transcripts with alternative 3' UTRs were first discovered in 1980, however, it was not until expressed sequence tag data analysis, and later the advent of RNA-Seq, that the extent of APA was revealed genome-wide [38, 60-64]. About 70% of human genes are subject to APA, while about 50% of genes in fly and worm have been observed to undergo APA [65-67].

1.7 Mechanisms of APA

There are several hypotheses regarding how a particular polyadenylation site is chosen. The proximal polyadenylation site of multi-UTR genes tend to be weaker than their more distal counterparts [68]. A weaker site generally means that the PAS in context of its USE and DSE does not exhibit as high of a binding affinity for the cleavage machinery. This finding is consistent with the observation that when the abundance of some cleavage factors, such as CstF64, are increased, a shift to the usage of the weaker, more proximal polyadenylation site occurs (the converse is also true) [69]. However this trend does not apply to all cleavage factors. It was found that an increase in CFI expression leads to usage of the more distal site [70].

It has also been shown that RNAP II kinetics can influence polyadenylation site choice [71]. In a *C4* fly mutant where RNAP II elongation rate is nearly half, it was seen that the proximal polyadenylation site for the gene *polo* was used preferentially when compared to the wild-type fly [72]. Additionally, this study demonstrates relevance for physiological splice site choice since flies mutant for the extended 3' UTR isoform of *polo* are lethal.

Another proposed mechanism suggests that RBPs are able to bind nascent transcripts co-transcriptionally to either competitively inhibit cleavage factors from binding to a particular polyadenylation site. In fly, one example of this is seen with the

pan-neuronally expressed RBP embryonic lethal abnormal vision (ELAV) [73]. The binding of ELAV to the nascent transcript causes a shift toward the use of the longer 3' UTR isoform of many genes. Similarly, other studies show that Sex-lethal (SXL), and PABPN1 also can competitively bind sequence motifs near the polyadenylation site to prevent cleavage at proximal sites [74, 75]. In contrast to this, RBPs such as cytoplasmic polyadenylation element binding protein 1 (CPEB1) and heterogeneous nuclear ribonucleoprotein H (hnRNP H1) have been shown to bind near the proximal polyadenylation site to recruit cleavage factors to preferentially generate short 3' UTR isoforms of transcripts [76, 77].

There is no reason to believe that any one of these regulatory mechanisms is exclusive. The regulation of polyadenylation site choice may be under the control of several regulatory pathways concurrently.

1.8 APA in different Tissues

While the exact function of the extended 3' UTR isoform of many genes is poorly understood, it does not appear to be stochastic [78]. Proliferative cells, as well as oncogenic cells, tend to express shorter 3' UTR isoforms of genes in contrast to differentiated healthy cells [79, 80]. Intuitively, it is thought that shorter 3' UTR variants promote higher expression of protein via escape of negative regulatory elements such as micro RNA target sites located within longer 3' UTRs [80, 81]. However there are many positive regulatory elements that may be contained within the 3' UTR as well and 3' UTR length has been shown to be weakly correlated with mRNA stability [82, 83]. Polyadenylation site choice among transcripts appears to be relatively conserved amongst tissue types, even between species [65, 66]. Tissue-specific APA of transcripts

allows ubiquitous expression of a gene, while maintaining the ability to tailor the final product to the tissue-type.

1.9 Neural APA

In *Drosophila*, it has been shown that there is an enrichment of extended 3' UTR isoforms of genes being used in neural-tissues, consistent with studies in mammals[84-86]. Additionally, 383 extended 3' UTR transcript isoforms were found to be expressed exclusively in fly head [66]. This stands in contrast with 3' UTR shortening events that appear to be enriched in proliferative tissues such as testis [66]. One hypothesis explaining this enrichment argues that a complex tissue, like neurons, may require additional regulation of gene expression. This regulation can in part be provided by transcripts with longer 3' UTRs.

Neurons are very polarized cells, some with axons that can be many times longer than the diameter of the cell body. Localizing transcripts before translation may be more energetically favorable than transporting the large amount of the protein. For example, one study shows that brain-derived neurotrophic factor (BDNF) is capable of undergoing APA to generate a long 3' UTR isoform that localizes specifically to dendrites [87]. Similarly (RanBP1) was shown to have a long 3' UTR variant that was able to promote axonal localization of a reporter construct [88].

1.10 ELAV: A regulator of 3' UTR extension

Embryonic lethal abnormal vision (ELAV) is a pan-neuronally expressed *Drosophila* RBP that has long been known to influence alternative splicing of genes such as *neuroglian* (*nrg)*, *erect wing (ewg)*, and *armadillo (arm) [89, 90]*. More recently than

this, it has also been found that ELAV is a regulator of APA in the central nervous system (CNS) [66, 73].

ELAV is a nuclear localized protein first expressed around embryonic stage 11, when the *Drosophila* nervous system is forming [91]. It is known that ELAV and its human orthologs (HuB, HuC, HuD, and HuR) contain 3 RNA recognition motifs that bind U-rich sequences of mRNA transcripts [92-94].

It has been suggested that ELAV binding U-rich elements within close proximity to the poly(A) site may block cleavage factors from binding there, causing RNAPII read-through and cleavage to occur at the more distal site [73, 95]. A second hypothesis proposes that ELAV stabilizes extended transcripts by binding U-rich elements found throughout the UTR and without ELAV binding the transcript becomes unstable and is degraded [96].

E*lav* mutants are embryonic lethal, but survive into early stages of neural development. One study showed that ELAV mutants exhibit severe axon guidance defects at the embryonic ventral nerve cord (VNC) [97]. Furthermore, the group shows that the *Elav* phenotype seen at the VNC is similar in manner to a *comm* full null mutant, a gene involved in axon guidance that expresses an extended 3' UTR isoform. Interestingly, several genes involved in axon guidance including *Dscam1*, c*omm*, and *Fas1*, express extended 3' UTR isoforms. This led to the hypothesis that the extended 3' UTR isoforms of these genes are necessary for proper embryonic neural development, and more specifically, axon guidance.

1.11 Axon Guidance

In the developing CNS of an organism, neurons project axons that must often traverse significant lengths to innervate their synaptic targets in a reliable manner [98]. The axonal growth cone is the fan-shaped axonal structure located at the tip of the elongating axon that is responsible for detecting and responding to extracellular guidance cues [99]. Guidance cues are primarily detected by receptors expressed at the cell-surface of the developing neuron and can either trigger an attractive or repulsive response [100]. Guidance molecules include adhesive glycoproteins that interact with the neuron as it navigates through its substrate, as well as diffusible ligands, which can be detected over long-distances.

In chick embryonic retina, it was shown that adhesive molecules expressed in the extracellular matrix (ECM), such as the glycoprotein laminin, are required to direct developing axons toward their target [101]. In the zebrafish midbrain, it was further shown that laminins in the ECM are bound by integrin receptors located on the growth cone surface that direct the initial outgrowth of neurons [102]. In addition to laminin, transiently expressed axonal glycoproteins (TAGs) such as TAG-1 have been shown to be important for adhesion of axons to other developing axons [102, 103].

Apart from these “local” interactions, several secreted ligands have been identified which provide a concentration gradient for longer range signaling. This process, known as chemotropism, had long been postulated to be a guidance mechanism for developing neurons. The hypothesis was confirmed in an *in vitro* assay using rat spinal explants [104]. It was shown that commissural axons from the spinal explants would exhibit growth toward nearby floor-plate explants. At the time, the secreted factors were unknown, but since then, multiple guidance molecules have been

identified [98]. These include the netrins, slits, semaphorins, and ephrins, along with their respective receptors, DCC, Robo, the plexin receptors, and eph [105-108]. In recent years it has been shown that some these ligands, such as netrin and slit, can bind multiple receptors[109, 110].

1.12 *Dscam1*: The Down syndrome cell adhesion molecule

The *Drosophila* Down syndrome cell adhesion molecule (*Dscam1*) is a transmembrane receptor that plays an important role in neurite self-avoidance, axon guidance, and maintenance of neural circuits [111]. While the gene itself belongs to the immunoglobulin superfamily, its role in immune function is less studied compared to its involvement in neural development. The gene name stems from the chromosome region in which the human homolog, *DSCAM,* is located. In human, the *DSCAM* locus is found in a region of chromosome 21 known as the "Down syndrome critical region," a region containing a group of genes overexpressed in trisomy 21 [112].

Dscam1 is the most extensively alternatively spliced gene known, capable of producing 304,128 unique transcript isoforms, and more than 38,000 different proteins, assuming every alternative splicing combination is possible [113, 114]. The primary source of this transcript diversity is attributed to alternative splicing of 3 hyper-variable exon clusters, 4, 6, and 9, containing 12, 48, and 33 alternative exons respectively. Additionally, exon 17 can be alternatively spliced, and exons 19 and 23 can be spliced out completely [114]. To further add to *Dscam1's* transcript diversity, the *Dscam1* gene is also capable of producing an extended 3' UTR. While I suggest more than 300,000 transcript isoforms are possible, determining how many exist *in vivo* is difficult due to read length limitations of traditional RNA-seq technology [115]. Using long-read sequencing with the MinIon nanopore sequencer, one group was able to confirm at least

7,874 unique splice-forms; however, this study only addressed splice patterns of the exon 4, 6, and 9 clusters, suggesting many more variants likely exist [113].

The Dscam protein contains 10 immunoglobulin domains and 6 fibronectin type III domains. The exon 4,6, and 9 clusters are responsible for coding the immunoglobulin 2, 3, and 7 domains [116]. These immunoglobulin domains are crucial for proper binding with other Dscam molecules and ligands; therefore, it is crucial proper splicing occurs for *Dscam1* to maintain proper function (Figure 3). These three exons are each spliced in a mutually exclusive manner, ensuring that only one exon from each group is included in the mature transcript.

Interestingly, multiple mechanisms may be working together to regulate mutualistic splicing in the exon 4, 6, and 9 exon clusters [117]. It was found that a highly conserved sequence motif (called the docking site) exists between the constant exon 5 and the most proximal alternative exon 6 in 16 different Dipteran species, including *Drosophila* melanogaster. Moreover, the same study identified highly conserved sequences (termed the selector sequences) that resided in the introns between the alternative exons of the exon 6 cluster [118]. It is proposed that through complementary binding of the selector sequence to the docking site spatially arranges the pre-mRNA into an orientation that allows the spliceosome to include exactly one exon 6 variant into the final transcript. Ten years later, it was found that a similar mechanism is acting at the exon 4 and 9 variable regions [119]. Additionally, a RNA secondary structure (called the iStem) 18nt downstream of exon 3 was found to be necessary for exon 4 inclusion [120]. While both of these mechanisms explain how mutually exclusive splicing of *Dscam1* alternative exons might be regulated, it does not describe how specific exons are chosen to be included in the mature mRNA.

Dscam1 splice choice appears to be, at least in part, developmentally regulated. Although most alternative exons in the exons 4, 6, and 9 clusters were found to be expressed throughout development, a study using *Dscam1* microarrays revealed a biased expression profile for certain exons at different developmental time-points [121]. The study also demonstrated that individual R3/R4 photoreceptor cells, and even individual S2 cells, each express between 14 and 50 variants of *Dscam1.* Furthermore, while individual cells express unique sets of *Dscam1* splice variants, similar groups of tissues tend to express a unique repertoire of *Dscam1* alternative exons [121]. Taken together, this suggests *Dscam1* splicing is at least in part, a developmentally regulated process, and that expression of multiple Dscam isoforms confers neurons unique cell identity at the cell surface. The exact mechanism determining which *Dscam1* isoforms are expressed in specific neurons remains unknown [122].

Dscam1 induces contact-dependent repulsion of neurites through homophilic binding of the Dscam ectodomain, requiring the 3 variable immunoglobulin domains to be identical on both Dscam receptors [123]. Repulsion through homophilic-binding is responsible for correct morphogenesis of many types of neurons, and relies on neighboring neurons expressing unique sets of *Dscam1* isoforms [122, 124].

Proper dendritic arborization of sensory neurons (dA neurons) in the larval body wall was found to be reliant on Dscam homophilic-binding [123]. *Dscam1* loss of function mutants result in a collapse of the dendritic arbor and self-crossing of sister dendrites. In mice, DSCAM mutants show excessive branching and shortening of dendritic spines of pyramidal cells in the developing cortex [125].

Dscam1 function is equally important for proper axon development as it is for dendritic arborization. *Dscam1* is required for proper segregation of axons in the *Drosophila* mushroom body, a brain structure heavily implicated in olfaction and possibly

sensory integration [126, 127]. In the same study, it was also observed that individual neurons of the ellipsoid body that were limited to expressing a single isoform of *Dscam1* failed to properly arborize [126].

To summarize, *Dscam1* is critical for proper morphogenesis of both dendritic and axonal branches in several different neural tissues. Individual neurons express tens of *Dscam1* splice variants which allow individual neurons to recognize sister branches, and the overall diversity of splice-forms possible allows one neuron to distinguish itself from other neurons. One study looked to investigate how much diversity is required for self-avoidance and neural identity mechanisms to occur [128]. Strikingly, they also found that in order to get proper mushroom body formation, which requires thousands of neurons to bifurcate at a specific developmental time-point, between 1,100 and 4,800 isoforms are required. This data suggests that different tissue types require different amount of *Dscam1* diversity to maintain proper function, but it does not directly address if specific isoforms play unique roles in certain tissues.

In the same study, it was confirmed that expression of a single *Dscam1* isoform was sufficient to rescue dendritic collapse of body-wall neurons that normally do not overlap; however, expression of single isoform *Dscam1* caused ectopic repulsion of class I and III neurons which normally display overlapping fields of dendrites. Using *Dscam1* mutant flies that allow expression of different amounts of *Dscam1* isoforms (between ten and thousands), they found that somewhere between 12 and 24 *Dscam1* isoforms was sufficient to rescue both phenotypes to be indistinguishable from wild type in the body wall.

Interestingly, mammalian *DSCAM* and *DSCAML1* are not alternatively spliced, nor are they alternatively polyadenylated, although function in neural development appears to be conserved [117]. Although DCC (Fra in fly) is thought to be the primary

receptor for Netrin-1, it was found that DCC likely acts in complex with DSCAM and binding of Netrin induces an overall attractive growth queue. Interestingly, while at least 7 of the most N-terminal immune-globulin domains are required for homophilic binding (Igs 1-7), immunoglubulins 7-9 (only one of which being variable) are sufficient for Netrin binding, suggesting splice diversity might be less important for Netrin-1 binding [129, 130].

Aside from self-avoidance and neurite segregation, Dscam also functions in axon guidance as a receptor for the ligand Netrin-1 in the mammalian spinal cord [130]. More recently, *Dscam1* was also found to be capable of binding Slit, a ligand known to bind the receptor Robo1 causing a repulsive signal [107, 110]. Normally, Slit binds Robo1, generating a repulsive queue preventing midline crossing and counteracting the attractive effects of Netrin/DCC binding [107]. In this new model, the N-terminal fragment of Slit was determined to bind the first 3 immunoglobulins of Dscam1, allowing Dscam to form a complex with Robo1 to prevent repulsive signaling in longitudinal axons of the midline [110]. Robo1 is predominantly expressed in longitudinal axons of the CNS midline, providing further evidence that *Dscam1* performs varying functions in a tissue-specific and developmentally relevant context [107].

While I outline many functions of *Dscam1* in neuron morphology and axon guidance, the function of the extended 3' UTR isoform of *Dscam1* remains unknown. One study showed that *Drosophila* fragile xXmental retardation protein (*dFMRP*) and Wallenda (Wnd) to two RBPs, act on *Dscam1* to down-regulate protein expression [131]. It was found that Dscam expression is a regulator of the presynaptic arbor size in sensory neurons and mutants in either of these RBPs increased Dscam expression levels, resulting in increased arbor size. Only Wnd was found to bind the 3' UTR of *Dscam1,* and the extended 3' UTR was not examined. Characterization of the extended

3' UTR and its motifs may help elucidate further how *Dscam1* is regulated and how it carries out its various functions.

1.13 CRISPR/CAS9: A non-disruptive approach to studying extended 3' UTR function

Most genome editing techniques rely on double-stranded cleavage of the target DNA followed by repair of the break. Two endogenous cellular repair pathways exist, non-homologous end joining (NHEJ), and homology directed repair (HDR).

NHEJ is the most commonly employed repair mechanism by the cell. First, a NHEJ protein complex recognizes the break and recruits other proteins required for repair. Usually the cut produces either blunt or incompatible overhangs. If overhangs exist, they are resected before DNA polymerases are able to incorporate new dNTPs. Lastly, DNA ligase IV is able to join the two ends. NHEJ results in random insertions or deletions (indels), which often cause frameshifts [132].

HDR is very useful because this repair pathway can be manipulated to cause implementation of exogenous DNA at the cleavage site [133]. HDR is able to occur when a donor template whose ends are homologous (homology arms) to the ends of the cleaved DNA, which is usually a double-stranded break, is present (DSB) [134]. By introducing exogenous DNA templates with homology arms to into cells that contain DSB in DNA, it is possible to insert entirely new portions of DNA into the genome [135].

Tools to specifically induce double-stranded breaks have been around since the 1990's after the advent of zinc finger nucleases (ZFNs) [136]. A zinc finger is a 30 amino acid long protein that is able to recognize a specific trinucleotide sequence. Different zinc fingers have been developed to recognize every different trinucleotide sequence possible and they are modular in fashion. Therefore, multiple zinc fingers are able to be

linked together via a zinc ion. Most importantly, an endonuclease domain (usually FokI) is then fused to the most 3' zinc finger which allows cutting at a specific genomic site when directed by the zinc fingers. Collectively, the zinc fingers with the endonuclease domain are known as ZFNs. In order to cause a DSB, the FokI domains must dimerize. Therefore, ZFNs are used in pairs; one directed against the positive strand upstream of the cut site, and the other against the negative strand downstream of the cut site. Once cleavage has occurred, the site may be repaired via NHEJ or HDR.

The next major breakthrough in genome-editing technology arose through the discovery and application of transcription activator-like effector nucleases (TALENS) [137]. Transcription activator-like effectors (TALEs) are a large family of secreted DNA-binding proteins, first discovered in *Xanthomonas,* which are able to bind DNA with single nucleotide specificity [138, 139]. Similarly to ZFNs, TALEs are able to be chained together modularly, and are fused to a non-specific nuclease domain, which when dimerized, is able to induce a double-stranded break at a specific locus. TALEs allow a greater flexibility in targeting due to the fact that each unit has single nucleotide specificity as compared to ZFN's trinucleotide recognition, however, due to the highly repetitive nature of TALEs, they are costly to manufacture [140].

The clustered regulatory interspaced short palindromic repeats (CRISPR) system is a new revolutionary genome editing technology that allows genome editing in a faster, cheaper, and more precise manner than that of ZFNs or TALENs. The CRISPR system was first identified as part of the adaptive immune system of bacteria [141].

The endogenous system can be described in three parts: Foreign DNA integration, CRISPR RNA expression, and target cleavage. Following viral infection, a bacterium such as *S. pyogenes,* is able to retain and integrate fragments of viral DNA into a specific CRISPR locus within its own genome. Next, these integrated fragments

are expressed and processed into mature CRISPR-derived RNAs (crRNAs) [142]. In the endogenous system, a transactivating RNA (tracrRNA) hybridizes the crRNA through partial complementarity and acts as a scaffold to bind a non-specific CRISPR-associated endonuclease (Cas), most commonly Cas9. The 5' end of the crRNA contains what is known as a protospacer element which is the sequence that confers specificity of Cas9 to the target DNA. For the purpose of simplifying the system for genome editing, the crRNA and tracrRNA are combined into a single construct called guideRNA (gRNA). In addition to containing the protospacer sequence, which gives target specificity, the target DNA must also contain the protospacer adjacent motif (PAM), which is recognized by the Cas9 protein. Upon target recognition and binding a double stranded break is induced by the endonucleolytic activity of Cas9.

The CRISPR/Cas9 system has been shown to work in many model organisms including HEK293T cells, zebrafish, mouse, fly, worm, rat, frog, as well as *Arabidopsis* [143]. Conveniently, all components required for the CRISPR/Cas9 system can be expressed via microinjection of a single vector. Like both ZFNs and TALENs, CRISPR double-stranded breaks can be repaired via NHEJ or HDR by co-expressing a DNA fragment with sequence homology around the cleaved site. Due to Cas9 being an RNA-guided nuclease opposed to a DNA-guided nuclease, it can easily be retargeted by simply changing the 20 nucleotide sequence of the gRNA, and therefore, no protein redesigning is necessary. Additionally, multiple regions can be targeted simultaneously by simply expressing multiple gRNAs. With these benefits in addition to its cost effectiveness, CRISPR/Cas9 has become the preferred genome editing tool of scientists.

Chapter Two: Materials and Methods

Primers Synthesis:

All primers and RNA probes were synthesized and obtained from Integrated DNA Technologies (IDT). All sequences are listed in appendix A.

Genome browser analysis:

The Integrative Genomics Viewer (The Broad Institute) was used for primer design validation, RNA-seq data analysis, and identification of Poly(A) sites along with their associated motifs.

S2 Cell Transfection:

Schneider 2 cells were obtained from Drosophila Genomics Resource Center (DGRC). Cells were cultured in Schneider's *Drosophila* medium (ThermoFisher CAT# 21720024), with the addition of 10% fetal bovine serum (Atlanta Biologicals CAT# S11150). Cells were 80% confluent the day of transfection in a T25 tissue culture flask (USA Scientific CAT# CC7682-4825). The flask was then brought up to 15ml with fresh media, and 2ml aliquots were transferred to 6-well plates (USA Scientific CAT# CC7682-7506). Two hours were allowed to elapse to allow cells to become semi-adherent. All transfections were performed using the Effectene transfection reagent (Qiagen CAT# 301425) according to the manufacturer's protocol for suspension cells. For co-transfections, 500ng of the driver (tubulin-gal4) and 500ng of UAS-ELAV and/or reporter was added, for a maximum of 1.5ug being transfected if three plasmids were being introduced. Cells were collected 48 hours post-transfection and pelleted in preparation for RNA extraction.

RNA Extraction: RNA extracted from cell culture and *Drosophila* was performed using TRIzol reagent (Invitrogen CAT# 15596026) following the manufacturer's protocol. Following extraction, and prior to cDNA synthesis, RNA was DNase treated using the DNA-*free* DNase Treatment and Removal Reagents (Ambion CAT# 1906) following the manufacture's "Routine DNase treatment" protocol with 7ug of input RNA.

***Dscam1* 3' UTR Cloning:** The *Dscam1* 3' UTR was amplified from cDNA isolated from W1118 *Drosophila* melanogaster using iProof High-fidelity PCR kit (Bio-Rad Cat#1725330) per the manufacturer's protocol (Primers used found in appendix). All products were sub-cloned into Topo TA vector (Invitrogen CAT# K4575J10) per the manufacturer's protocol for Sanger sequencing prior to cloning into the terminal vector. The Topo-insert vector was grown up in TOP10 cells included with the kit using the supplied protocol. Resultant colonies were inoculated into 25ml of LB broth (Fisher Scientific CAT#BP1426-500), left to grow overnight, and were midi prepped (Qiagen CAT# 12945) the next day. After sequence confirmation, the insert, and terminal vector were digested with the appropriate restriction enzymes (found in appendix, obtained from New England Biolabs), and run in a 1% agarose gel for 45 minutes at 120v. The digested insert and vector were recovered from the gel using a gel extraction kit (Qiagen CAT# 28704). The prepared vector and insert were then ligated together using T7 DNA ligase (New England BioLabs CAT# M0318) in a molar ratio of 1:3 vector to insert using the manufacturer's protocol. The ligated vector was then grown in Top10 cells and midi prepped as described earlier.

Site-directed Mutagenesis: Mutagenesis of the putative proximal and distal ELAV binding sites was performed using the QuikChange II XL site-directed mutagenesis kit (Agilent CAT#200521) using the manufacturer's protocol. The mutagenic primers were designed using Agilent's primer design program (Primer sequence found in the appendix).

CRISPR/CAS9 Deletion: The CRISPR/CAS9 deletion was performed by Well Genetics. Two gRNAs were designed flanking the extended 3' UTR of *Dscam1* which caused double-stranded breaks at positions chr2R: 3,204,852 and chr2R: 3,206926 generating a nearly 2kb deletion. A 134 bp region between the proximal *Dscam1* short 3' UTR cleavage site and the first double-stranded break was left in tact in order to not disturb any potential DSEs important to proper cleavage and polyadenylation of the short 3' UTR isoform. HDR was used to knock-in a 1.8kbp RFP cassette containing loxP sites. RFP was used as a visible marker for screening successful mutants. Three successful *Dscam1* dUTR mutants were generated with the expected deletion and knock-in. Flies were balanced over CyO or GFP CyO to create stable stocks.

Genomic DNA extraction and Sanger sequencing: Genomic DNA extraction was performed using the DNeasy Blood and Tissue kit (Qiagen CAT#69504) per the manufacturer's protocol. Endpoint PCR amplification (as described above) was carried out using the Dscam_deletion_Seq1&2 primers, which flank the deletion locus (Appendix II). 500ng of cDNA and 1ul of a 10uM sequencing primer was used as input for Sanger sequencing. Both ends of the deletion locus were sequenced for all three *Dscam1* dUTR fly lines. Sanger sequencing was performed by the Nevada Genomics center using the 3730 DNA Analyser (Applied Biosystems Model# 3730).

cDNA Preparation and qRT-PCR: Dnase treated RNA was reverse transcribed into cDNA using iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad CAT# 1708840) as per manufacturer's protocol using 1ug of RNA as input. cDNA was then diluted 1:5 in ddH_2O. qPCR reactions were set up in clear low-profile 0.2ml 8-tube strips (Bio-Rad CAT# TLS0801) with 0.2ml flat PCR optical caps (Bio-Rad CAT# TCS0803). A diluted master mix was made with the ratio of of 10ul SYBR Green PCR Master Mix (Invitrogen CAT# 4309155) to 6.5ul ddH2O per reaction was first made. Next, a master mix for each pair of primers being tested was made using .75ul of each individual primer (at 10um) per reaction. Each individual reaction consisted of 16.5ul diluted master mix, 1.5ul primer mix, and 2ul of 1:5 diluted cDNA. Thermocycling took place in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad CAT#1855195). All data was analyzed on Bio-Rad's CFX software.

Longevity Assay: Male flies were collected within 8 hours of eclosion and separated into groups of 10 by genotype using minimal CO2. Flies were transferred to vials with standard food and were kept at 25 degrees Celsius, 55% relative humidity, and a 12 hour light-dark cycle. Flies were checked every two to three days for deaths and survivors flipped to fresh vials containing new food. Every group of 10 flies constituted an N=1. Data was plotted using excel.

ELAV Protein Vector Creation, Expression, and Purification

Vector Creation: The coding region of ELAV was PCR amplified from cDNA isolated from w1118 *Drosophila* using AccuPrime Pfx DNA polymerase (Invitrogen CAT# 12344024) per the manufacturer's protocol. The product was then cloned into the pENTR/D-TOPO vector (Invitrogen CAT# K240020) donor vector per the manufacturer's

protocol. The ELAV-pENTR vector was then grown up in One Shot TOP10 cells (Invitrogen CAT#C404003) per the manufacturer's protocol. The destination vector chosen was pET-60-DEST (Novagen CAT# 71851-3) because it contained both a C-terminal GST tag as well as a N-terminal 6xHis Tag. The destination vector was grown up in One Shot *ccd*B Survival 2T1^{R} cells (Invitrogen CAT# A10460). Both vectors were then prepared using midi-prep (Qiagen CAT# 12945). The final expression vector was made using LR Clonase II (Invitrogen CAT#11791-020) enzyme per the manufacturer's protocol, which flipped the ELAV coding region into the pDEST60 vector with the N and C-terminal tags.

Protein Expression and Purification: BL21(DE3) Competent E. coli (New England Biolabs CAT# C2527I) were transformed with pDEST60-ELAV construct according to the manufacturer's protocol. Cells were collected by centrifugation in 50ml centrifuge tubes (USA Scientific CAT# 1500-1811) for 15 minutes at 3000rpm. Cells were then suspended in 5ml of water and centrifuged again to rinse off leftover media. The water was decanted off and the cell pellet was resuspended in 30ml of a buffer containing 25mM Tris HCl pH 7.5, 150mM NaCl, 20mM imidazole and 10mM $MgCl_2$ with the addition of a protease inhibitor tablet (Thermo Scientific CAT#88666) and 30mg of lysozyme (Sigma CAT# L6876). The mixture was left to incubate for 30 minutes at room temperature. The cells were then sonicated (Branson Sonicator 450) on ice six times for 15 second intervals at a power level of four. The sample was then centrifuged at 28,000rpm for 1 hour at 4°C. 20ul of supernatant was collected to be later ran on a gel for quality analysis. Following sonification, 900ml of a buffer containing 25mM Tris HCl pH 7.5, 150mM NaCl, 20mM imidazole and 10mM $MgCl_2$ and 290ul of 2-mercaptoethanol was prepared. 2ml of glutathione agarose (Thermofisher CAT# 16100)

was then mixed with 48ml of the buffer just prepared, and was allowed to sit so the beads settled. Supernatent was then removed leaving only the beads at the bottom of the tube. The supernatant from the 1-hour centrifugation step was then added to the beads and incubated for 2 hours at 4°C with rocking. The rest of the purification took place at 4°C in a cold room. A chromatography column (Bio-Rad CAT# 7374021) was rinsed with the imidazole buffer containing the 2-mercaptoethanol, and all 50ml of the supernatant-glutathione bead mix was added to the column, allowing the beads to compact in the column. Some of the flow-through was kept for quality analysis performed later; the rest of the flow-through was discarded. To remove any material not bound to the agarose beads, 150ml of the buffer containing 2-mercapoethanol was added to the column to wash the beads, and the flow-through was discarded. To remove the ELAV bound protein, 0.28g of reduced L-glutathione (Sigma CAT# G4251) was added to 30ml of the buffer containing 2-mercaptoethanol. 10 ml of the reduced L-glutathione solution was added to the column and left to incubate for 5 minutes. Following incubation, 8x1ml aliquots were collected. To determine which fraction contained the highest concentration of ELAV, 5ul from each fraction (and also the flow-through samples collected) was aliquoted and added to new tubes each containing 5ul of SDS loading buffer. The samples were heated at 95°C for five minutes and then all samples were ran in a tris glycine gel for 30 minutes. The gel was then stained for 30 minutes while shaking in a solution containing 3.75ml of acetic acid, 46.25ml of ddH_2O and 10ul of sypro orange (ThermoFisher CAT# S6650). The gel was then scanned on a gel imager (GE Healthcare Typhoon Trio) to determine which fraction contained the most protein. That fraction with the most protein was then quantified using a BSA standard curve protocol, using ImageJ to quantify pixel intensity of the Flamingo-stained(Bio-Rad CAT#161-0490) TrisGlycine Gel.

Electrophoretic Mobility Shift Assay

Polyacrylamide Gel purification of Probes: All RNA oligos were ordered from Integrated DNA Technologies (Appendix II). The lipholized RNA was rehydrated to a concentration of 1nmol/ul. The oligos were then purified using a 12-15% denaturing polyacrylamide gel (ex. 15%: 14.2ml DEPC treated water, 48g Urea, 37.5 ml of 40% 19:1 acryl:bis, 500ul 10%APS, 50ul Temed). Denaturing gels were left to polymerize for at least two hours, and then pre-ran for five minutes at twenty watts. 10nmol of RNA sample was mixed with an equal volume of Thermo Fisher loading buffer II composed of 95% formamide, 18mM EDTA, .025% SDS, small amounts of bromophenol blue, and xylene cyanol. Samples were incubated at 95°C for five minutes. Samples were then loaded into the gel and ran for six hours at twenty watts with 1x TBE used as running buffer. RNA was detected in the gel using short wavelength UV and extracted for gel purification.

Gel Purification: RNA samples excised from the gel were mixed with 500ul of acrylamide crush and soak buffer (0.5M ammonium acetate, .1%SDS, and .1mM EDTA) and pulverized. They were then left to incubate overnight at 4°C. Samples were then run through a .22um cellulose acetate centrifuge filter column for 1 minute at 14,000 RPM.

Phenol Chloroform extraction: 500ul of Phenol:Chloroform:isoamyl alchohol (25:24:1) was added to the sample, vortexed, and centrifuged for 1 minute at 14,000RPM. The bottom layer was removed and discarded. 500ul of chloroform:isoamyl alcohol (24:1) was then added to the sample. Samples were once again vortexed and centrifuged for 1 minute at 14,000RPM. The bottom layer is removed once more and discarded.

Ethanol Precipitation: 1500ul of ethanol and 50ul of 3M sodium acetate was mixed with the sample. The sample was left at -20°C for 1 hour and centrifuged for 30 minutes at 14,000RPM at 4°C. Supernatant was discarded and the sample was washed with 500ul 70% cold ethanol. The samples are once again centrifuged at 14,000RPM. Supernatant is removed and the samples were left to dry.

Quantification: Samples were then diluted with water until the UV spectrophotometer reading A_{260} was between 0.2 and 0.8. The concentration was then determined using the absorbance value and the extinction coefficient of the RNA sample.

P32 5' end labeling: 2.5pmol of each sample was 5' labeled with γ^{32}-P. 2.5pmol was dried to a solid using a speedvac. A master mix containing 1ul of T4 Polynucleotide Kinase buffer, 5.6ul Water, and .4ul of T4 polynucleotide kinase was prepared. 7ul was aliquoted into each sample of dried RNA and resuspended. 3ul of [γ^{32}-P] ATP was added to each sample and left to incubate at 37°C for one hour.

Spin Column purification: Illustra Micropsin G-25 columns were prepared by centrifugation at 735 RCF for one minute. 250ul of water was then added to the columns and centrifuged again to remove trace amounts of ethanol. The columns were then placed in microcentrifuge tubes that contained 25ul of Thermo Fisher loading buffer II. 15ul of water was added to each sample and then loaded into the G-25 spin column. Columns were then centrifuged at 735 RCF for two minutes. Columns were then discarded and the sample was incubated at 95°C for 5 minutes.

Repurification: All samples were purified in a denaturing polyacrylamide gel, phenol-chloroform extracted, and precipitated as described previously. To determine the location of the RNAs in the gel use a phosphor screen instead of short wavelength UV.

Native gel and RNA sample preparation: A 4% acrylamide:bisacrylamide 80:1 native gel was prepared (27.25ml DEPC treated water, 1.65ml 10xTBE, 3.3mL 40% acrylamide, 830ul 2% bisacylamide, 133ul of 10% APS, and 66.6ul of Temed) and left to polymerize for at least 2 hours. The RNA samples were then resuspended in hybridization buffer (10mM Tris HCl pH7.5, 100mM NaCl, and .1mM EDTA) so that the final concentration was 10 times less than the molar concentration of the ELAV protein sample. Samples were then incubated at 65°C for 5 minutes.

ELAV-RNA hybridization: A master mix was prepared consisting of 1ul 10x reaction buffer (450mM Tris HCl ph 7.5, 5mM NaCl, and 400 mM KCl), 1ul of 250ug/ml tRNA, 1ul 5mM DTT, 1ul 500ug/ml BSA, and 1ul of 6units/ul RNAse inhibitor. 5ul of the hybridization mix was aliquoted per sample and 4.35ul of ELAV protein, or GST buffer without protein was added to each tube. Then .65ul (154nM) of RNA sample was added to its respective tube. Hybridizaion was allowed to take place for 20 minutes at room temperature, and then 3ul of loading buffer was added to each sample.

EMSA: The gel was pre-run for 15 minutes at 250V. Samples were then loaded into the gel while still running at 250V. The gels were then ran for 2 hours and 20 minutes in a cold room at 4°C. Gels were then dried on a gel dryer at 80°C for 1 hour with vacuum. A phosphor screen was exposed to the gel overnight and imaged the following day.

Negative Geotaxis assay:

The negative geotaxis assay performed was adapted from the RING assay[144]. Flies were collected flies within 24hr of eclosion and sorted into groups of 10 by gender and genotype with minimal use of CO2. Flies were allowed to recover for 18 hours on regular food in the incubator. Flies were transferred to the climbing apparatus, which consisted of two polystyrene tubes (Genesee Scientific CAT# 32-110) and allowed to acclimate for at least 5 minutes. The test was performed by mechanically tapping them down 5 times rapidly with consistent force. A picture was taken after 10 seconds to determine how many had climbed more than 3cm. Flies were allowed to rest for 1 minute before 4 more trials were completed. A total of 5 trials on the same group of flies constituted a N of 1. This was performed for a N of 5 for each gender/genotype.

Embryo Collection and Fixation: To analyze the embryonic ventral nerve cord, flies with the genotype of interest were placed in cages containing grape agar and thin layer of yeast paste to lay eggs. Grape agar plates were made using 350ml ddH20, 120ml grape juice, 15g agar gelidium (MOORAGAR CAT# 41054) and 13g of sugar. The mix was allowed to boil for 5 min and 2ml of propionic acid (Sigma CAT# 402907) was added after the mix had cooled. Grape agar mix was then poured into 60mm petri dishes (Genesee Scientific CAT# 32-105). Yeast paste was prepared mixing standard bread yeast with water until a peanut butter-like consistency was reached. Changing the plate two times after over the course of 24 hours synchronized the embryos on the plates. Following synchronization, flies were allowed to lay eggs for 2 hours per plate. Grape plates were then allowed to age for ~16 hours to ensure the majority of embryos collected are at stage 17 embryos. Approximately 200ul of embryos were collected and placed in a sieve and washed 3X with .02% triton-X 100 (Sigma CAT# T8787). Embryos

were then dechorionated by incubation for five minutes in a 50% bleach solution and then rinsed again 3X with .02% triton-X. Embryos were then transferred to a 1.7ml microcentrifuge tube (Genesee Scientific CAT# 22-284) containing equal parts heptane (Fisher Scientific CAT# H350) and 4% paraformaldehyde/PBS (Boster CAT# AR1068) and allowed to incubate with shaking for 20 minutes. The lower aqueous phase was then removes and replaced with an equal volume of methanol (Fisher Scientific CAT# A452). Embryos were then shook vigorously for 1 minute to remove vitelline membrane and then allowed to settle. The upper phase was then removed and embryos were washed 3X with methanol.

1D4 (Fasciclin II) Immunostaining: To inspect the embryonic VNC for defects, stage 17 *Dscam1* dUTR /CyO GFP embryos were stained by 1D4 immunolabeling. Approximately 25ul of fixed embryos were washed with a 1XPBS (Fisher Scientific CAT# 20012027) .1% Triton solution (PT) 3X for five minutes. Embryos were then incubated in 750ul of PT + 5% normal goat serum (Invitrogen CAT# 31873) for 30 minutes at room temperature with rocking. Primary 1D4 anti-Fasciclin II mouse antibody (Developmental Studies Hybridoma Bank CAT# AB_528235) was added to achieve a 1:5 dilution. Rabbit Anti-GFP antibody (Life Technologies CAT# A11122) was also added at a dilution of 1:400 in order to screen out heterozygotes. The embryos were incubated overnight at 4 degrees Celsius with rocking. Embryos were then washed with PT 3X for 5 minutes with shaking, followed by another set of 3 washes in PT for 30 minutes each. After washing 750ul of PT + 5% normal goat serum was added and secondary goat anti-mouse HRP antibody (Jackson ImmunoReasearch laboratories CAT# 115-035-003) along with HRP Goat-anti rabbit (Jackson ImmunoReasearch laboratories CAT# 111-035-003) were added to dilutions of 1:250 and 1:10,000, respectively. Secondary antibodies were

incubated for 3 hours with shaking at room temperature. Embryos were then washed according the same regimen after primary antibody incubation. Development of embryos took place in 1ml of 3,3'-diaminobenzidine (Sigma CAT# D8001) After two minutes of embryos equalizing to the solution, 3ul of 30% hydrogen peroxide (Sigma CAT# H1009) was added. Development was allowed to take place until the desired color was achieved, usually 30 minutes. The development was stopped by rinsing 3X with PT, followed by rinsing 3X with PBS. Embryos were then transferred into 70% glycerol for clearing. Correctly staged embryos were selected out using a light microscope and the VNC was dissected using custom-made tungsten needles. The VNC was mounted on glass slides and imaged on a Leica DM 5000B.

Chapter Three: Results

Alternative polyadenylation of transcripts is pervasive in many metazoan genomes, including *Drosophila*, mouse, and human [65, 66]. Recent studies provide evidence for an enrichment of thousands of genes with extended 3' UTR isoforms in the CNS, and in *Drosophila,* at least 383 genes have been shown to have extended 3' UTR isoforms exclusively expressed in neural tissues [66]. In hopes of elucidating a broad role for these extensions, especially amongst genes involved in axon guidance, we chose to study the function of the *Dscam1* extended 3' UTR. *Dscam1* is a relatively well-characterized gene due to being the most extensively alternatively spliced gene known, however, specific function of the extended 3' UTR isoform remains unknown.

3.1 Regulation of *Dscam1* 3´ UTR extension by ELAV

In *Drosophila,* ELAV is the most well studied RBP that mediates 3' UTR extension [73, 95]. To determine if *Dscam1* extension was regulated by ELAV, we analyzed RNA -seq tracks of Schneider 2 cells that were induced to express ELAV. RNA seq tracks provide visualization of relative gene expression of a particular genomic region by showing the pile-up of individual reads generated from RNA-sequencing. The data revealed that extended isoform variants of *Dscam1* are upregulated upon ELAV overexpression (Figure 4.A). As a positive control we also show 3' UTR switching of *ewg,* a known target of ELAV (Figure 4.B) [145]. To confirm this, quantitative real-time PCR (qRT-PCR) was performed to measure expression of extended *Dscam1* 3' UTR isoforms in S2 cells transiently transfected with ELAV (Figure 5.A). Expression of the extended transcript is detected by a set of primers specific to the extended isoform, while another set of universal primers detects all *Dscam1* isoforms. The extended 3' UTR expression is shown as a ratio of extended isoform normalized to the housekeeping

gene *Rpl32*. As a positive control, expression of the extended 3' UTR variant of *erect wings* (*ewg)* was measured in the same manner. Analysis reveals that ELAV expression in S2 cells increases expression of extended *Dscam1* 4.46 compared to mock transfected cells, suggesting *Dscam1* is a target of ELAV.

The exact mechanism by how ELAV influences 3' UTR extension is not known; however, ELAV has been proposed to compete with cleavage factors binding U-rich elements near the proximal poly(A) site. In order to determine if putative ELAV binding sites (EBS) are located within the *Dscam1* 3' UTR we first established the location of the proximal and distal poly(A) sites. Using the integrative genomics viewer, poly(A) sites were identified manually by scanning for USEs, DSEs, and most importantly the PAS (Figure 6). Consistent with the idea that proximal poly(A) sites tend to be weaker than their downstream counterparts, the proximal poly(A) site of *Dscam1* does not contain a canonical PAS of A[A/U]UAAA . To identify putative EBSs, U-rich regions were scanned for in the vicinity of the poly(A) site. A roughly 52 bp long region with a uracil content of 63.4% was identified near the proximal poly(A) site and we predicted it to be a possible EBS. Interestingly, another potential EBS was found immediately downstream of the distal poly(A) site, although there is limited evidence showing ELAV binding near distal polyadenylation sites in vivo.

To test if ELAV was capable of binding these sites, we performed an electrophoretic mobility shift assay (EMSA) using ^{32}P 5' radiolabeled RNA probes and recombinant purified ELAV (Figure 7.C). The proximal and distal EBS RNA probes consisted of the putative EBS sequences identified above, and were 43 and 58 basepairs long, respectively (Figure 7.B). The addition of purified ELAV with the RNA probes resulted in an upward shift of the band for both proximal and distal probes,

suggesting RNA-protein complex formation. ELAV has been shown to bind at least one of its targets, *ewg,* as a dodecamer, and we hypothesize the minor band shift seen in ELAV positive lanes may be caused by ELAV binding the substrate as a smaller multimer. [146]. To control for non-specific binding of ELAV to the probes, two additional probes were generated by disrupting U-rich sequence integrity of the test probes via U to G substitution. Addition of ELAV to the mutant RNA probes with disrupted U-rich stretches did not result in a shift of the major band. This supports the idea that ELAV binding is specific to U-rich sequence motifs. Notably, both mutant probes appeared to run through the gel at a faster rate than their U-rich counterparts. We believe that this is caused by a reduction in secondary structure of the RNA probes as a result of the U to G substitutions.

We have shown that ELAV overexpression is sufficient to cause *Dscam1* extension in S2 cells and that it is capable of binding at least two regions of the *Dscam1* 3' UTR *in vitro;* however, it is not clear if the two binding sites tested are necessary for *Dscam1* 3' UTR extension. In order to test this, we created four different luciferase-*Dscam1* 3' UTR fusion reporters using the psicheck-2 vector (Figure 8.A). The luciferase reporters either contained the short *Dscam1* 3' UTR, the full-length 3' UTR, or the full-length 3' UTR with either the proximal or distal EBS mutated via site-directed mutagenesis. With this reporter system we also hoped to determine the effect(s) that the extended *Dscam1* 3' UTR has on transcript stability and translational control.

To specifically detect the expression levels of the short and extended reporter isoforms, it was necessary to perform a northern blot targeting luciferase because the qRT-PCR primers previously used would be incapable of distinguishing between endogenous *Dscam1* and the reporter construct. The psicheck-2 vector is under the

control of the constitutively active SV40 promoter and preliminary experiments showed expression of the reporter in transfected S2 cells was below detectable levels for northern blot (data not shown). We decided to switch to HeLa cells, a mammalian line, in an attempt to improve expression of the reporter. In a preliminary test, HeLa cells were transfected with the reporter containing either no 3' UTR, the *Dscam1* short 3' UTR, or the full-length *Dscam1* 3' UTR, and a northern blot for luciferase was performed. Although HeLa cells yielded high expression of the reporter, multiple splice variants were detected which could not be explained (Figure 8.B). This suggests that the Psicheck2 vector is not appropriate for studying APA of *Dscam1*, and calls into question its suitability as a reporter construct for studies of *Drosophila* sequences in S2 transfection experiments.

It was our opinion that the simplest solution to this problem was to switch back to the S2 cell system with the use of a previously described UAS-driven EGFP reporter [131]. Generation of these constructs is currently underway (Figure 9.A). A preliminary test transfection of the empty EGFP construct and the EGFP reporter containing the short *Dscam1* 3' UTR was performed to determine efficacy of the new system. A northern blot targeting EGFP was able to detect sufficient levels of the reporter, and yielded the expected band sizes suggesting this reporter system is adequate for testing the effects the *Dscam1* 3' UTRs on transcription.

3.2 Function of *Dscam1* 3´ UTR extension

3.2.1 Generation of the *Dscam1* extended 3' UTR deletion

Dscam1 plays several crucial roles during embryonic neural development including axon guidance and establishing unique neuron identity. Although many of the roles of *Dscam1* in development have been well characterized, the role of the extended

3’ UTR in these developmental functions remain unknown. In order to directly study the biological importance of the *Dscam1* extended 3’ UTR *in vivo*, we utilized the CRISPR/CAS9 system to create a mutant fly incapable of making the extended 3’ UTR isoform (hereinafter referred to as *Dscam1* dUTR mutant).

To generate the *Dscam1* dUTR mutant, two guide RNAs were designed flanking the *Dscam1* extended 3’ UTR region for Cas9 mediated cleavage (Figure 10. A). Exactly 134bp downstream of the *Dscam1* proximal cleavage site was left intact to maintain U-rich DSEs important for proper cleavage of the short 3’ UTR isoform. Following the two double-stranded breaks, a 1.8kb RFP cassette containing loxP sites was knocked-in at the deletion locus using HDR. The RFP served as a visible marker in order to screen for successful deletion mutants. This method generated 3 *Dscam1* dUTR lines, A,B, and C.

To ensure the correct deletion was made, the deletion locus of all three lines was verified via Sanger sequencing. Genomic DNA was extracted from five adult flies of each line and primers flanking the deletion locus were used for PCR. Sequencing primers designed to have coverage through the expected 5’ and 3’ UTR/RFP cassette junction. All three lines were confirmed to have an identical deletion on chromosome 2R from position 3,204,852 to 3,206,962 as expected (Figure 10.B). In addition to our Sanger sequencing, line A was independently sequenced by Well Genetics. All three lines appeared to be similar phenotypically, and all experiments hereafter were performed on line A. To confirm that *Dscam1* dUTR mutants were not generating an artificial long transcript induced by the CRISPR deletion, we performed a northern blot targeting universal *Dscam1* transcript (Figure 10.C). The northern blot resulted in only the short 3’ UTR isoform being detected in the *Dscam1* dUTR mutants, as expected.

3.2.2 Phenotype analysis of the *Dscam1* dUTR mutant

Dscam1 dUTR homozygous flies were determined to be viable, but adults were scarce compared to their CyO balanced counterparts. A longevity assay was performed to determine if the deletion had an effect on lifespan (Figure 11.). Adult male flies were collected within 8 hours of eclosion and seperated into groups of 10 by genotype into vials. The vials contained standard food and all flies were housed at 25°C, 55% relative humidity, and a 12 hour light-dark cycle. Strikingly, approximately 50% of *Dscam1* dUTR homozygotes died by day 14 post-eclosion, significantly more than wild type and *Dscam1* dUTR heterozygous flies. Upon closer inspection, it was seen that mortality was almost exclusively due to homozygotes becoming stuck in the food, potentially indicating poor locomotor activity.

It became apparent that locomotor activity is severely diminished in homozygous *Dscam1* dUTR flies, which creates a propensity for the flies to “drown” in their food following eclosion. To better assess the extent of this locomotor defect, a negative geotaxis assay was performed to evaluate the climbing ability of homozygous mutants following mechanical stimulation (Figure 12.). Groups of 10 flies at a time, seperated by genotype and gender, were transferred to the testing apparatus, which consisted of two polystyrene vials taped together. Only ~10% of homozygous flies were able to climb more than 3cm within the 10 second testing window following mechanical stimulation, a roughly 86% reduction in negative geotactic response.

We hypothesized these locomotor defects were the result of abherrant axon guidance during development. To investigate this possibility we examined the ventral nerve cord (VNC) of stage 17 embryos for abnormalities in the longitudinal axon tracts at the midline (Figure 13.). Defects of the VNC have been previously characterized in *Dscam1* null mutants by 1D4 (anti-fasciclin II) immuno-staining [147]. Remarkably, we

identified several defects in homozygous *Dscam1* dUTR mutants similar in nature to *Dscam1 null* mutants. Although not as severe, evidence of axon kinking, defasciculation, and breaks in the outermost fascicles were identified.

The simplest explanation for the phenotypes displayed was an overall change in the amount of Dscam expressed as a result of the deletion. First, we investigated the amount of total *Dscam1* mRNA being produced via qPCR in adult heads (Figure 14.A). Using a “universal” primer that detects both long and short 3’ UTR isoforms of *Dscam1*, we found that the *Dscam1* dUTR mutants actually made 1.82 fold more overall transcript than wild type flies. This was very interesting, and suggested that a compensatory mechanism may exist. We expected to see an increase of *Dscam1* protein, but this was not the case. Western blot for Dscam protein revealed little to no change in the amount of protein expression (Figure 14.B).

Chapter Four: Discussion

We have shown that ELAV is sufficient to cause *Dscam1* 3' UTR extension in S2 cells. Furthermore, we were able to show that ELAV is capable of binding at least two different loci on the *Dscam1* 3' UTR immediately downstream of the proximal and distal PASs.

Unfortunately, expression problems with our reporter system have delayed us from determining the effects that the different 3' UTRs of *Dscam1* have on reporter transcription and translation rates in S2 cells. Once the new GFP-*Dscam1* 3' UTR constructs are made we will also be able to test which U-rich elements are necessary for proper *Dscam1* 3' UTR extension with the use of the the mutagenized ELAV binding site constructs by coexpressing ELAV. If the proximal mutagenized EBS construct is still capable of generating the extended 3' UTR reporter, it woud suggest another EBS within the 3' UTR may be present. Alternatively, this might also suggest a different RBP is capable of causing *Dscam1* 3' UTR extension. *Elav* has two homologs in fly, RBP9, and FNE [148]. We have data (not shown) that *fne* is capable of causing 3' UTR extension of at least 3 genes that are targets of ELAV in S2 cells.

Generation of the *Dscam1* 3' dUTR mutant was a success and demonstrated that the extended *Dscam1* 3' UTR serves a functional role. Initially it was dissapointing to see that unlike *Dscam1* null mutants, which don't survive into adulthood, our mutants appeared to be adult viable, however, the mutant flies recapitulated a resemblence of the axong guidance defects seen in the embryonic VNC. Homozygous mutants also showed a severe defect in negative geotaxis and overall locomotor activity. This defect resulted in a significantly increased mortality rate when compared to wild type flies. 1D4 immunostaining of the embryonic VNC revealed that longitudinal axon bundles

demonstrated breaking, stalling, and increased waviness, suggesting a disruption of proper axon guidance during embryonic development. Recent 1D4 immunostains of the adult *Dscam1* 3' dUTR mutant brain consistently show a major defect with the ellipsoid body, a structure known to be involved in locomotor activity (data not shown). Interestingly, this result appears similar to the result from Wang et al., 2002 showing that flies expressing a reduced diversity of *Dscam1* transcripts have ellipsoid body neurons that fail to properly segregate [126]. This data suggests that the extended 3' UTR has an important role in the overall function of *Dscam1.*

As an additional measure to confirm the mutant phenotypes identified, a shRNA knockdown fly for the extended 3' UTR of *Dscam1* was generated. Although not all experiments have yet been replicated with this fly, thus far, it recapitualtes the phenotyptypes observed in the *Dscam1* 3' dUTR mutants. This information suggests that the phenotypes displayed are not due to off-target effects of CRISPR.

Interestingly, our western blot against Dscam in the dUTR mutant indicated almost no change in Dscam1 protein levels. This suggests that the defects seen at the embryonic VNC are not simply the result of an effect on gene dosage. Since gene dosage was unlikely to be responsible for these axon guidance defects, we chose to investigate if the extended 3' UTR isoform differentially localized subcellularly.

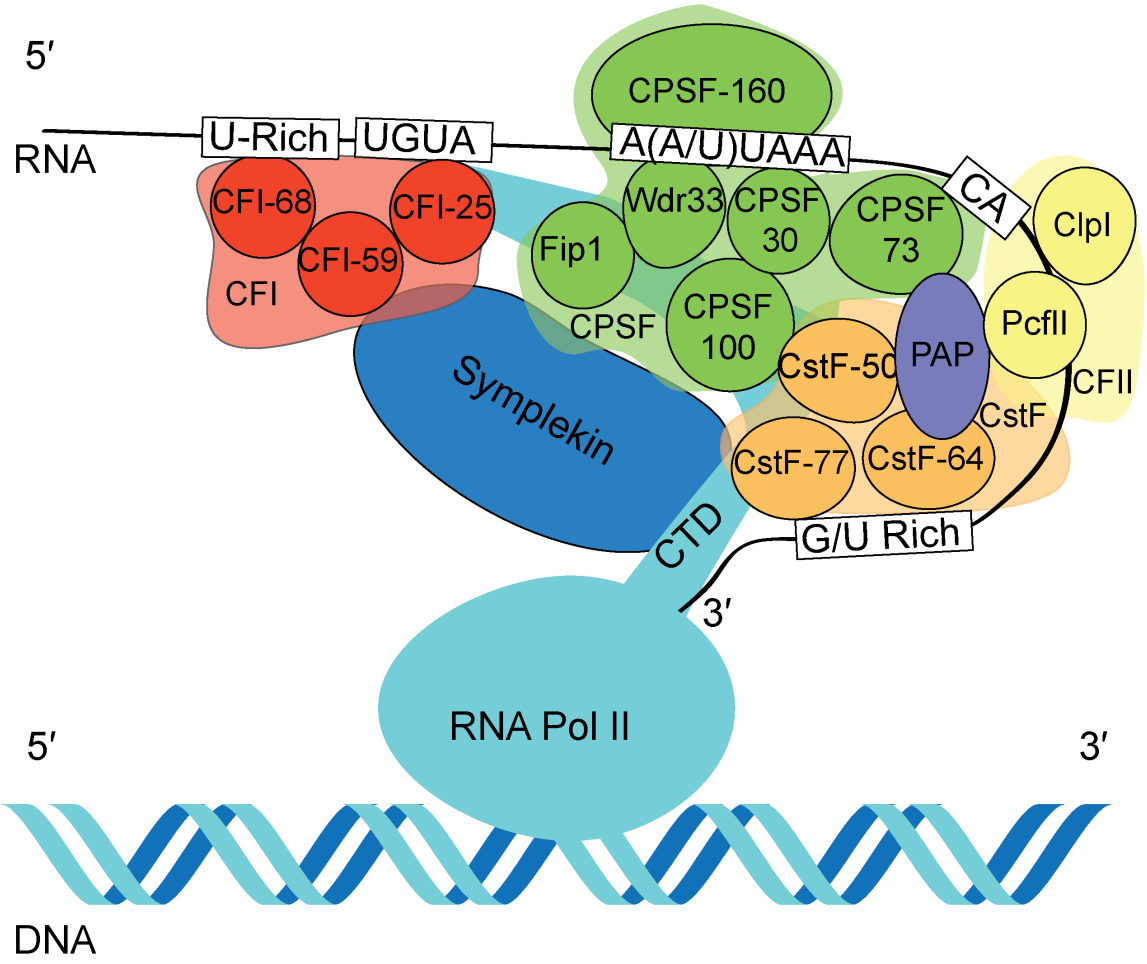
In a study examining the extended 3' UTR of brain-derived neurotrophic factor (BDNF) in mouse, it was found that the extended 3' UTR isoform of BDNF was necessary to target BDNF transcripts dendrites [87]. This was shown by generating mutant mice expressing a truncated version of the extended 3' UTR of the BDNF transcript. These mutant mice were unable to target BDNF to apical dendrites which reduced the overall amount of BDNF protein found in dendritic spines. The loss of BDNF in these spines was shown to cause abnormal development of CA1 pyramidal distal

dendritic spines. The targeting of BDNF to dendrites is thought to allow localized translation and differential regulation BDNF subcellularly. In a similar study in fly, it was shown the the extended 3' UTR of another gene, *calcium/calmodulon-dependent protein kinase II* (*CaMKII)* is necessary to maintain the correct basal activity of miniature excitatory junctional potentials in the larval neuromuscular junction. While it was shown that the extended 3' UTR of CamKII was not essential for dendritic localization, it was shown to be required fo translation-dependent plasticity of spontaneous release [149]. To determine if the extended 3' UTR of *Dscam1* was differentially localizing subcellularly to neurites, we performed *in situ* hybridization in wild type flies (data not shown) to specifically detect the extended 3' UTR isoform, however, we were unable to find any evidence of this occuring.

Our latest hypothesis is that the extended 3' UTR may be associated with specific alternatively spliced versions of *Dscam1.* As mentioned in the introduction, certain RNA-RNA interactions have been shown to be involved in splice choice of alternative exons in the 3, 6, and 9 exon regions. It could be that secondary structure in the extended 3' UTR participates in these types of interactions to influence upstream splicing of exons. To investigate this, we plan to utilize long-read RNA sequencing with the new nanopore technology. Current limitations of RNA-seq stem from read-lengths which are often too short to cover multiple exon junctions. Oxford Nanopore allows sequencing of single-stranded pieces of cDNA that is multiple kilobases long [150]. It is the same technology that was utilized to uniquely identify more than 7,800 transcript variants of *Dscam1* [113]. If we are able to identify a biased usage of alternatively spliced exons of associated with the exteneded *Dscam1* 3' UTR isoform, we may be able to uncover yet another role of the 3' UTR.

Although major defects are observed in the *Dscam1* 3'-dUTR mutants, the mechanism remains unclear. Elucidating the mechanisms responible for the observed defects may provide insight into extended 3' UTR function of other axon guidance genes with extended 3' UTRs.

Figures



Adapted from Tian & Manley, 2017

Figure 1. Cleavage machinery and polyadenylation site sequence motifs
The CTD of RNA Pol II along with symplekin provide the scaffold for many of the cleavage factors responsible for 3' end processing. Upstream elements such as UGUA are recognized by Cleavage Factor I, while downstream elements are recognized by CstF. The primary sequence motif of the Poly(A) site is the hexameric PAS A(A/U)UAAA, which is recognized by CPSF. The CPSF73 subunit possesses RNA endonuclease activity that allows for cleavage of the 3' end, often at a CA dinucleotide downstream of the PAS. Following cleavage, PolyA polymerase initiates formation of the Poly(A) tail at the site of cleavage.

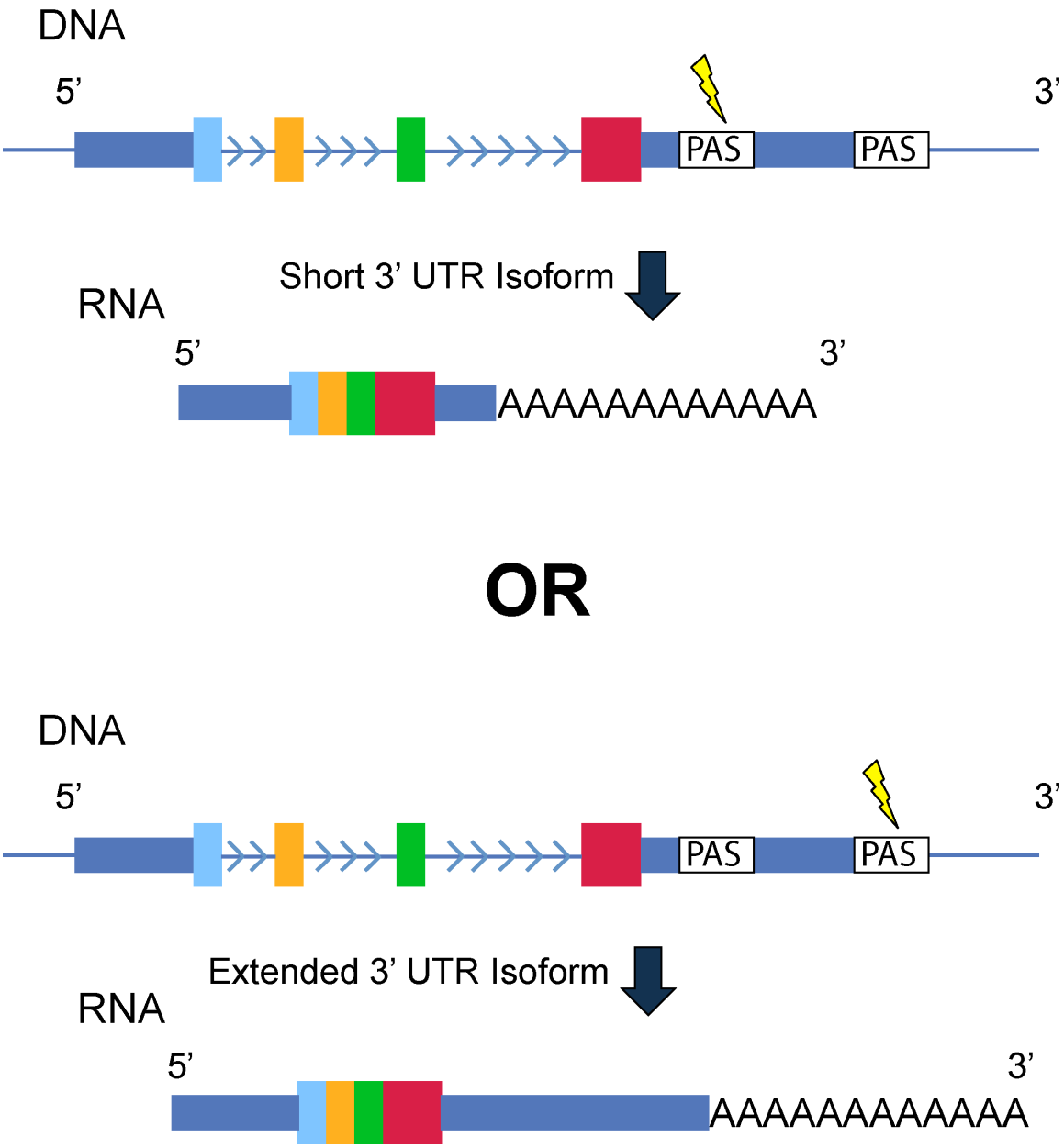
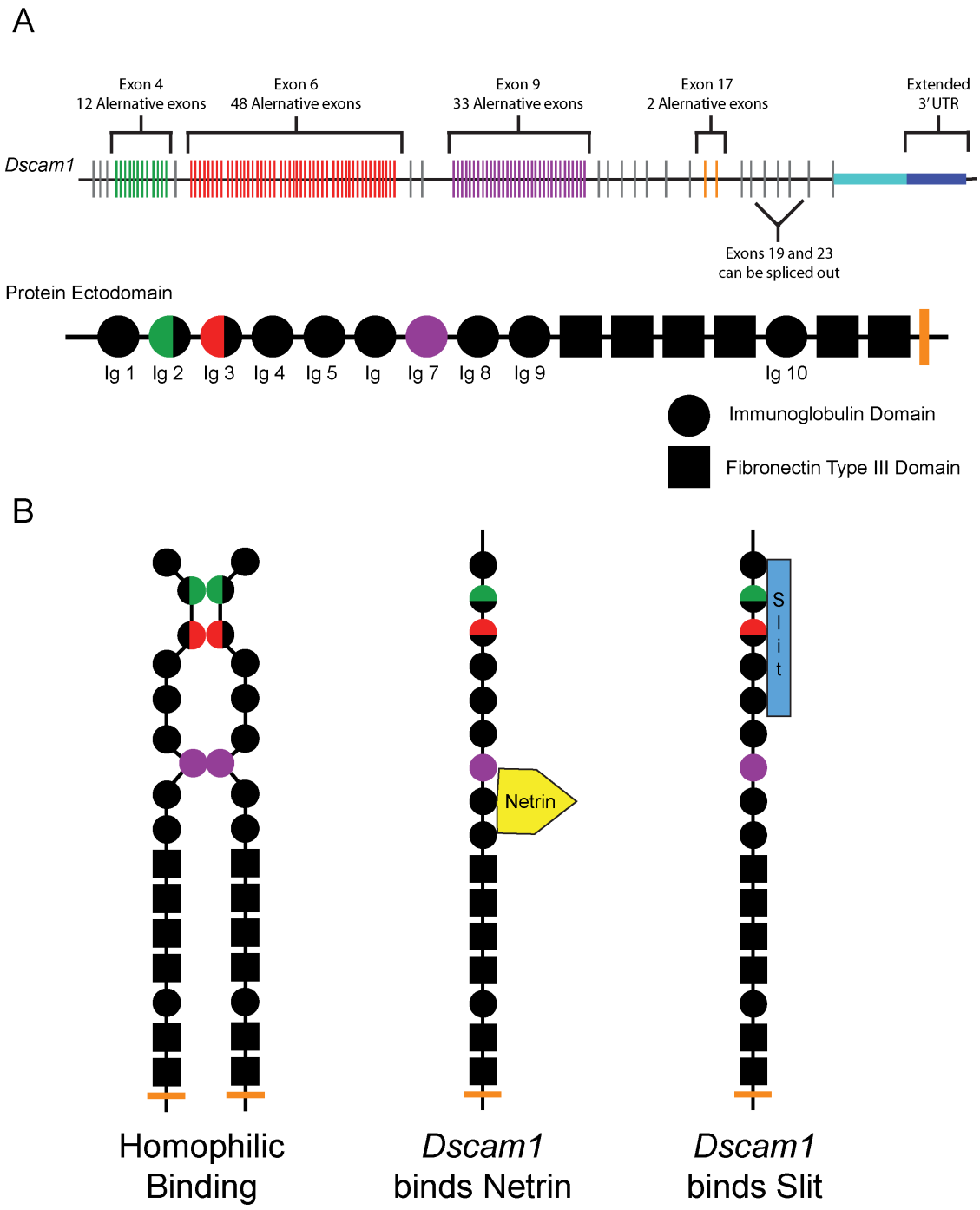


Figure 2. Usage of alternative PAS leads to transcripts with alternative 3’ UTR. Cleavage and polyadenylation occurring at the proximal PAS yields a shorter 3’ UTR Isoform, while cleavage at the distal PAS generates the longer 3’ UTR isoform. Longer 3’ UTRs can harbor additional regulatory motifs, altering gene expression while not affecting the protein coding region.



Adapted from Graveley et al., 2013

Figure 3. *Dscam1* gene structure and protein function. *Dscam1* is the most alternatively spliced gene known, able to generate up to 152,064 protein isoforms and 304,128 unique mRNA isoforms assuming all combinations are possible. This diversity primarily stems from the 3 hypervariable exon clusters 4, 6, and 9, which encode a part or an entire immunoglobulin domain. Dscam1 functions as a receptor protein on the surface of neurons. It is capable of homophilic binding, leading to neural self-avoidance, or heterophilic binding, which provide attractive queues during neural development.

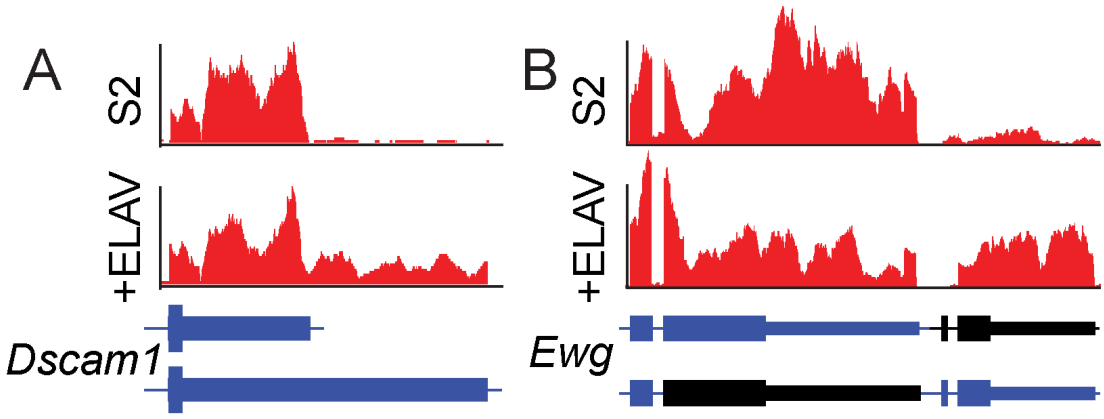


Figure 4: ELAV regulates 3´ UTR extensions of neural genes.
A) RNA-seq data of S2 cells overexpressing ELAV shows an increase in the usage of the longer 3' UTR isoform of *Dscam1.* B) *Ewg,* a gene known to be under the control of ELAV, is shown to switch to usage of the more distal 3' UTR isoform. EWG undergoes a slightly different version of 3' UTR APA where in addition to 3' UTR switching, an alternative last terminal exon used. Black denotes the part of the gene spliced out.

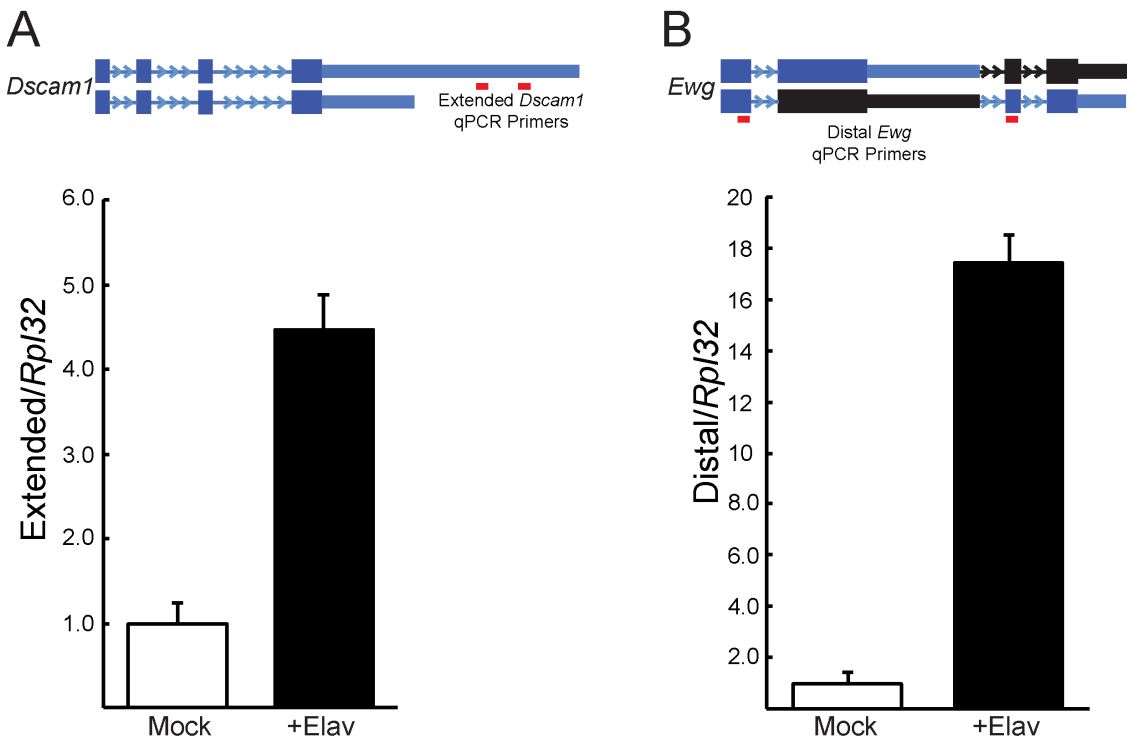


Figure 5. Overexpression of ELAV in S2 cells influences APA.
A) qPCR reavels S2 cells overexpressing ELAV show an 4.46 fold increase of *Dscam1* extended 3' UTR usage. B) Similarly, our positive control, *EWG,* exhibits a 17.42 fold increase in usage of the alternative distal 3' UTR. In both analysis, fold change of the alternative 3' UTR is normalized to the housekeeping gene Rpl32.

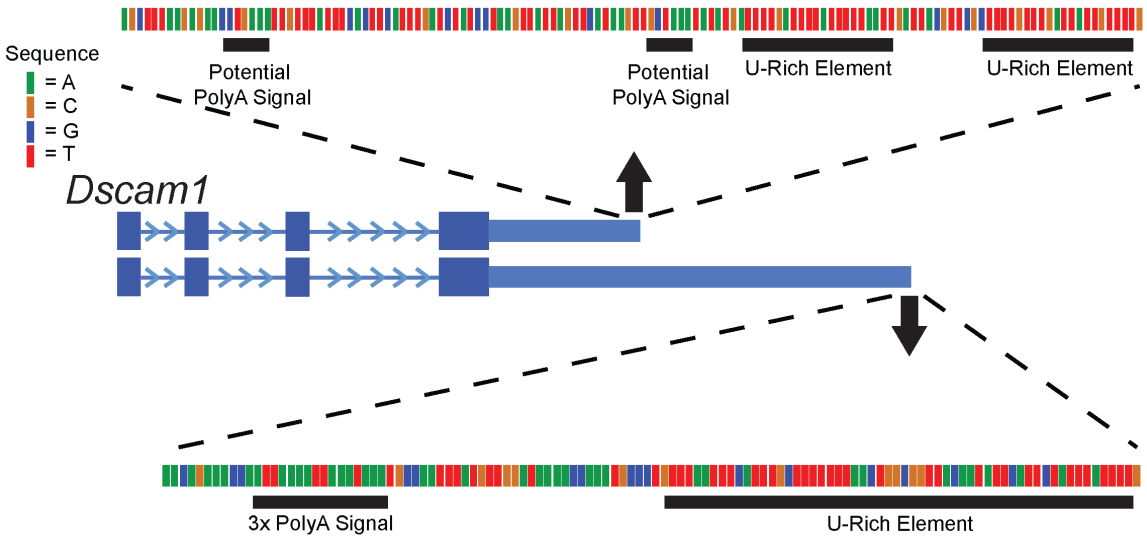


Figure 6. Poly(A) sequence motifs identified at the proximal and distal *Dscam1* 3' UTR cleavage sites. The proximal *Dscam1* 3' UTR does not contain a canonical PAS, but we identified two potential non-canonical signals just upstream of the expected cleavage site. Immediately downstream, two U-rich elements were identified as possible ELAV binding sites. The more proximal element was used as the proximal ELAV binding site probe for the EMSA. The extended 3' UTR contained 3 overlapping PAS and following the expected cleavage site was a long U-rich sequence. This U-rich sequence was used as the distal ELAV binding site probe in the EMSA.

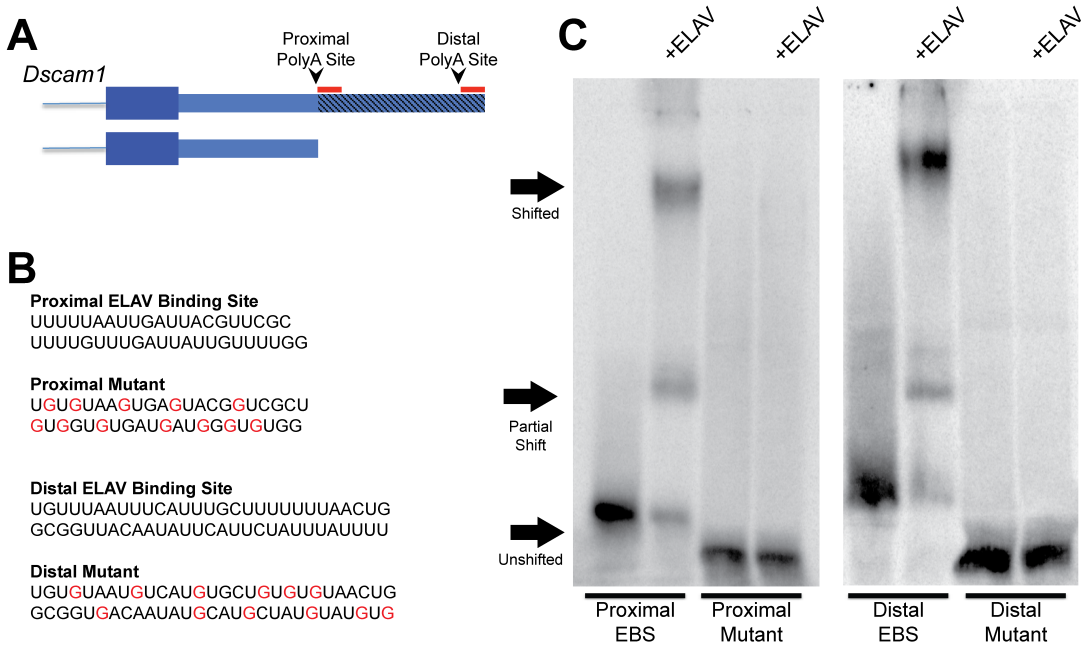
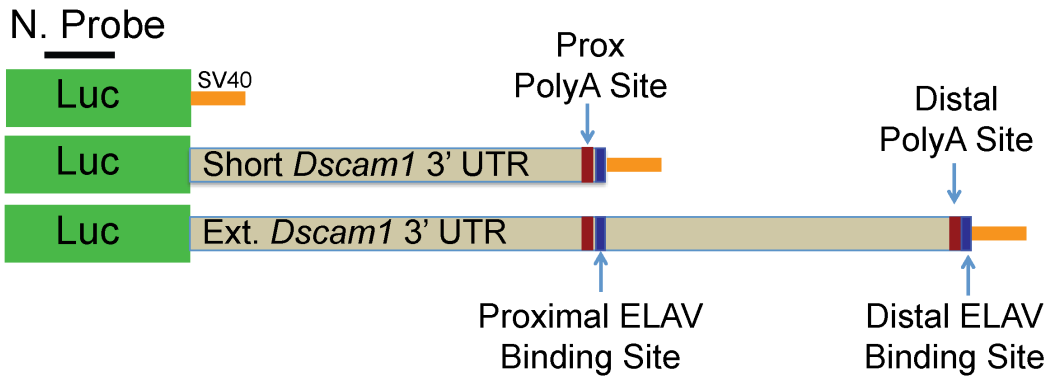


Figure 7. ELAV binds the *Dscam1* 3' UTR. A) Schematic showing the general locations of the proximal and distal ELAV probes used in electrophoretic mobility shift assays (EMSA) the relative to the predicted PolyA sites. B) U-rich RNA probes used to determine ELAV binding capability. To control for non-specific binding, an additional set of probes were generated by disrupting the integrity of U-rich stretches via U to G mutation (indicated in red text). C) EMSAs revealed that upon the addition of recombinant purified ELAV, both proximal and distal radiolabeled probes showed a shift upwards, indicating ELAV had bound the probes. Because ELAV binds as a dodecamer, the partial shift may be explained as the binding of ELAV in a smaller multimeric complex. Both mutated probes were incapable of ELAV binding, evident by the lack of shift.

A



B

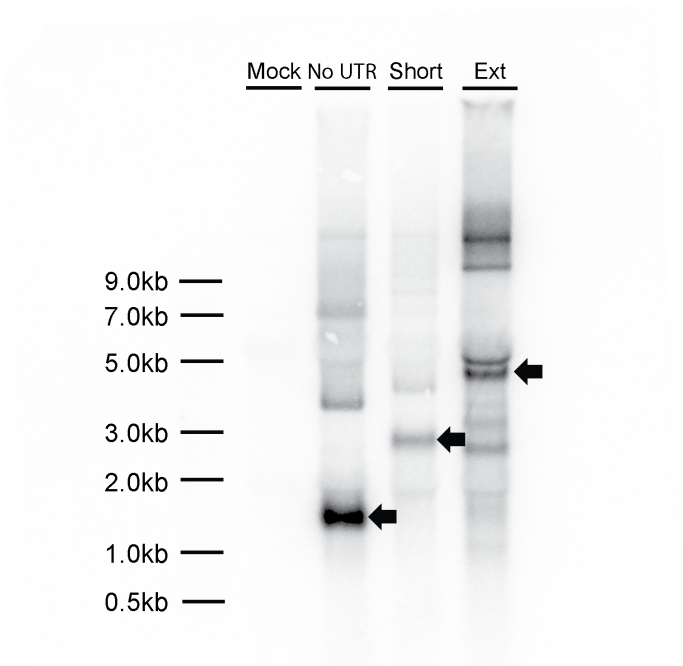


Figure 8. Schematic of Luciferase-*Dscam1* 3' UTR Constructs.
A) *Dscam1* 3' UTR was fused to luciferase in the psicheck-2 vector. By using this reporter system in HeLa cells, we would be able to determine the relative levels of each isoform present via northern blot. Performing a luciferase assay would allow us to determine the amount of luciferase protein made and correlate translation rates to how much of a particular 3' UTR isoform was present. B) Preliminary data showed that expression of the luciferase reporter system produced bands of the expected sizes (arrows), but also yielded extra bands with sizes that could not be explained.

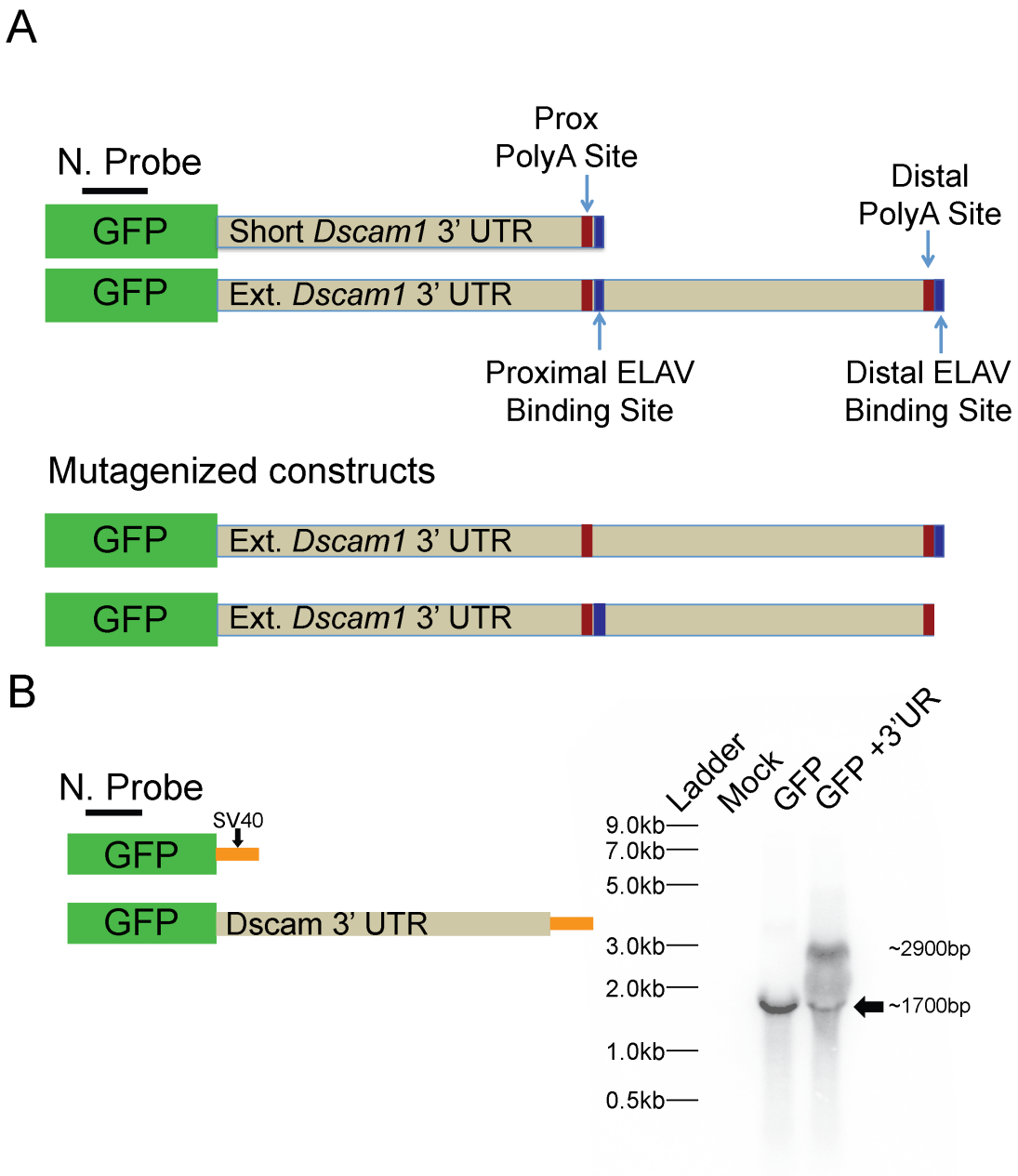


Figure 9. Schematic of GFP-*Dscam1* 3' UTR constructs.
A) Proposed design of *Dscam1* 3' UTR sequence fused with eGFP in a pUAST vector. A second set of constructs are planned to be made by disrupting U-rich regions within the predicted proximal and distal ELAV binding sites via U to C site-directed mutagenesis. We hypothesize that disruption of ELAV binding sites may prevent the extended isoform from being made, or destabilize the transcript. B) Preliminary northern blot data shows sufficient expression levels and expected product bands in S2 cells.

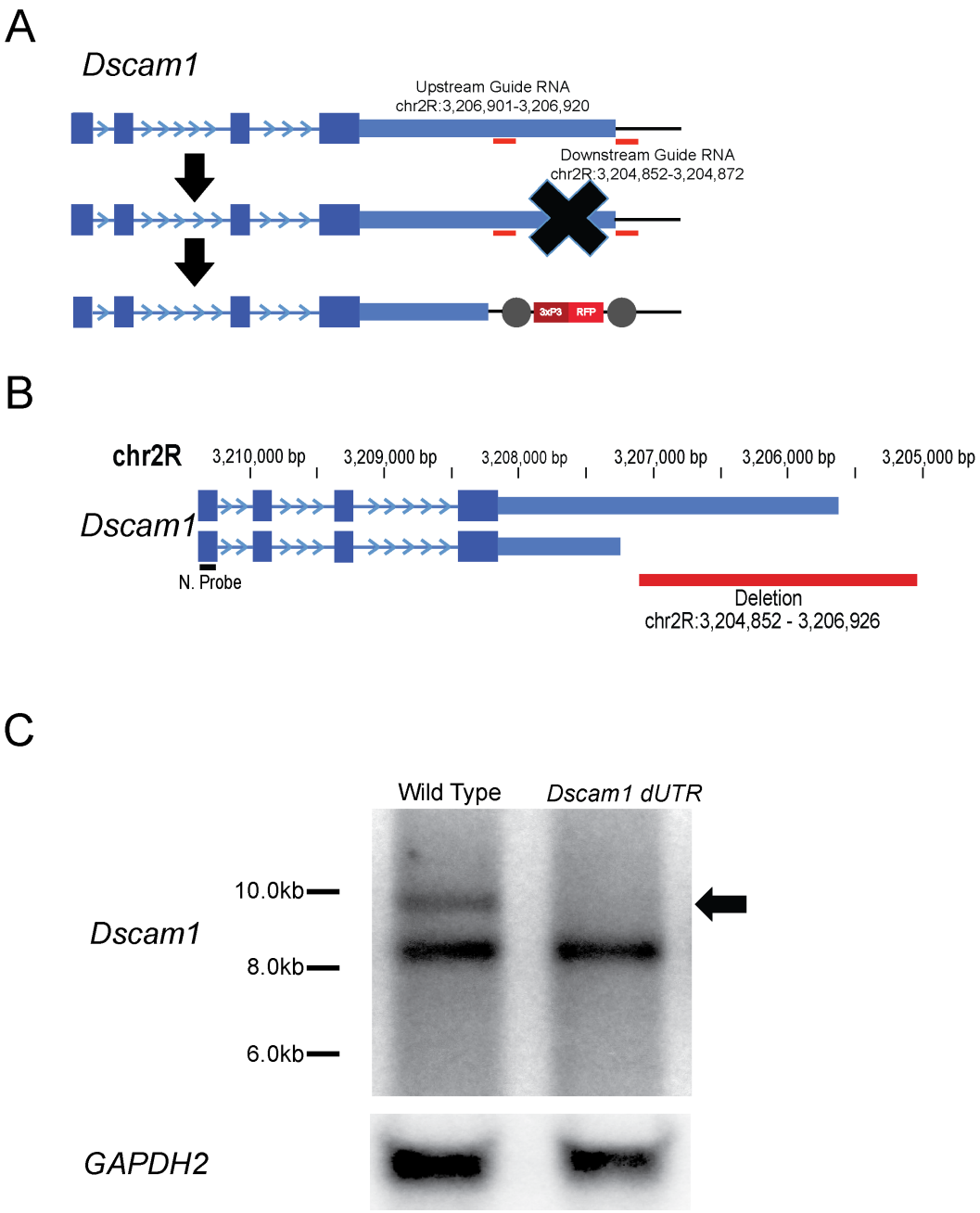


Figure 10. CRISPR/CAS9 deletion strategy and deletion validation.
A) Guide RNAs were designed flanking the extended Dscam1 3' UTR region. Following deletion, a RFP cassette was knocked-in via homology directed repair for use as a visible marker. B) Gene models for the long and short *Dscam1* 3' UTR isoforms. All 3 homozygous Dscam1-dUTR mutant lines generated were verified by Sanger sequencing and produced identical deletions. C) Northern blot for universal *Dscam1* transcript in adult head shows that *Dscam1* dUTR homozygotes do not express the long extended isoform (arrow). GAPDH2 used as a loading control.

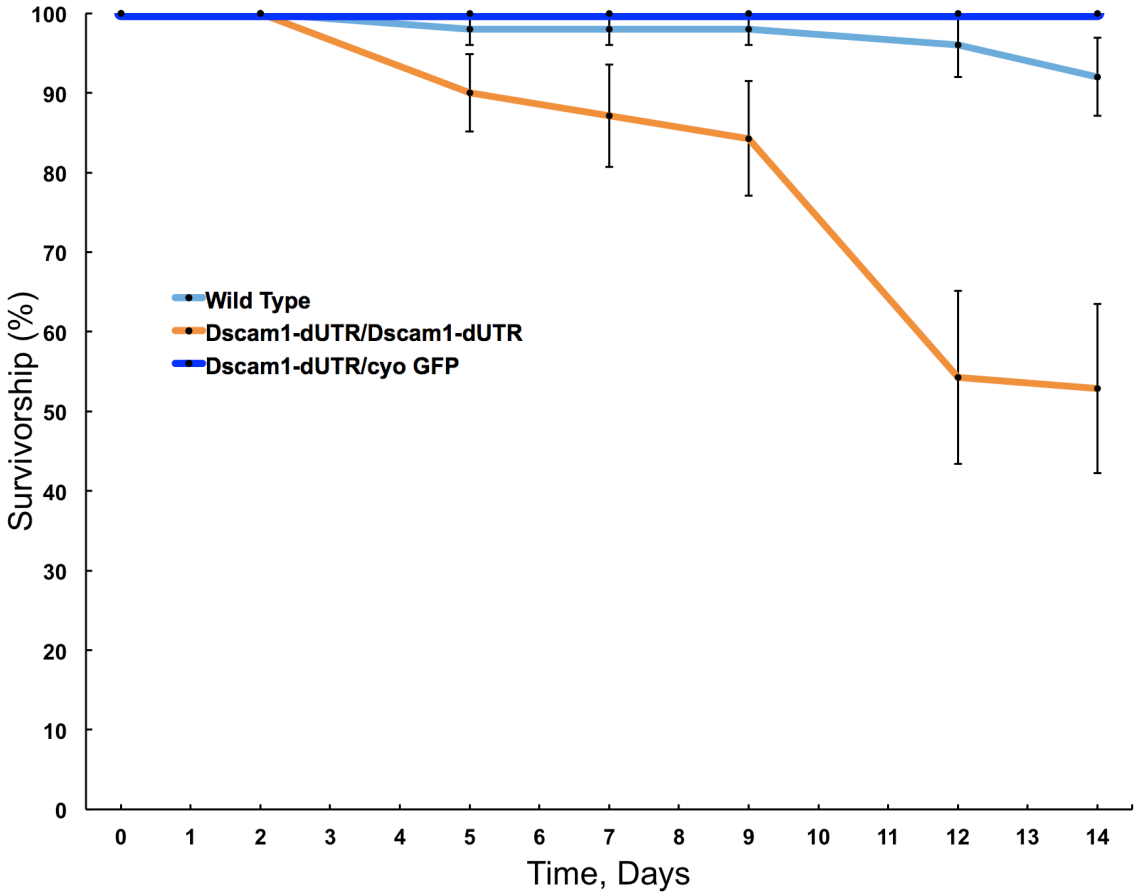


Figure 11. Survivorship of *Dscam1*-dUTR homozygous mutants. *Dscam1*-dUTR mutants showed a severe increase in mortality compared to wild type and balanced lines. Curves represent pooled data from independent experiments, with a total N= 5 per genotype. Assay assisted by Henry Ng.

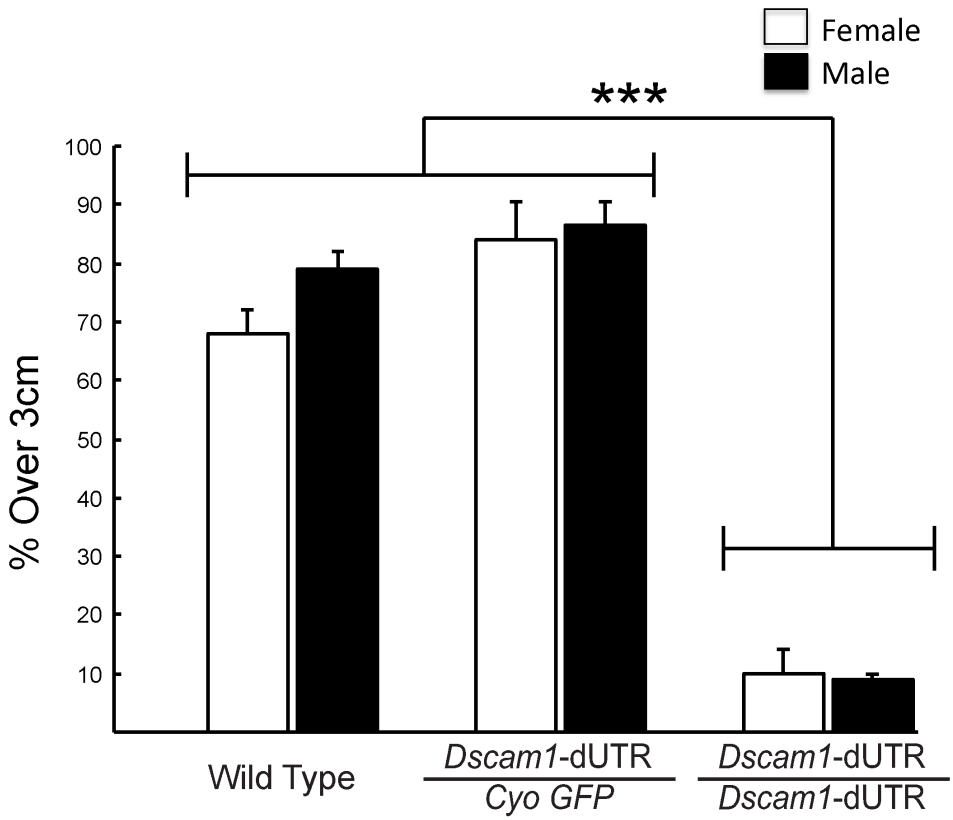


Figure 12. Negative geotactic response is impaired in *Dscam1*-dUTR flies. The data shows a ~86.67% reduction in the ability of *Dscam1*-dUTR(-/-) flies to vertically climb 3cm within the 10 second testing period following mechanical stimulation compared to wild type flies. Each genotype/sex represents pooled data from 2-4 independent experiments where each experiment consisted of a group of 10 flies undergoing 5 repeated trials. There was no significant difference detected between sexes. A Post hoc comparison using the Tukey HSD test indicated a significant difference between the *Dscam1*-dUTR homozygotes and all other genotypes/sexes with a $p < .001$.

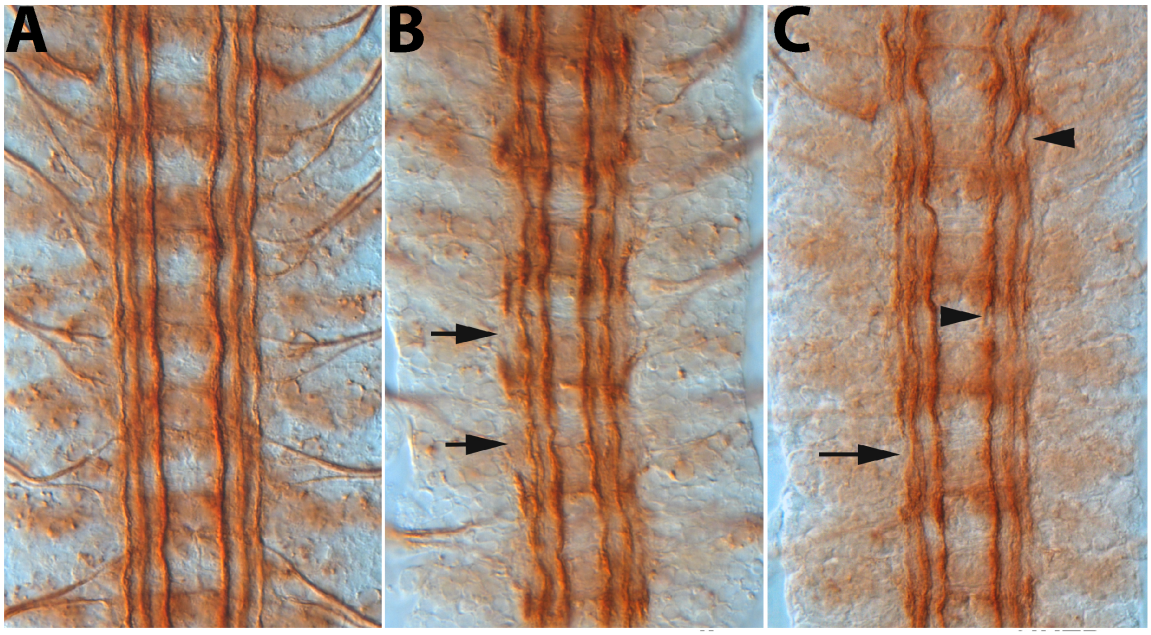


Figure 13. *Dscam1*-dUTR mutants show defects at the ventral nerve cord.
A) Wild type ventral nerve cord at embryonic stage 17. B) Dscam1 null ventral nerve cord shows increased waviness of the longitudinals and severe breaks of the outermost fascicles (arrows). C) Dscam1 extended 3' UTR mutants also displays increased waviness of the longitudinals as well as defasciculation (arrowhead) and some breaks of the outermost fascicles. Images taken by Dr. Thomas Kidd.

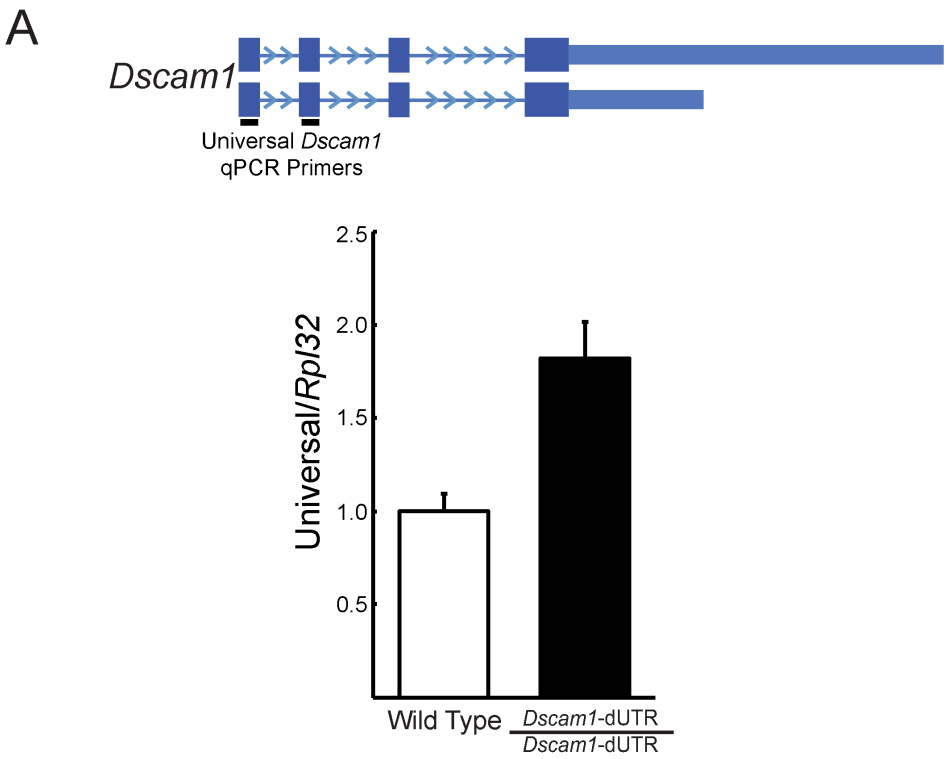


Figure 14. *Dscam1*-dUTR mutants exhibit increased expression of *Dscam1* mRNA, but no change in protein expression.
A) Schematic showing qPCR primers that detect both extended and short 3' UTR isoforms of *Dscam1*. qPCR indicates a 1.82 fold increase of overall *Dscam1* transcript being produced in the *Dscam1*-dUTR mutant compared to wild type in adult head.
B) Western blot analysis of adult head tissues shows almost no change in protein levels of *Dscam1* when normalized to tubulin. Western Blot and qPCR performed by Dr. Zhiping Zhang.

Appendix

I. Plasmid Construct List

Construct Name	Backbone	Inserted Sequence	Restriction Sites Used
pDest60-ELAV	pDest60	ELAV Coding Region	gateway
Psicheck2-Dscam1 Short	Hong Psicheck-2	Short Dscam1 3' UTR	NotI, XhoI
Psicheck2-Dscam1 Long	Hong Psicheck-2	Extended Dscam1 3' UTR	NotI, XhoI
GFP-*Dscam1* Short	pUASTattb eGFP	Short Dscam1 3' UTR	XhoI, XbaI
GFP-*Dscam1* Long	pUASTattb eGFP	Extended Dscam1 3' UTR	XhoI, XbaI
GFP-*Dscam1* Long-prox mut	pUASTattb eGFP	Extended Dscam1 3' UTR w/ mutated proximal ELAV binding site	XhoI, XbaI
GFP-*Dscam1* Long-distal mut	pUASTattb eGFP	Extended Dscam1 3' UTR w/ mutated distal ELAV binding site	XhoI, XbaI

II. List of Primers and Probes

Primer/Probe Name	Sequence 5' to 3'	Description
rp20_DscamUni_CLN1	GATCGCGGCCGCTAATTTGTAAGCGCCCTCTGCG	Amplify *Dscam1* short 3' UTR
rp21_DscamShort_CLN2	GATCCTCGAGTGGTGGTGTCTCCACAGATTTC	
rp20_DscamUni_CLN1	GATCGCGGCCGCTAATTTGTAAGCGCCCTCTGCG	Amplify *Dscam1* Extended 3' UTR
rp22_DscamLong_Cln3	GATCCTCGAGACCGGAATTGACTGAACGATTC	
ELAV_Forward	CACCGACTTTATTATGGCAAATACCGGA	Amplify ELAV for Protein Expression
ELAV_Reverse	CTTGGCTTTGTTGGTCTTGAAGCT	
Dscam_prox_elav site3	caaaacaataatcaaacaaaagcgaacgtaatcaatttaaaatgagatcgagatcatatatttagcaacttatgaactaataacttaggatagat	Mutagenesis Primers for 1st half of proximal ELAV binding site
Dscam_prox_elav site4	atctatcctaagttattagttcataagttgctaaatatatgatctcgatctcattttaaattgattacgttcgcttttgtttgattattgttttg	
Dscam_prox_elav_site5	aataatcaaacaaaagcgaacgtaatcagtttgagatgagatcgagatcatatatttagcaacttatg	Mutagenesis Primers for 2nd half of proximal ELAV binding site
Dscam_prox_elav_site6	cataagttgctaaatatatgatctcgatctcatctcaaactgattacgttcgcttttgtttgattatt	
Dscam_ELAV_Dist_mut1	gactgaacgattccttttgaatcagagatagataggatggatattgtgaccgccagttaaaaaaagcaaatgaaattaaacagggc	Mutagenesis Primers for 1st half of distall ELAV binding site
Dscam_ELAV_Dist_mut2	gccctgtttaatttcatttgctttttttaactggcggtcacaatatccatcctatctatctctgattcaaaaggaatcgttcagtc	
Dscam_ELAV_Dist_mut3	atagataggatggatattgtgaccgccagttagagagagcagatgagattagacagggcatttggttttatccaacatataaattg	Mutagenesis Primers for 2nd half of distal ELAV binding site
Dscam_ELAV_Dist_mut4	caatttatgttggataaaaccaaatgccctgtctaatctcatctgctctctctaactggcggtcacaatatccatcctatctat	
Proximal *Dscam1 EMSA* RNA probe	UUUUUAAUUGAUUACGUUCGCUUUUGUUUGAUUAUUGUUUUGG	Proximal RNA probe to test ELAV binding in EMSA
Proximal *Dscam1* mutated EMSA RNA Probe	UGUGUAAGUGAGUACGGUCGCUGUGGUGUGAUGAUGGGUGUGG	Proximal U to G mutated RNA probe to test ELAV binding in EMSA
Distal *Dscam1 EMSA* RNA probe	UGUUUAAUUUCAUUUGCUUUUUUUAACUGGCGGUUACAAUAUUCAUUCUAUUUAUUUU	Distal RNA probe to test ELAV binding in EMSA
Distal *Dscam1* mutated EMSA RNA Probe	UGUGUAAUGUCAUGUGCUGUGUGUAACUGGCGGUGACAAUAUGCAUGCUAUGUAUGUG	Distal U to G mutated RNA probe to test ELAV binding in EMSA
rp_qf_67_DscamUni2_Left	TCCGGAGTACAGGCTACCG	qPCR primer set to detect universal *Dscam1* transcript
rp_qf_68_DscamUni2_Right	GGACAGTCCTCAATCTACACG	
rp_qf_6_dscam_ex3UTR	GCGTTTTAAACTGCCTGTCC	qPCR primer set to detect extended *Dscam1 3' UTR* transcript
rp_qf_7_dscam_ex3UTR	CAACTTCAACGCACATCAGG	
rp_NF_1_DscamUni	GGTCTTGATCACTCCGTTGG	Used to make *Dscam1* Universal Northern Probe
rp_NF_2_DscamUni	CACCTACAACATTCGCATCG	
N.f.709_GAPDH2	GGCATCCACTCACTTGAAGG	Used to make *GAPDH2* Northern Probe
N.f.710_GAPDH2	TCAGCTTCACGAACTTGTCG	
R.Luc probe Forward	TGATCGGAATGGGTAAGTCC	Used to make Luciferase Northern Probe
R.Luc probe Reverse	GCTATTGTCGAGGGAGCTAA	
rp_N_1_GFP	TAAACGGCCACAAGTTCAGC	Used to make eGFP Northern Probe
rp_N_2_GFP	CTTGTACAGCTCGTCCATGC	
Proximal *Dscam1 EMSA* RNA probe	UUUUUAAUUGAUUACGUUCGCUUUUGUUUGAUUAUUGUUUUGG	Proximal RNA probe to test ELAV binding in EMSA
Proximal *Dscam1* mutated EMSA RNA Probe	UGUGUAAGUGAGUACGGUCGCUGUGGUGUGAUGAUGGGUGUGG	Proximal U to G mutated RNA probe to test ELAV binding in EMSA
Distal *Dscam1 EMSA* RNA probe	UGUUUAAUUUCAUUUGCUUUUUUUAACUGGCGGUUACAAUAUUCAUUCUAUUUAUUUU	Distal RNA probe to test ELAV binding in EMSA
Distal *Dscam1* mutated EMSA RNA Probe	UGUGUAAUGUCAUGUGCUGUGUGUAACUGGCGGUGACAAUAUGCAUGCUAUGUAUGUG	Distal U to G mutated RNA probe to test ELAV binding in EMSA
Dscam_deletion_Seq1	AATGAGCCGCAATTAGTTGA	Sequencing primers for *Dscam1* dUTR ends
Dscam_deletion_Seq2	GAACATGAGTTATACGCATT	
rp_seq73_DSCAM_UTR	AGCAACTGAGATGGATTCTT	Sequencing primers for *Dscam1* 3' UTR Sanger sequencing
rp_seq74_DSCAM_UTR	CCAAAACCCCTAACCCGAAA	
rp_seq75_DSCAM_UTR	CACATCACGAACTGCATTTA	
rp_seq76_DSCAM_UTR	TTTGAAGAAGAAAACACTTA	
rp_seq77_DSCAM_UTR	AGAGAGAATCAATACCCCAG	

III. *Dscam1* 3' UTR Sequence

chr2R:3,204,518 - 3,208,190

```
                  20                  40                  60
TAATTTGTAAGCGCCCTCTGCGGCGGTGGG CGTGGCATTTTAGCTTAGTGTTTTAGGTGC
ATTAAACATTCGCGGGAGACGCCGCCACCC GCACCGTAAAATCGAATCACAAAATCCACG

                  80                 100                 120
ATGAAGTGCGTGGGAGTTGAATGACACTTA GAGACACAGGACACAGATACACACAGACAG
TACTTCACGCACCCTCAACTTACTGTGAAT CTCTGTGTCCTGTGTCTATGTGTGTCTGTC

                 140                 160                 180
ATCATCAGGACACAAGGAAAGAGATAGAGA AAGAGAGGGAGTTATATACCGAATATACAC
TAGTAGTCCTGTGTTCCTTTCTCTATCTCT TTCTCTCCCTCAATATATGGCTTATATGTG

                 200                 220                 240
GTAAAATATTCAACTACTTCGAGTAATTAT AGCGAAGGAATGCAAGTTTCAAGAAACATA
CATTTTATAAGTTGATGAAGCTCATTAATA TCGCTTCCTTACGTTCAAAGTTCTTTGTAT

                 260                 280                 300
TTTATATTGTCTATGCAAGCAGCAAACCAA ACTTACATAAAGTATATAAGTACATTCAAT
AAATATAACAGATACGTTCGTCGTTTGGTT TGAATGTATTTCATATATTCATGTAAGTTA

                 320                 340                 360
GGATAACTACTATAAGAGAGTGCTGGAGAC ACTGAAGGAAATTACTCAAGTTATTCGAGA
CCTATTGATGATATTCTCTCACGACCTCTG TGACTTCCTTTAATGAGTTCAATAAGCTCT

                 380                 400                 420
GGAACAGAAAACGAAATCTCAAGATCAAGA TCATACTTACGTATTCGAACTAATTTTCAA
CCTTGTCTTTTGCTTTAGAGTTCTAGTTCT AGTATGAATGCATAAGCTTGATTAAAAGTT

                 440                 460                 480
TAATGTACTTTAGTTTTAAGCGGTGTCTAC CGAAATATCAGAGCAACAGCAACTGAGATG
ATTACATGAAATCAAAATTCGCCACAGATG GCTTTATAGTCTCGTTGTCGTTGACTCTAC

                 500                 520                 540
GATTCTTGTGTGTAAGAGGAGATGATTCGA AATCTGTTTGTGTAAACTTTTGCCATCTAT
CTAAGAACACACATTCTCCTCTACTAAGCT TTAGACAAACACATTTGAAAACGGTAGATA

                 560                 580                 600
GCCAGGAGCAGAAAACTCATCCTTTGCCTA GCCTAAGTCAAAGTCATAATATTAACTCGT
CGGTCCTCGTCTTTTGAGTAGGAAACGGAT CGGATTCAGTTTCAGTATTATAATTGAGCA

                 620                 640                 660
CTAATTGATATATTTAGTCGTAACATATGC GTAATATGTATGATTTCTTCCTATGCATTT
GATTAACTATATAAATCAGCATTGTATACG CATTATACATACTAAAGAAGGATACGTAAA
```

```
                  680                 700                 720
TATAATTCGTATGGACAGCGCACATTTGGA GTTTTGTTTGCGTTTTGCGTTGACTTGTGA
ATATTAAGCATACCTGTCGCGTGTAAACCT CAAAACAAACGCAAAACGCAACTGAACACT

                  740                 760                 780
TGATTGTTAATTTACAGAGCTTTGTGAGTG AGTACTTACGATTGTGCTCATTACTCGCTT
ACTAACAATTAAATGTCTCGAAACACTCAC TCATGAATGCTAACACGAGTAATGAGCGAA

                  800                 820                 840
TGTTCTTTATTTACAGACAAGTGATTGTCC TATCTTAGTTCGCCCTTATTTTTAATTCAA
ACAAGAAATAAATGTCTGTTCACTAACAGG ATAGAATCAAGCGGGAATAAAAATTAAGTT

                  860                 880                 900
AAATGCGCTTCCAACAGAACAAATCGAAAC TGTTCAATGCACAAATCGCAGCCACAACAA
TTTACGCGAAGGTTGTCTTGTTTAGCTTTG ACAAGTTACGTGTTTAGCGTCGGTGTTGTT

                  920                 940                 960
TTGAATGTATTATACAAGATCCCCAGGACA CCGAAAAACCAAAACCCCTAACCCGAAACA
AACTTACATAATATGTTCTAGGGGTCCTGT GGCTTTTTGGTTTTGGGGATTGGGCTTTGT

                  980                1000                1020
CAGATACACTCAAGCACGATTTATTAAACC AATTCTGATATAGACCGTTCTGAACATGAG
GTCTATGTGAGTTCGTGCTAAATAATTTGG TTAAGACTATATCTGGCAAGACTTGTACTC

                 1040                1060                1080
TTATACGCATTTGAAGCTTTAAGAGAACCT GAAATTGTTTATTTCACTTCATTTGATCTA
AATATGCGTAAACTTCGAAATTCTCTTGGA CTTTAACAAATAAAGTGAAGTAAACTAGAT
                    Possible PolyA Sginal       Prox ELAV binding site
                 1100                1120                1140
TCCTAAGTTATTAGTTCATAAGTTGCTAAA TATATGATTTTGATTTTATTTTTAATTGAT
AGGATTCAATAATCAAGTATTCAACGATTT ATATACTAAAACTAAAATAAAAATTAACTA
              U-Rich element                Proximal Cleavage Site
                 1160                1180                1200
TACGTTCGCTTTTGTTTGATTATTGTTTTG GGGTTCCGCCCATTGCGGCCAGAAATCTGT
ATGCAAGCGAAAACAAACTAATAACAAAAC CCCAAGGCGGGTAACGCCGGTCTTTAGACA

                 1220                1240                1260
GGAGACACCACCAACACTTCAAAGACTGAC TAATGCAAGCTAAAAACGAAGCCGGATCAA
CCTCTGTGGTGGTTGTGAAGTTTCTGACTG ATTACGTTCGATTTTTGCTTCGGCCTAGTT

                 1280                1300                1320
TCCAAGTTTTGCGGGATGAAACCGATGCAA TGGAGCTACAGAATATCCAGTATTTTGCTC
AGGTTCAAAACGCCCTACTTTGGCTACGTT ACCTCGATGTCTTATAGGTCATAAAACGAG

                 1340                1360                1380
AACACGTAGAATAGCAGGGCATTTCCGGGT GAATCAACTGCCATAAGATCGGAACTGGAT
TTGTGCATCTTATCGTCCCGTAAAGGCCCA CTTAGTTGACGGTATTCTAGCCTTGACCTA
```

```
                1400                1420                1440
TTTGATCCTCGTCCCTTAAGCACATCACGA ACTGCATTTATCCAATGGGAGATTTATTGT
AAACTAGGAGCAGGGAATTCGTGTAGTGCT TGACGTAAATAGGTTACCCTCTAAATAACA

                1460                1480                1500
CAGGTTTATCAGGCGTTTAGTTTCACTTCA ATCTTCTTCTCCTTAAGTTCCCACTTAACC
GTCCAAATAGTCCGCAAATCAAAGTGAAGT TAGAAGAAGAGGAATTCAAGGGTGAATTGG

                1520                1540                1560
TATCCGAACTAATACTCACTCTTACCCAGA CTATACCAGTTTTATATAGATATATACGTA
ATAGGCTTGATTATGAGTGAGAATGGGTCT GATATGGTCAAAATATATCTATATATGCAT

                1580                1600                1620
TATAAATATATAAATGTAGTGTTGTGAAAA AATAAGATTGCTTACCTATATGAGTGTACT
ATATTTATATATTTACATCACAACACTTTT TTATTCTAACGAATGGATATACTCACATGA

                1640                1660                1680
ACAAATTTTCATGGCAAACGAACTACAAAT ATTAAGTAGTGTATAGTAAAGGAAATGGCA
TGTTTAAAAGTACCGTTTGCTTGATGTTTA TAATTCATCACATATCATTTCCTTTACCGT

                1700                1720                1740
TTTAATCAATAAGCAAACGAAAGGAATACA TCTACAAATTTTCAAGTAATTAAAATAGCA
AAATTAGTTATTCGTTTGCTTTCCTTATGT AGATGTTTAAAAGTTCATTAATTTTATCGT

                1760                1780                1800
AGCAATATTCTAGAGTCAAAAGAACGTTGG ATTTCACAAATTGTAACGAACAAACACGAT
TCGTTATAAGATCTCAGTTTTCTTGCAACC TAAAGTGTTTAACATTGCTTGTTTGTGCTA

                1820                1840                1860
TTTAATTACACTCGTAAAATTTATAAACTT TTCGGCAAATCAAAATCAAAACAAAAACTT
AAATTAATGTGAGCATTTTAAATATTTGAA AAGCCGTTTAGTTTTAGTTTTGTTTTTGAA

                1880                1900                1920
TTGGCATTAAAAAGAAAAAGAAAACGAAAA CAAAAACATAAACAAAGCGGTTTGAAGAAG
AACCGTAATTTTTCTTTTTCTTTTGCTTTT GTTTTTGTATTTGTTTCGCCAAACTTCTTC

                1940                1960                1980
AAAACACTTAGGCTAAGTACGCTAATATAA CATAATCCATTTTTCATACTATATATCATG
TTTTGTGAATCCGATTCATGCGATTATATT GTATTAGGTAAAAAGTATGATATATAGTAC

                2000                2020                2040
CATGAATAATGTATACTCGTAATTCGAATC AATCCGAGTCGGATGCTTTTTGAGTTGCTT
GTACTTATTACATATGAGCATTAAGCTTAG TTAGGCTCAGCCTACGAAAAACTCAACGAA

                2060                2080                2100
CATTGTCAAAAGGAATATGGCGAAATTAAC GTTTTGAATTCATTAGACGCGCTTCATACT
GTAACAGTTTTCCTTATACCGCTTTAATTG CAAAACTTAAGTAATCTGCGCGAAGTATGA
```

```
                2120                2140                2160
TGTAATTATTTAAGCGATAAGCGTACATGC ATATACAAACATATACAGATAATTTTTTAC
ACATTAATAAATTCGCTATTCGCATGTACG TATATGTTTGTATATGTCTATTAAAAAATG

                2180                2200                2220
TTAACTTTACATAAAACGCATTAATCAAAA TTCAACTCCCAAGAAACATTAACCAGAAGG
AATTGAAATGTATTTTGCGTAATTAGTTTT AAGTTGAGGGTTCTTTGTAATTGGTCTTCC

                2240                2260                2280
TCAGACACAAAATCCAAGACACTCTCACAC GAAGGACATTCAACTTCAACGCACATCAGG
AGTCTGTGTTTTAGGTTCTGTGAGAGTGTG CTTCCTGTAAGTTGAAGTTGCGTGTAGTCC

                2300                2320                2340
CTCGACATCGAGGATAACTAAGCACGATCA GCACACATTTGAATTCTGTAACCAGGCGGA
GAGCTGTAGCTCCTATTGATTCGTGCTAGT CGTGTGTAAACTTAAGACATTGGTCCGCCT

                2360                2380                2400
CAGGCAGTTTAAAACGCATTTATACTAATC ACTCGCTCCTTCTTAGCAGAACGAATATTA
GTCCGTCAAATTTTGCGTAAATATGATTAG TGAGCGAGGAAGAATCGTCTTGCTTATAAT

                2420                2440                2460
TATGAAATTAAGAGAGAATCAATACCCCAG TGGAAACATTCGTAATCATTAAGCGCGGCA
ATACTTTAATTCTCTCTTAGTTATGGGGTC ACCTTTGTAAGCATTAGTAATTCGCGCCGT

                2480                2500                2520
TATACGGTAAACCCCAAATTGAATACTACA AATATGTGTATTACTAATTACATTTACGAA
ATATGCCATTTGGGGTTTAACTTATGATGT TTATACACATAATGATTAATGTAAATGCTT

                2540                2560                2580
AGCATACATATATTAACTATATTAAATATT GAATGTATTGAATTGACAACAGTCGGCATG
TCGTATGTATATAATTGATATAATTTATAA CTTACATAACTTAACTGTTGTCAGCCGTAC

                2600                2620                2640
CTGGGGATTCACTATCCTGAACCGCTTCAC GTGGAACGGATTGGTAATCCCGCCCCAAAC
GACCCCTAAGTGATAGGACTTGGCGAAGTG CACCTTGCCTAACCATTAGGGCGGGGTTTG

                2660                2680                2700
TCGATCACTTTGCATGTCCGACCAGCGAAG AAAGAATTTGTCAGTTTCCAGTTTCAGTGT
AGCTAGTGAAACGTACAGGCTGGTCGCTTC TTTCTTAAACAGTCAAAGGTCAAAGTCACA
                    Distal PolyA Signals
                2720                2740                2760
CATAAAAGTCAAACAGAAACCAATTAAAAT TAAATAAATGCCAATTTATGTTGGATAAAA
GTATTTTCAGTTTGTCTTTGGTTAATTTTA ATTTATTTACGGTTAAATACAACCTATTTT

                2780                2800                2820
CCAAATGCCCTGTTTAATTTCATTTGCTTT TTTTAACTGGCGGTTACAATATTCATTCTA
GGTTTACGGGACAAATTAAAGTAAACGAAA AAAATTGACCGCCAATGTTATAAGTAAGAT
Distal Cleavage Site                Distal ELAV Binding Site
                2840                2860
TTTATTTTGGATTCAAAAGGAATCGTTCAG TCAATTCCGGT
AAATAAAACCTAAGTTTTCCTTAGCAAGTC AGTTAAGGCCA
```

References

1. Lodish H, B.A., Zipursky SL, et al., *Molecular Definition of a Gene*, in *Molecular Cell Biology*. 2000, W. H. Freeman: New York.
2. Alberts B, J.A., Lewis J, et al., *An Overview of the Cell Cycle*, in *Molecular Biology of the Cell*. 2002, Garland Science: New York.
3. Maduro, M.F., *Cell fate specification in the C. elegans embryo.* Dev Dyn, 2010. **239**(5): p. 1315-29.
4. Guo, J., *Transcription: the epicenter of gene expression.* J Zhejiang Univ Sci B, 2014. **15**(5): p. 409-11.
5. Mignone, F., et al., *Untranslated regions of mRNAs.* Genome Biol, 2002. **3**(3): p. Reviews0004.
6. Brody, Y., et al., *The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing.* PLoS Biol, 2011. **9**(1): p. e1000573.
7. Mercer, T.R., et al., *Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome.* Genome Res, 2010. **20**(12): p. 1639-50.
8. Inukai, S., K.H. Kock, and M.L. Bulyk, *Transcription factor-DNA binding: beyond binding site motifs.* Curr Opin Genet Dev, 2017. **43**: p. 110-119.
9. Carlberg, C. and F. Molnár, *Mechanisms of gene regulation*. 2014: Springer.
10. Spitz, F. and E.E. Furlong, *Transcription factors: from enhancer binding to developmental control.* Nat Rev Genet, 2012. **13**(9): p. 613-26.
11. Barrett, L.W., S. Fletcher, and S.D. Wilton, *Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements.* Cell Mol Life Sci, 2012. **69**(21): p. 3613-34.
12. Matoulkova, E., et al., *The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells.* RNA Biol, 2012. **9**(5): p. 563-76.
13. Berkovits, B.D. and C. Mayr, *Alternative 3' UTRs act as scaffolds to regulate membrane protein localization.* Nature, 2015. **522**(7556): p. 363-7.
14. Martin, K.C. and A. Ephrussi, *mRNA localization: gene expression in the spatial dimension.* Cell, 2009. **136**(4): p. 719-30.
15. Fabian, M.R., N. Sonenberg, and W. Filipowicz, *Regulation of mRNA translation and stability by microRNAs.* Annu Rev Biochem, 2010. **79**: p. 351-79.
16. Andreassi, C. and A. Riccio, *To localize or not to localize: mRNA fate is in 3'UTR ends.* Trends Cell Biol, 2009. **19**(9): p. 465-74.
17. Friedlander, M.R., et al., *Evidence for the biogenesis of more than 1,000 novel human microRNAs.* Genome Biol, 2014. **15**(4): p. R57.
18. Schertel, C., et al., *Functional characterization of Drosophila microRNAs by a novel in vivo library.* Genetics, 2012. **192**(4): p. 1543-52.
19. Cannell, I.G., Y.W. Kong, and M. Bushell, *How do microRNAs regulate gene expression?* Biochem Soc Trans, 2008. **36**(Pt 6): p. 1224-31.

20. Lu, J. and A.G. Clark, *Impact of microRNA regulation on variation in human gene expression.* Genome Res, 2012. **22**(7): p. 1243-54.
21. Giraldez, A.J., et al., *MicroRNAs regulate brain morphogenesis in zebrafish.* Science, 2005. **308**(5723): p. 833-8.
22. Yue, B., et al., *Expression Profiles Analysis and Functional Characterization of MicroRNA-660 in Skeletal Muscle Differentiation.* J Cell Biochem, 2017. **118**(8): p. 2387-2394.
23. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**(6772): p. 901-6.
24. Miska, E.A., et al., *Most Caenorhabditis elegans microRNAs are individually not essential for development or viability.* PLoS Genet, 2007. **3**(12): p. e215.
25. Bejarano, F., et al., *A genome-wide transgenic resource for conditional expression of Drosophila microRNAs.* Development, 2012. **139**(15): p. 2821-31.
26. Carthew, R.W., P. Agbu, and R. Giri, *MicroRNA function in Drosophila melanogaster.* Semin Cell Dev Biol, 2017. **65**: p. 29-37.
27. Seliger, B., *Immune modulatory microRNAs as a novel mechanism to revert immune escape of tumors.* Cytokine Growth Factor Rev, 2017.
28. Lunde, B.M., C. Moore, and G. Varani, *RNA-binding proteins: modular design for efficient function.* Nat Rev Mol Cell Biol, 2007. **8**(6): p. 479-90.
29. Colgan, D.F. and J.L. Manley, *Mechanism and regulation of mRNA polyadenylation.* Genes Dev, 1997. **11**(21): p. 2755-66.
30. Lewis, J.D., S.I. Gunderson, and I.W. Mattaj, *The influence of 5' and 3' end structures on pre-mRNA metabolism.* J Cell Sci Suppl, 1995. **19**: p. 13-9.
31. Wahle, E., *Poly(A) tail length control is caused by termination of processive synthesis.* J Biol Chem, 1995. **270**(6): p. 2800-8.
32. Eckmann, C.R., C. Rammelt, and E. Wahle, *Control of poly(A) tail length.* Wiley Interdiscip Rev RNA, 2011. **2**(3): p. 348-61.
33. Ustyantsev, I.G., et al., *[Canonical and noncanonical RNA polyadenylation].* Mol Biol (Mosk), 2017. **51**(2): p. 262-273.
34. Shi, Y., et al., *Molecular architecture of the human pre-mRNA 3' processing complex.* Mol Cell, 2009. **33**(3): p. 365-76.
35. Shi, Y. and J.L. Manley, *The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site.* Genes Dev, 2015. **29**(9): p. 889-97.
36. McCracken, S., et al., *The C-terminal domain of RNA polymerase II couples mRNA processing to transcription.* Nature, 1997. **385**(6614): p. 357-61.
37. Mandel, C.R., Y. Bai, and L. Tong, *Protein factors in pre-mRNA 3'-end processing.* Cell Mol Life Sci, 2008. **65**(7-8): p. 1099-122.
38. Tian, B., et al., *A large-scale analysis of mRNA polyadenylation of human and mouse genes.* Nucleic Acids Res, 2005. **33**(1): p. 201-12.
39. Beaudoing, E., et al., *Patterns of variant polyadenylation signal usage in human genes.* Genome Res, 2000. **10**(7): p. 1001-10.

40. Chan, S.L., et al., *CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing.* Genes Dev, 2014. **28**(21): p. 2370-80.
41. Schonemann, L., et al., *Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33.* Genes Dev, 2014. **28**(21): p. 2381-93.
42. Chen, F., C.C. MacDonald, and J. Wilusz, *Cleavage site determinants in the mammalian polyadenylation signal.* Nucleic Acids Res, 1995. **23**(14): p. 2614-20.
43. Sheets, M.D., S.C. Ogg, and M.P. Wickens, *Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro.* Nucleic Acids Res, 1990. **18**(19): p. 5799-805.
44. Mandel, C.R., et al., *Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease.* Nature, 2006. **444**(7121): p. 953-6.
45. Chou, Z.F., F. Chen, and J. Wilusz, *Sequence and position requirements for uridylate-rich downstream elements of polyadenylation signals.* Nucleic Acids Res, 1994. **22**(13): p. 2525-31.
46. Gil, A. and N.J. Proudfoot, *Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation.* Cell, 1987. **49**(3): p. 399-406.
47. MacDonald, C.C., J. Wilusz, and T. Shenk, *The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location.* Mol Cell Biol, 1994. **14**(10): p. 6647-54.
48. Ruegsegger, U., K. Beyer, and W. Keller, *Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors.* J Biol Chem, 1996. **271**(11): p. 6107-13.
49. Venkataraman, K., K.M. Brown, and G.M. Gilmartin, *Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition.* Genes Dev, 2005. **19**(11): p. 1315-27.
50. Chan, S., E.A. Choi, and Y. Shi, *Pre-mRNA 3'-end processing complex assembly and function.* Wiley Interdiscip Rev RNA, 2011. **2**(3): p. 321-35.
51. Bienroth, S., W. Keller, and E. Wahle, *Assembly of a processive messenger RNA polyadenylation complex.* Embo j, 1993. **12**(2): p. 585-94.
52. Ezkurdia, I., et al., *Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes.* Hum Mol Genet, 2014. **23**(22): p. 5866-78.
53. Church, D.M., et al., *Lineage-specific biology revealed by a finished genome assembly of the mouse.* PLoS Biol, 2009. **7**(5): p. e1000112.
54. Hillier, L.W., et al., *Genomics in C. elegans: so many genes, such a little worm.* Genome Res, 2005. **15**(12): p. 1651-60.
55. Lin, M.F., et al., *Revisiting the protein-coding gene catalog of Drosophila melanogaster using 12 fly genomes.* Genome Res, 2007. **17**(12): p. 1823-36.
56. Richards, E.J. and S.C. Elgin, *Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects.* Cell, 2002. **108**(4): p. 489-500.
57. Liu, G., J.S. Mattick, and R.J. Taft, *A meta-analysis of the genomic and transcriptomic composition of complex life.* Cell Cycle, 2013. **12**(13): p. 2061-72.

58. Deveson, I.W., et al., *The Dimensions, Dynamics, and Relevance of the Mammalian Noncoding Transcriptome.* Trends Genet, 2017.
59. Elkon, R., A.P. Ugalde, and R. Agami, *Alternative cleavage and polyadenylation: extent, regulation and function.* Nat Rev Genet, 2013. **14**(7): p. 496-506.
60. Alt, F.W., et al., *Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends.* Cell, 1980. **20**(2): p. 293-301.
61. Early, P., et al., *Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways.* Cell, 1980. **20**(2): p. 313-9.
62. Rogers, J., et al., *Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin mu chain.* Cell, 1980. **20**(2): p. 303-12.
63. Setzer, D.R., et al., *Size heterogeneity in the 3' end of dihydrofolate reductase messenger RNAs in mouse cells.* Cell, 1980. **22**(2 Pt 2): p. 361-70.
64. Weber, A.P., *Discovering New Biology through Sequencing of RNA.* Plant Physiol, 2015. **169**(3): p. 1524-31.
65. Derti, A., et al., *A quantitative atlas of polyadenylation in five mammals.* Genome Res, 2012. **22**(6): p. 1173-83.
66. Smibert, P., et al., *Global patterns of tissue-specific alternative polyadenylation in Drosophila.* Cell Rep, 2012. **1**(3): p. 277-89.
67. Jan, C.H., et al., *Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs.* Nature, 2011. **469**(7328): p. 97-101.
68. Legendre, M. and D. Gautheret, *Sequence determinants in human polyadenylation site selection.* BMC Genomics, 2003. **4**(1): p. 7.
69. Yao, C., et al., *Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation.* Proc Natl Acad Sci U S A, 2012. **109**(46): p. 18773-8.
70. Hardy, J.G. and C.J. Norbury, *Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation.* Biochem Soc Trans, 2016. **44**(4): p. 1051-7.
71. Yonaha, M. and N.J. Proudfoot, *Specific transcriptional pausing activates polyadenylation in a coupled in vitro system.* Mol Cell, 1999. **3**(5): p. 593-600.
72. Pinto, P.A., et al., *RNA polymerase II kinetics in polo polyadenylation signal selection.* Embo j, 2011. **30**(12): p. 2431-44.
73. Hilgers, V., S.B. Lemke, and M. Levine, *ELAV mediates 3' UTR extension in the Drosophila nervous system.* Genes Dev, 2012. **26**(20): p. 2259-64.
74. Gawande, B., et al., *Drosophila Sex-lethal protein mediates polyadenylation switching in the female germline.* Embo j, 2006. **25**(6): p. 1263-72.
75. Jenal, M., et al., *The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites.* Cell, 2012. **149**(3): p. 538-53.
76. Bava, F.A., et al., *CPEB1 coordinates alternative 3'-UTR formation with translational regulation.* Nature, 2013. **495**(7439): p. 121-5.
77. Katz, Y., et al., *Analysis and design of RNA sequencing experiments for identifying isoform regulation.* Nat Methods, 2010. **7**(12): p. 1009-15.

78. Zhang, H., J.Y. Lee, and B. Tian, *Biased alternative polyadenylation in human tissues.* Genome Biol, 2005. **6**(12): p. R100.
79. Sandberg, R., et al., *Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites.* Science, 2008. **320**(5883): p. 1643-7.
80. Mayr, C. and D.P. Bartel, *Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells.* Cell, 2009. **138**(4): p. 673-84.
81. Ranganathan, G., et al., *Tissue-specific expression of human lipoprotein lipase. Effect of the 3'-untranslated region on translation.* J Biol Chem, 1995. **270**(13): p. 7149-55.
82. Oikonomou, P., H. Goodarzi, and S. Tavazoie, *Systematic identification of regulatory elements in conserved 3' UTRs of human transcripts.* Cell Rep, 2014. **7**(1): p. 281-92.
83. Spies, N., C.B. Burge, and D.P. Bartel, *3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts.* Genome Res, 2013. **23**(12): p. 2078-90.
84. Shepard, P.J., et al., *Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq.* Rna, 2011. **17**(4): p. 761-72.
85. Miura, P., et al., *Alternative polyadenylation in the nervous system: to what lengths will 3' UTR extensions take us?* Bioessays, 2014. **36**(8): p. 766-77.
86. Miura, P., et al., *Widespread and extensive lengthening of 3' UTRs in the mammalian brain.* Genome Res, 2013. **23**(5): p. 812-25.
87. An, J.J., et al., *Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons.* Cell, 2008. **134**(1): p. 175-87.
88. Yudin, D., et al., *Localized regulation of axonal RanGTPase controls retrograde injury signaling in peripheral nerve.* Neuron, 2008. **59**(2): p. 241-52.
89. Koushika, S.P., M.J. Lisbin, and K. White, *ELAV, a Drosophila neuron-specific protein, mediates the generation of an alternatively spliced neural protein isoform.* Curr Biol, 1996. **6**(12): p. 1634-41.
90. Koushika, S.P., M. Soller, and K. White, *The neuron-enriched splicing pattern of Drosophila erect wing is dependent on the presence of ELAV protein.* Mol Cell Biol, 2000. **20**(5): p. 1836-45.
91. Robinow, S. and K. White, *Characterization and spatial distribution of the ELAV protein during Drosophila melanogaster development.* J Neurobiol, 1991. **22**(5): p. 443-61.
92. Levine, T.D., et al., *Hel-N1: an autoimmune RNA-binding protein with specificity for 3' uridylate-rich untranslated regions of growth factor mRNAs.* Mol Cell Biol, 1993. **13**(6): p. 3494-504.
93. Abe, R., et al., *Two different RNA binding activities for the AU-rich element and the poly(A) sequence of the mouse neuronal protein mHuC.* Nucleic Acids Res, 1996. **24**(24): p. 4895-901.

94. Lisbin, M.J., et al., *Function of RRM domains of Drosophila melanogaster ELAV: Rnp1 mutations and rrm domain replacements with ELAV family proteins and SXL.* Genetics, 2000. **155**(4): p. 1789-98.
95. Oktaba, K., et al., *ELAV links paused Pol II to alternative polyadenylation in the Drosophila nervous system.* Mol Cell, 2015. **57**(2): p. 341-8.
96. Peng, S.S., et al., *RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein.* Embo j, 1998. **17**(12): p. 3461-70.
97. Simionato, E., et al., *The Drosophila RNA-binding protein ELAV is required for commissural axon midline crossing via control of commissureless mRNA expression in neurons.* Dev Biol, 2007. **301**(1): p. 166-77.
98. Dickson, B.J., *Molecular mechanisms of axon guidance.* Science, 2002. **298**(5600): p. 1959-64.
99. Dent, E.W. and F.B. Gertler, *Cytoskeletal dynamics and transport in growth cone motility and axon guidance.* Neuron, 2003. **40**(2): p. 209-27.
100. Tessier-Lavigne, M. and C.S. Goodman, *The molecular biology of axon guidance.* Science, 1996. **274**(5290): p. 1123-33.
101. Cohen, J., et al., *The role of laminin and the laminin/fibronectin receptor complex in the outgrowth of retinal ganglion cell axons.* Dev Biol, 1987. **122**(2): p. 407-18.
102. Wolman, M.A., et al., *Transient axonal glycoprotein-1 (TAG-1) and laminin-alpha1 regulate dynamic growth cone behaviors and initial axon direction in vivo.* Neural Dev, 2008. **3**: p. 6.
103. Dodd, J., et al., *Spatial regulation of axonal glycoprotein expression on subsets of embryonic spinal neurons.* Neuron, 1988. **1**(2): p. 105-16.
104. Tessier-Lavigne, M., et al., *Chemotropic guidance of developing axons in the mammalian central nervous system.* Nature, 1988. **336**(6201): p. 775-8.
105. Serafini, T., et al., *The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC-6.* Cell, 1994. **78**(3): p. 409-24.
106. Kennedy, T.E., et al., *Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord.* Cell, 1994. **78**(3): p. 425-35.
107. Kidd, T., K.S. Bland, and C.S. Goodman, *Slit is the midline repellent for the robo receptor in Drosophila.* Cell, 1999. **96**(6): p. 785-94.
108. Raper, J.A., *Semaphorins and their receptors in vertebrates and invertebrates.* Curr Opin Neurobiol, 2000. **10**(1): p. 88-94.
109. Andrews, G.L., et al., *Dscam guides embryonic axons by Netrin-dependent and -independent functions.* Development, 2008. **135**(23): p. 3839-48.
110. Alavi, M., et al., *Dscam1 Forms a Complex with Robo1 and the N-Terminal Fragment of Slit to Promote the Growth of Longitudinal Axons.* PLoS Biol, 2016. **14**(9): p. e1002560.
111. Zipursky, S.L., W.M. Wojtowicz, and D. Hattori, *Got diversity? Wiring the fly brain with Dscam.* Trends Biochem Sci, 2006. **31**(10): p. 581-8.
112. Yamakawa, K., et al., *DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system.* Hum Mol Genet, 1998. **7**(2): p. 227-37.

113. Bolisetty, M.T., G. Rajadinakaran, and B.R. Graveley, *Determining exon connectivity in complex mRNAs by nanopore sequencing.* Genome Biol, 2015. **16**: p. 204.
114. Yu, H.H., et al., *Endodomain diversity in the Drosophila Dscam and its roles in neuronal morphogenesis.* J Neurosci, 2009. **29**(6): p. 1904-14.
115. Ozsolak, F. and P.M. Milos, *RNA sequencing: advances, challenges and opportunities.* Nat Rev Genet, 2011. **12**(2): p. 87-98.
116. Schmucker, D., et al., *Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity.* Cell, 2000. **101**(6): p. 671-84.
117. Schmucker, D. and B. Chen, *Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes.* Genes Dev, 2009. **23**(2): p. 147-56.
118. Graveley, B.R., *Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures.* Cell, 2005. **123**(1): p. 65-73.
119. Yue, Y., et al., *Long-range RNA pairings contribute to mutually exclusive splicing.* Rna, 2016. **22**(1): p. 96-110.
120. Kreahling, J.M. and B.R. Graveley, *The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the Drosophila Dscam pre-mRNA.* Mol Cell Biol, 2005. **25**(23): p. 10251-60.
121. Neves, G., et al., *Stochastic yet biased expression of multiple Dscam splice variants by individual cells.* Nat Genet, 2004. **36**(3): p. 240-6.
122. Miura, S.K., et al., *Probabilistic splicing of Dscam1 establishes identity at the level of single neurons.* Cell, 2013. **155**(5): p. 1166-77.
123. Hughes, M.E., et al., *Homophilic Dscam interactions control complex dendrite morphogenesis.* Neuron, 2007. **54**(3): p. 417-27.
124. Millard, S.S. and S.L. Zipursky, *Dscam-mediated repulsion controls tiling and self-avoidance.* Curr Opin Neurobiol, 2008. **18**(1): p. 84-9.
125. Maynard, K.R. and E. Stein, *DSCAM contributes to dendrite arborization and spine formation in the developing cerebral cortex.* J Neurosci, 2012. **32**(47): p. 16637-50.
126. Wang, J., et al., *Drosophila Dscam is required for divergent segregation of sister branches and suppresses ectopic bifurcation of axons.* Neuron, 2002. **33**(4): p. 559-71.
127. Heisenberg, M., *What do the mushroom bodies do for the insect brain? an introduction.* Learn Mem, 1998. **5**(1-2): p. 1-10.
128. Hattori, D., et al., *Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms.* Nature, 2009. **461**(7264): p. 644-8.
129. Sawaya, M.R., et al., *A double S shape provides the structural basis for the extraordinary binding specificity of Dscam isoforms.* Cell, 2008. **134**(6): p. 1007-18.
130. Ly, A., et al., *DSCAM is a netrin receptor that collaborates with DCC in mediating turning responses to netrin-1.* Cell, 2008. **133**(7): p. 1241-54.
131. Kim, J.H., et al., *Dscam expression levels determine presynaptic arbor sizes in Drosophila sensory neurons.* Neuron, 2013. **78**(5): p. 827-38.

132. Chang, H.H.Y., et al., *Non-homologous DNA end joining and alternative pathways to double-strand break repair.* Nat Rev Mol Cell Biol, 2017.
133. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.
134. Pardo, B., B. Gomez-Gonzalez, and A. Aguilera, *DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship.* Cell Mol Life Sci, 2009. **66**(6): p. 1039-56.
135. Li, K., et al., *Optimization of genome engineering approaches with the CRISPR/Cas9 system.* PLoS One, 2014. **9**(8): p. e105779.
136. Klug, A., *The discovery of zinc fingers and their applications in gene regulation and genome manipulation.* Annu Rev Biochem, 2010. **79**: p. 213-31.
137. Joung, J.K. and J.D. Sander, *TALENs: a widely applicable technology for targeted genome editing.* Nat Rev Mol Cell Biol, 2013. **14**(1): p. 49-55.
138. Moscou, M.J. and A.J. Bogdanove, *A simple cipher governs DNA recognition by TAL effectors.* Science, 2009. **326**(5959): p. 1501.
139. Boch, J. and U. Bonas, *Xanthomonas AvrBs3 family-type III effectors: discovery and function.* Annu Rev Phytopathol, 2010. **48**: p. 419-36.
140. Gaj, T., C.A. Gersbach, and C.F. Barbas, 3rd, *ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering.* Trends Biotechnol, 2013. **31**(7): p. 397-405.
141. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna, *RNA-guided genetic silencing systems in bacteria and archaea.* Nature, 2012. **482**(7385): p. 331-8.
142. Wang, H., M. La Russa, and L.S. Qi, *CRISPR/Cas9 in Genome Editing and Beyond.* Annu Rev Biochem, 2016. **85**: p. 227-64.
143. Ma, D. and F. Liu, *Genome Editing and Its Applications in Model Organisms.* Genomics Proteomics Bioinformatics, 2015. **13**(6): p. 336-44.
144. Gargano, J.W., et al., *Rapid iterative negative geotaxis (RING): a new method for assessing age-related locomotor decline in Drosophila.* Exp Gerontol, 2005. **40**(5): p. 386-95.
145. Soller, M. and K. White, *ELAV inhibits 3'-end processing to promote neural splicing of ewg pre-mRNA.* Genes Dev, 2003. **17**(20): p. 2526-38.
146. Soller, M. and K. White, *ELAV multimerizes on conserved AU4-6 motifs important for ewg splicing regulation.* Mol Cell Biol, 2005. **25**(17): p. 7580-91.
147. Hattori, D., et al., *Dscam diversity is essential for neuronal wiring and self-recognition.* Nature, 2007. **449**(7159): p. 223-7.
148. Samson, M.L., *Rapid functional diversification in the structurally conserved ELAV family of neuronal RNA binding proteins.* BMC Genomics, 2008. **9**: p. 392.
149. Alon, S., et al., *The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing.* Elife, 2015. **4**.
150. Branton, D., et al., *The potential and challenges of nanopore sequencing.* Nat Biotechnol, 2008. **26**(10): p. 1146-53.