Mechanisms of MiRNA-based Gene Regulation in *C. elegans* and Human Cells

by

Kasuen Indrajith Bandara Kotagama

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2019 by the
Graduate Supervisory Committee:

Marco Mangone, Chair
Joshua LaBaer
Jason Newbern
Alan Rawls

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Multicellular organisms use precise gene regulation, executed throughout

development, to build and sustain various cell and tissue types. Post-transcriptional gene

regulation is essential for metazoan development and acts on mRNA to determine its

localization, stability, and translation. MicroRNAs (miRNAs) and RNA binding proteins

(RBPs) are the principal effectors of post-transcriptional gene regulation and act by

targeting the 3'untranslated regions (3'UTRs) of mRNA. MiRNAs are small non-coding

RNAs that have the potential to regulate hundreds to thousands of genes and are

dysregulated in many prevalent human diseases such as diabetes, Alzheimer's disease,

Duchenne muscular dystrophy, and cancer. However, the precise contribution of

miRNAs to the pathology of these diseases is not known.

MiRNA-based gene regulation occurs in a tissue-specific manner and is

implemented by an interplay of poorly understood and complex mechanisms, which

control both the presence of the miRNAs and their targets. As a consequence, the precise

contributions of miRNAs to gene regulation are not well known. The research presented

in this thesis systematically explores the targets and effects of miRNA-based gene

regulation in cell lines and tissues.

I hypothesize that miRNAs have distinct tissue-specific roles that contribute to the

gene expression differences seen across tissues. To address this hypothesis and expand

our understanding of miRNA-based gene regulation, 1) I developed the human

3'UTRome v1, a resource for studying post-transcriptional gene regulation. Using this

resource, I explored the targets of two cancer-associated miRNAs *miR-221* and *let-7c*. I

identified novel targets of both these miRNAs, which present potential mechanisms by which they contribute to cancer. 2) Identified *in vivo*, tissue-specific targets in the intestine and body muscle of the model organism *Caenorhabditis elegans*. The results from this study revealed that miRNAs regulate tissue homeostasis, and that alternative polyadenylation and miRNA expression patterns modulate miRNA targeting at the tissue-specific level. 3) Explored the functional relevance of miRNA targeting to tissue-specific gene expression, where I found that miRNAs contribute to the biogenesis of mRNAs, through alternative splicing, by regulating tissue-specific expression of splicing factors. These results expand our understanding of the mechanisms that guide miRNA targeting and its effects on tissue-specific gene expression.

DEDICATION

To Dishna and Hemersiri Kotagama. Ammi and appuchi, all of this is made possible because of the sacrifices you've made to give me everything that I have.

To my brother Praveen. You've always been by my side supporting me. You are my tether to the real world, and a reason to make it a better place.

To Marco Mangone, and to Justin Wolter. When I write for science it is always with the two of you in mind.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

xv

CHAPTER 1

INTRODUCTION

RNA - not just a messenger

The genome of an organism contains all the information necessary to build, sustain, and replicate that organism. Understanding how this information produces complex life forms is a fundamental goal of genetics and genomics. The central dogma of molecular biology described by Francis Crick in 1956, explains the transfer of genetic information, from its storage as a nucleic acid code in DNA, through RNA which acts as a messenger, into proteins which carry out cellular functions [1]. The human body is estimated to consist of 37 trillion cells divided into various complex cell and tissue types [2], and based on the central dogma, it was expected that the human genome would contain 30k-100k protein-coding genes [3, 4]. Surprisingly, sequencing the human genome revealed that there are only 19-25k protein-coding genes [3-6] and that only 5% of the genome codes for proteins [7]. This number is surprisingly low given that the widely studied simple nematode *Caenorhabditis elegans* (*C. elegans*), with only 959 cells and 10 tissues, has 20k protein-coding genes [8]. Additionally, 40% of the protein-coding genes in both organisms perform similar functions [7]. These findings suggest that focusing on the proteome alone may be insufficient to understand the differences in complexity between organisms.

RNA, initially described in the central dogma as an intermediary messenger molecule for genetic information, is involved in a variety of cellular functions including

enzymatic catalysis, and gene regulation (Figure 1.1). The ability of RNA to perform all

these functions provides evidence for the RNA world hypothesis, the idea that at one

point in evolution RNA existed as the sole molecule required for life [9]. This hypothesis

suggests that a central role for RNA in cell biology pre-dates much of the organismal

complexity seen today (Figure 1.1). As such, it is not surprising that RNA-based

mechanisms contribute to tissue differentiation and have expanded throughout evolution

[10-12]. In metazoans, RNA contributes to gene regulation at the epigenetic [13],

transcriptional [14], post-transcriptional [15] and translational stages [16], and is essential

for development (Figure 1.1).

Post-transcriptional regulation is a hypernym used to indicate a series of

mechanisms that act on RNA and produces a disparity between the transcriptomes and

proteomes of organisms [17-20]. While post-transcriptional regulation acts on almost all
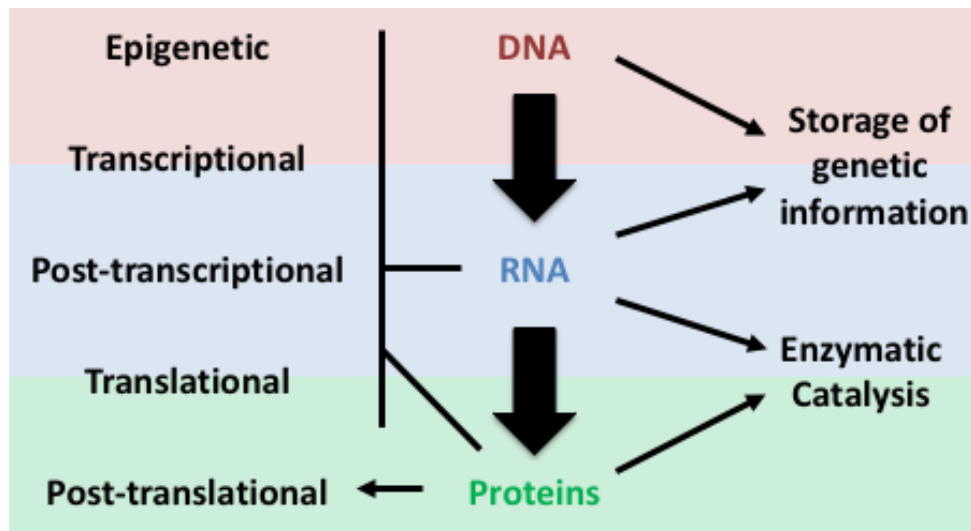
Figure 1.1 - RNA has a central role in cell biology. RNA is a highly active molecule,
that can store genetic information similar to DNA, and also perform enzymatic catalysis
similar to proteins. RNA contributes to gene regulation throughout expression, from
epigenetic control of DNA states to the translation of proteins.

2

RNA, it is perhaps best characterized for its effects on mRNA, where it can impact cellular localization [21], translation and degradation [22] (Figure 1.1). Many prevalent human diseases such as cancer, Alzheimer's disease, Duchene muscular dystrophy, and diabetes have disruptions in post-transcriptional gene regulation. However, the causality of these observations and the precise roles of post-transcriptional regulation in these conditions is not fully understood. Understanding the regulation of RNA in cells can reveal mechanisms that produce and maintain tissues and potentially reveal avenues of therapeutic intervention for human diseases. The research presented in this thesis explores some of these RNA-based regulatory networks, with a particular focus on post-transcriptional regulation and how it contributes to tissue-specific gene expression.

3'untranslated regions mediate post-transcriptional regulation.

The 3'untranslated regions (3'UTRs) of genes, unfortunately, named for what they do not do, are sequences that are found between the STOP codons and poly-A tails of transcripts and interact with many RNA binding proteins (RBPs) and non-coding RNAs to regulate mRNA post-transcriptionally [23] (Figure 1.2).

While 3'UTRs are generally less conserved than coding sequences [24], they contain conserved regions that are recognized by regulatory elements and determine localization [21], translation [25] or degradation [26, 27] of the transcript (Figure 1.2). The first reports of conserved elements in 3'UTRs arose from homology studies using

3

actin genes. Isoforms of actin expressed in similar tissues, across related vertebrate species, shared conserved 3'UTRs, while actin isoforms expressed in different tissues did not, implying a tissue-specific role for 3'UTRs [28, 29]. 3'UTRs are essential for embryonic development where they contribute to the maternal to zygotic transition in many model organisms [30-33], and then continue to produce tissue-specific gene regulation throughout development [34-36]. In *C. elegans* the 3'UTR can determine gene expression even over the promoters of genes, exemplifying the role of 3'UTRs even beyond embryogenesis [37].

The 3'ends of 3'UTRs are formed by the recognition of the polyadenylation sequence (PAS), a hexameric sequence that is canonically AAUAAA, by the cleavage and polyadenylation complex, which directs cleavage of the transcript ~21nt downstream of the PAS (Figure 1.2). The importance of the PAS was first identified in thalassemia [38], a condition characterized by decreased hemoglobin production, where a single point
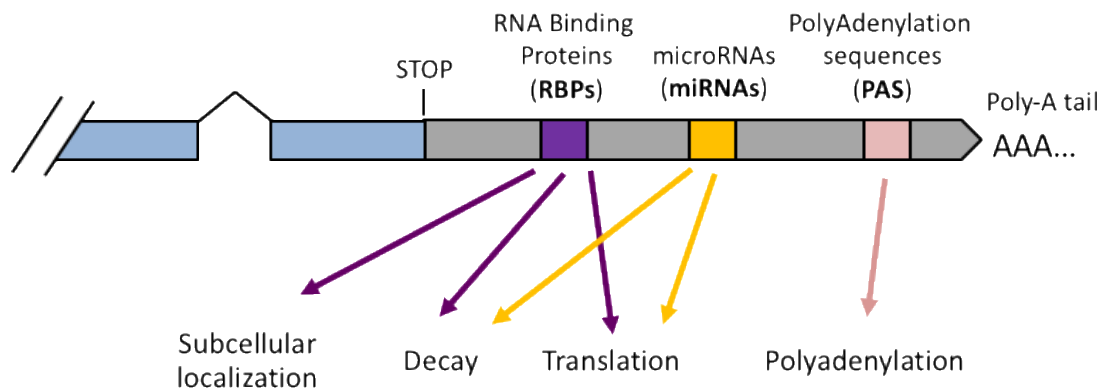
Figure 1.2 – The 3' untranslated regions (3'UTRs) of genes contain many regulatory elements that can direct mRNA fate. The 3'UTR of a transcript is the region between the STOP codon and poly-A tail and has binding sites of RNA binding proteins, miRNA and other regulatory factors.

mutation in the PAS of the α2-globin drastically reduced the mRNA levels of this gene in erythrocytes [38].

To study the role of 3'UTRs in diseases and throughout development, many studies have sequenced and annotated the 3'UTRs of various model organisms and found that a majority of 3'UTRs in metazoans are polyadenylated at multiple sites [39-42]. The mechanism by which the cleavage and polyadenylation machinery distinguishes between different PAS elements, to produce a single coding sequence with different 3'UTR isoforms, is known as alternative polyadenylation (APA) (Figure 1.3). Due to APA, a single coding sequence is expressed with different 3'UTR isoforms, which alters the post-transcriptional regulation of the gene (Figure 1.3). Accordingly, changes in 3'UTR length due to APA occur during cell differentiation [36, 40, 43-46], proliferation [47], and diseased states such as cardiomyopathy [48-50] and cancer [51].
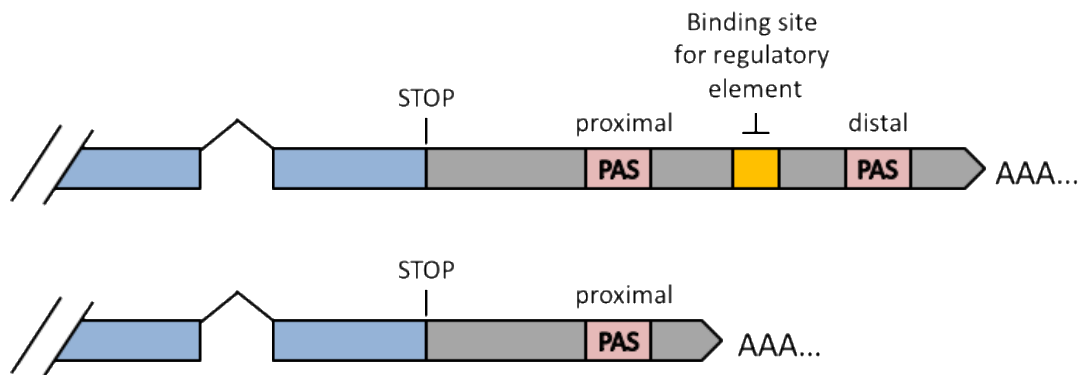
Figure 1.3 - APA can determine the availability of binding sites for regulatory elements. Alternative polyadenylation (APA) is a mechanism by which the cleavage and polyadenylation machinery distinguishes between PAS sites to produce a single coding sequence with different 3'UTR isoforms. The PAS site used determines the 3'UTR and regulatory sequences effecting the upstream coding sequence.

The mechanisms behind cleavage and polyadenylation have been studied in depth, and the core protein complex, capable of cleavage and polyadenylation *in vitro*, consists of approximately 20 proteins [52, 53]. In addition to the PAS, the cleavage and polyadenylation complex recognizes sequences both up and downstream of it, and the presence of these other sequences are believed to strengthen the usage of particular PAS elements [54]. The *in vivo*, active polyadenylation complex is estimated to contain upwards of 80 proteins [52] and, outside of the core complex, the roles of many of these proteins in cleavage and polyadenylation remains unknown. The polyadenylation complex is well conserved across eukaryotes [55] and the size of the complex, when compared to its relatively simple task in mRNA processing is perhaps an indication of the importance of regulating sites of polyadenylation in organisms.

A recent study of nine tissue-specific transcriptomes in *C. elegans* revealed that most genes commonly expressed between tissues have different 3'UTR isoforms due to APA, and changes in 3'UTR length allow transcripts to be differentially expressed in a tissue-specific manner [56, 57]. There is evolutionary conservation of the tissue-specific APA patterns in genes with more than one UTR isoform, implying that APA may be functional [18]. These studies provide evidence that at least in specific instances, 3'UTRs are modulated by APA, and interact with small non-coding RNAs and RNA binding proteins (RBPs) in ways that are important for gene regulation.

Discovery of miRNAs - essential regulators of metazoan development

MicroRNAs(miRNAs) are small non-coding RNAs that are ~22nt in length and play an indispensable role in the development of multicellular organisms [15]. Contributions of the miRNA pathway to development were apparent from its initial, discovery in *C. elegans* where the efforts of Victor Ambrose and Gary Ruvkun in 1993, showed that the interactions between the non-coding gene *lin-4* and the protein-coding gene *lin-14* were essential for the developmental progression of the worm through its larval stages [58, 59] (Figure 1.4). They found that *lin-4* has significant sequence complementarity to the 3'UTR of *lin-14* which led to an interaction that decreased the translation of *lin-14* [58, 59]. The decrease of *lin-14* protein levels caused by *lin-4* directs the developmental transition from the L1 to the L2 larval stage [58, 59] (Figure 1.4).

A second discovery in the miRNA pathway was made in 1998 by Bohmert et al., who found that a mutation in a specific family of proteins resulted in developmental defects to tissue differentiation in the plant *Arabidopsis thaliana*. They named this family of proteins Argonaute due to similarities in the phenotypes of the mutants to the octopus *Argonauta Argo* (Figure 1.5). Further studies of the Argonaute proteins found that they are the effectors in small RNA pathways, and deficiencies in the miRNA pathway likely caused the mutant phenotypes with defects to tissue differentiation. The mammalian Argonaute protein AGO-2 is predominantly responsible for the miRNA pathway, and studies in mice have shown that mutations leading to disruptions in this protein are embryonic lethal, demonstrating its importance in early development.

7
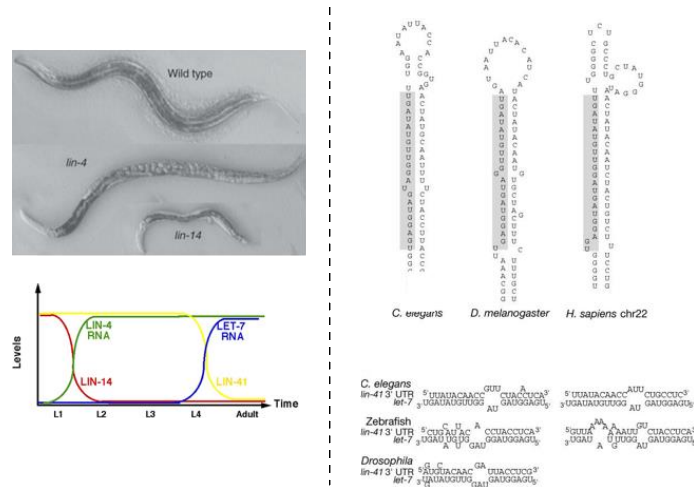
Figure 1.4 – The initial discovery of miRNAs revealed that they play a role in regulating developmental timing in *C. elegans*. Left panel – Wild type N2 worms compared to heterochronic mutants showing the *lin-4* and *lin-14* phenotypes. The *lin-4* worms develop slowly and do not produce all of the adult tissues while the *lin-14* mutants develop too rapidly and mature before reaching their full size. The graph below illustrates the changes, across developmental stages, in the levels of the miRNAs *lin-4* and *let-7*, and the corresponding changes in the protein levels of their targets LIN-14 and LIN-41. Right panel – The *let-7* miRNA is conserved in *C. elegans*, *D. melanogaster* and humans. The binding site for the miRNA *let-7* in the protein-coding gene *lin-41* gene is conserved between *C. elegans*, *Drosophila*, and zebrafish. Figures adapted from: Left panel – Wormbook, *C. elegans* microRNAs. Vella, M. C, and Slack, F. J. 2005 [60]. Right panel – Pasquinelli, A.E. et al. 2000 [61].

Wild type
Arabidopsis

Argonaute
mutant
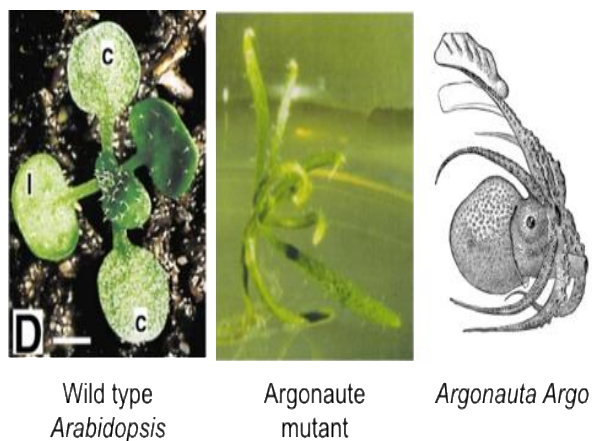
Argonauta Argo

Figure 1.5 – Mutations in the proteins that carry out the miRNA pathway were discovered to display developmental defects in *Arabidopsis thaliana.* When compared to the wild type the Argonaute mutants were unable to differentiate its tissues to develop stems and leaves. Due to similarities of the phenotype to *Argonauta Argo,* the protein family was named Argonaute. Figure adapted from Bohmert et al.

MiRNAs were believed to be an oddity of gene regulation in the worm until another study performed at the Ruvkun lab revealed that a different miRNA, *let-7* and its target in the 3'UTR of *lin-41* is conserved across species (Figure 1.4). Following this discovery, the field of miRNA research gathered interest and over 2000 miRNAs have been annotated in the human genome [63], and miRNAs have become well established in post-transcriptional gene regulation.

MicroRNA biogenesis and mechanisms of action

A significant effort has been made to study the biogenesis of miRNAs and how they carry out their cellular functions in metazoans (Figure 1.6) [15, 64, 65]. These studies have revealed that miRNAs are transcribed from the genome by RNA polymerase II to produce a primary transcript known as the pri-miRNA [66], which contains a complementary stretch of nucleotides that fold it into a hairpin structure comprising a double-stranded region at the miRNA sequence (Figure 1.6). The hairpin structure is recognized by enzymes which cleave the pri-miRNA and release the pre-miRNA. Pre-miRNAs are transported out of the nucleus by the nuclear membrane RNA transporter exoprotein-5, and once in the nucleus, the double-stranded region of the pre-miRNA is recognized by the enzyme Dicer which cleaves the hairpin sequence to produce a double-stranded RNA that is ~21nt in length (Figure 1.6). The double-stranded RNA, composed of the miRNA and miRNA* is then loaded into an Argonaute family protein. The Argonaute protein is the catalytic component of a large protein complex known as the RNA induced silencing complex (RISC), which executes the miRNA pathway (Figure

9

1.6). The RISC discriminates between the two miRNA strands based on thermodynamic properties of the strands [67], then degrades one of them and uses the other as a guide to identify target sequences in the 3'UTRs of mRNAs.
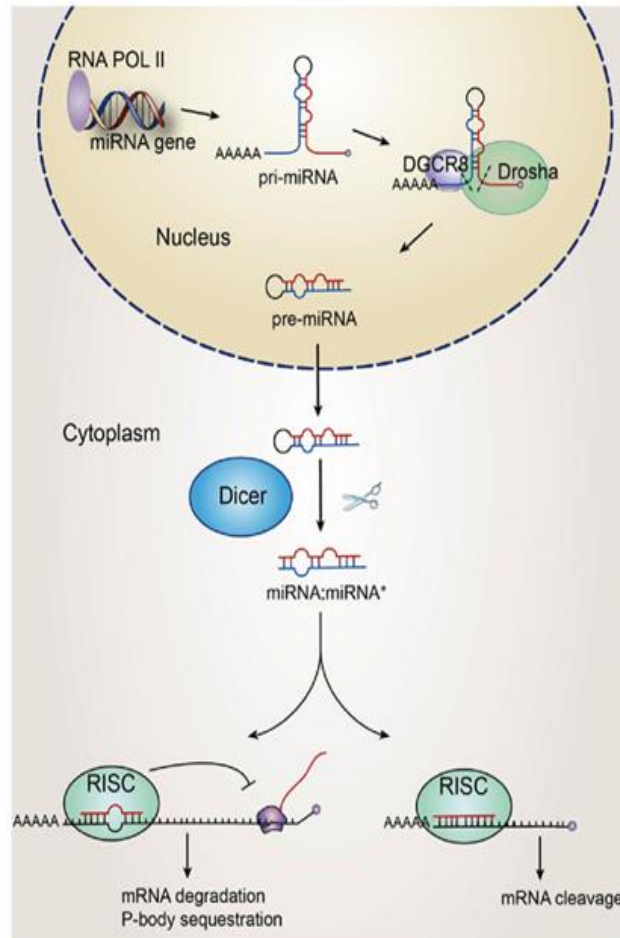
Figure 1.6 - The biogenesis and functions of miRNAs in metazoans. MiRNAs are transcribed from the genome by RNA polymerase II to produce pri-miRNAs. Due to sequence complementarity within the transcript, they fold to form a characteristic hairpin structure that is recognized by DGCR8 and Drosha which cleaves just below the stems of the hairpins to produce the pre-miRNAs. Once the pre-miRNAs are transported out of the nucleus, the enzyme Dicer recognizes their structure due to its double-stranded nature and cleaves the hairpin to release a ~22nt miRNA:miRNA* duplex which then binds to the RNA induced silencing complex (RISC). The RISC degrades one of the miRNA strands and uses the other as a guide to identify miRNA targets. Targets of miRNAs are held in translational repression or cleaved and degraded. Figure adapted from Barca-Mayo and Lu, 2012 [69]

The RISC is guided to target sites by the miRNA based on sequence complementarity and Watson and Crick base-pairing (Figure 1.7). Initial efforts looking at requirements for miRNA targeting indicated that sequence complementarity at a region of the miRNA, nucleotides 2-7, referred to as the "seed" region was essential for target recognition. The discovery of the crystal structure of an Argonaute protein complex with a miRNA clarified the importance of the seed for miRNA target recognition. The predominant Argonaute protein involved with the miRNA pathway in humans, AGO-2 was crystallized bound to a guide miRNA [68] (Figure 1.8) and revealed that the seed region of the miRNA extends into a cleft of the protein which allows probe and form base pairs with the 3'UTRs of mRNAs (Figure 1.8). The initial base pairing of the seed region of the miRNA with the target 3'UTR is believed to facilitate further interactions between the 3'UTR and the 3'end of the miRNA (Figure 1.8).
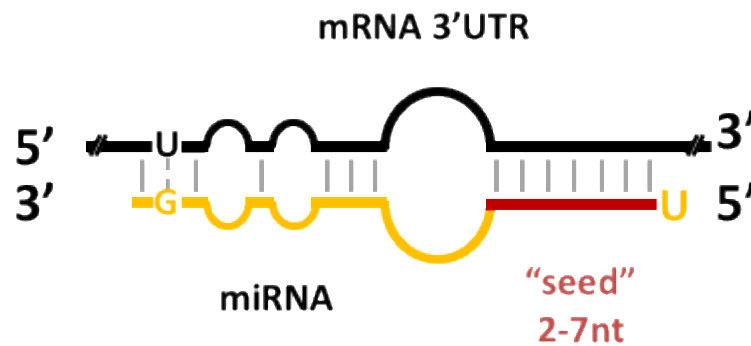
Figure 1.7 – A model representing a miRNA interacting with its target 3'UTR.
Base pairing between the miRNA and the 3'UTR facilitates target identification. Strict complementarity at the "seed" region, nucleotides 2-7 at the 5' end of the miRNA, is the best indicator of miRNA targeting. Mismatches are tolerated downstream of the seed and compensatory base pairing at the 3'end can strengthen site recognition.

The result of a miRNA targeting a 3'UTR is decreased protein expression. However, the mechanism that produces this effect can vary between organisms [70-72]. In plants, targeting by miRNAs requires perfect base pairing and leads to the degradation of the target [70]. While in metazoans miRNA targeting can result in the mRNA being held in translational repression or to degradation [71].
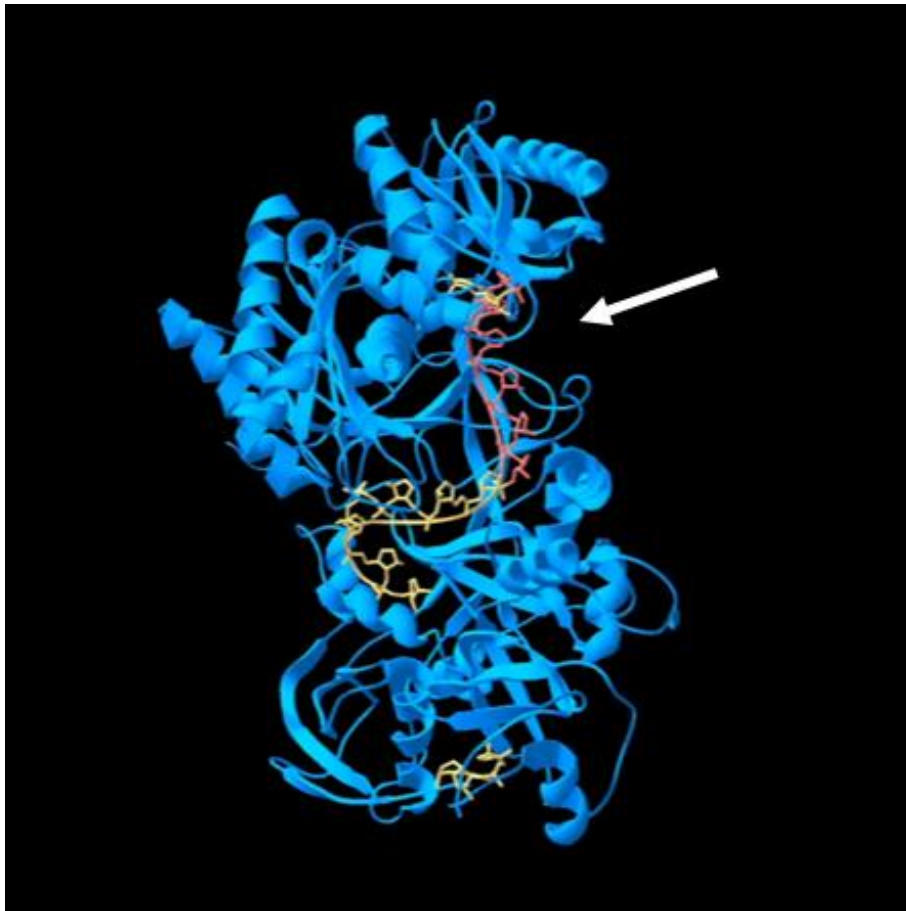
Figure 1.8 – The crystal structure of human AGO-2 in complex with a guide RNA. The Argonaute protein AGO-2 is shown in blue; and the miRNA is shown in red and yellow. The crystallography data potentially reveal the structural basis for miRNA target recognition. The white arrow indicates a cleft in the protein structure to which the bases of the seed region (red) of the guide miRNA extend. This allows the seed region to form initial base pairing with target 3'UTRs and is important for target recognition. Crystallography data was adapted from Schirle, N.T. et al. 2012 [68]

There are two proposed mechanisms of action by which miRNAs carry out their functions. Experiments done with cell lines show that miRNAs interfere with the formation of the translation initiation complex at the 5'cap [73]. This mechanism is further evidenced by *in vitro* studies that show miRNA-based repression does not act on IRES-dependent translation [74]. Evidence that the RISC associates with the CCR4-NOT deadenylation complex [89] suggest that miRNA targeting leads to deadenylation of the poly-A tail of mRNAs[75]. The effects of miRNA based deadenylation of targets have been seen in zebrafish [76], and *C. elegans* [77] and ultimately leads to the degradation of the target mRNA. MiRNA targeting can also lead to re-localization of mRNA to P-bodies, though if this is for degradation of the mRNA, or for it to be held translationally repressed to then later be released is unknown [78]. Perhaps these mechanisms are staggered, and that miRNA targets are held translationally silenced before degradation, though this implies that miRNA targeting is an irreversible process [79].

In animals, knocking out the miRNA pathway leads to lethality during embryonic development [80-82] with defects in cell differentiation [80, 83, 84], indicating an essential role for the miRNA pathway in producing tissues. Efforts to profile the miRNAs expressed in a tissue-specific manner have revealed that miRNAs can be tissue-restricted, even in fully developed tissues [85, 86], indicating that perhaps miRNAs may have tissue-specific roles beyond embryogenesis. Despite the withstanding questions about the precise mechanisms by which miRNAs act, their impact on gene regulation and its consequences for development is established and highlights the need to identify miRNA targets.

Methods for identifying microRNA targets

Due to the seed region being just six nucleotides, miRNAs are believed to be able to target many hundreds to even thousands of genes [87]. The most widely used method for identifying miRNA targets is through bioinformatic means using predictive software[88]. Initial predictive software identified miRNA target sites by locating regions of seed complementarity in genomes and requiring these sites to be evolutionarily conserved [87, 89, 90]. While complementarity of the seed element remains the best identifier for miRNA binding sites, many recent studies have revealed that miRNA targeting can occur even in the presence of mismatches in the seed, suggesting that the seed region is not a sufficient predictor of miRNA targets [11, 91-94]. An analysis performed by Wolter et al. 2014 [94] on three widely used miRNA target prediction software, TargetScan [95], PicTar [96], and DIANA-microT [97], revealed that each of the software failed to identify ~30% of the experimentally validated miRNA targets. The most likely cause for this discrepancy is the application of conservation filters that prevent the identification of species-specific interactions.

More recent predicative software allow for mismatches at the seed and utilize thermodynamic properties [99] or a combinatorial approach [100] to identify miRNA binding sites and ignore conservation requirements. The consequence of removing the conservation filter is that these algorithms predict far more targets for each miRNA (over 7000 for some *let-7* family members [101]), although the false positive rates are yet to be established for these predictions. While bioinformatic predictions can be a useful starting

point for studying miRNAs, in its current state identifying functional miRNA targets requires validation by wet bench approaches.

Wet bench efforts to identify miRNA targets can be divided into, biochemical approaches, high-throughput sequencing-based approaches, and functional reporter assays [102]. Biochemical methods such as PCRs, Western and Northern blotting have been used since the very discovery of miRNAs to test miRNA-gene interactions and its consequences. While these methods are exceptionally well established, and accurate, they are not easily applied to test miRNA targeting in high-throughput.

Many of the high-throughput approaches utilize next-generation sequencing along with cross-linking and immunoprecipitation (CLIP) based methods to identify miRNA targets from cell lines, tissue samples and in the case of *C. elegans* the entire organism [103-106]. While these techniques can detect a large number of miRNA targets at the genomic scale, there are two stipulations to these techniques, 1) it is difficult to distinguish transient and non-functional interactions from functional and biologically relevant interactions 2) it does not identify the miRNA responsible for targeting and is reliant on predictive software to do so.

Another sequencing-based approach is to overexpress particular miRNAs in cell lines using DNA constructs and perform transcriptome or proteome sequencing. This approach avoids the pitfall of identifying functional targets, however, is limited in 1) the capacity to study *in vivo* biology as it overexpresses a miRNA which changes the system being investigated 2) its inability to distinguish indirect interactions, where

15

downregulation of particular genes by a miRNA can then have secondary downstream effects on the transcriptome or proteome.

Reporter-based approaches overcome some of the limitations of sequencing and biochemical techniques as they can measure *in vivo,* functional interactions. Reporter systems make use of easily quantifiable proteins by attaching them upstream of query 3'UTRs then measuring changes to protein production in the presence of a test miRNA. One of the most widely used reporter systems to identify miRNA targets are luciferase assays [107], which utilize the ability of the luciferase enzymes to produce accurately quantifiable levels of luminescence in the presence of their respective substrates. Generally, the luciferase assays have been used to screen a small number of individual miRNA-target interactions, but recently Wolter et al. 2014, adapted it to the high-throughput 96-well format [108]. One of the caveats to applying this approach at a genomic scale is that it needs a large number of easily manipulated, cloned 3'UTRs.

Most studies that identify miRNA targets use a combination of the approaches described above, where bioinformatic predictions and high-throughput sequencing efforts producing genomic miRNA-target databases, validated with reporter-based experiments and biochemical means.

Dysregulation of microRNAs in cancer

Given the expansive roles of miRNAs in regulating gene expression, it is perhaps unsurprising that dysregulation of miRNAs is seen in all the phenotypes associated with cancer[109] (Figure 1.9).

The first reported deregulation of miRNAs in cancer was in chronic lymphocytic leukemia, which indicated that the miRNAs *miR-15* and *miR-16* are in a region of chromosome 13 deleted in more than half of B-cell chronic lymphocytic leukemia cases [110]. These two miRNAs target oncogenes such as BCL2 [111], CCND1 and WNT3a [112] and the deletion of both miRNAs allow these oncogenes to be overexpressed [113].

A second tumor suppressing miRNA *let-7* was found to be broadly involved with many cancers and is deleted in lung, breast, urothelial, ovarian, and cervical cancers
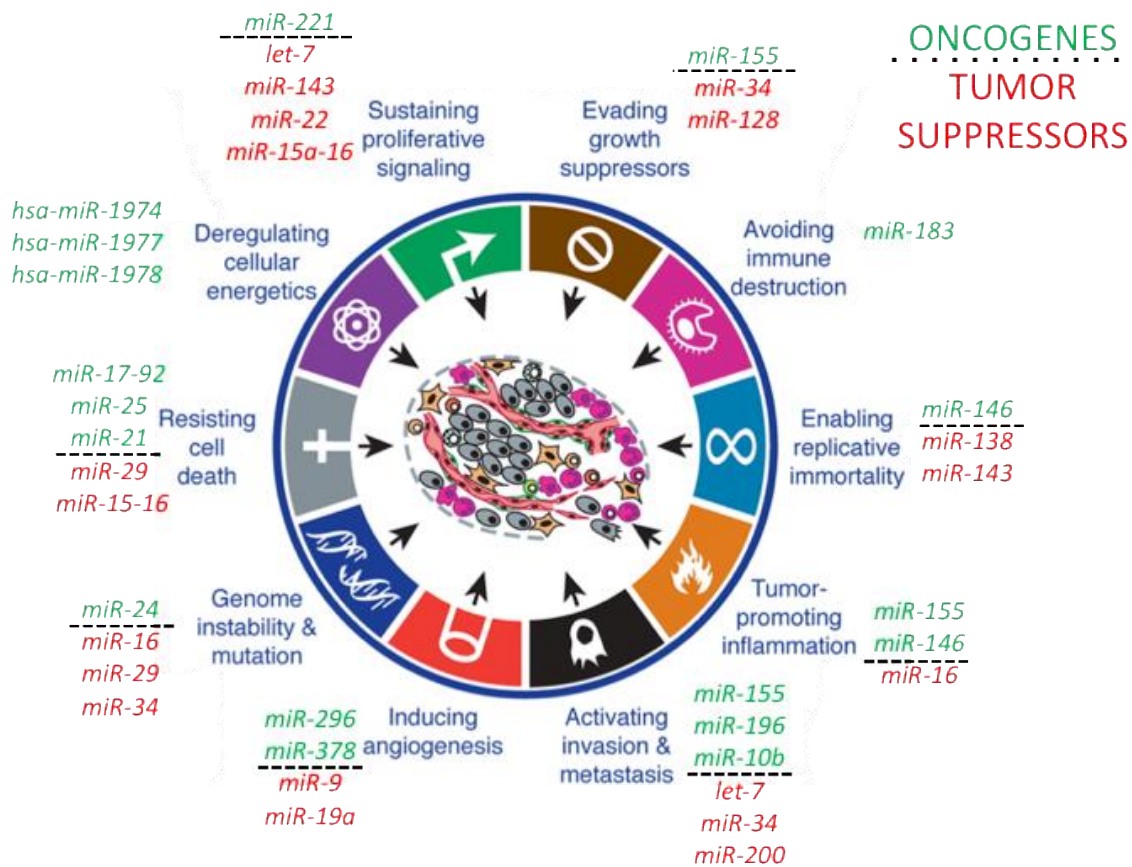
Figure 1.9 − MiRNAs are implicated in all the hallmarks of cancer. Based on the targets of a miRNA, it can act as either an oncogene (green) or a tumor suppressor (red). The precise contributions of most of these miRNAs to the phenotypes are unknown Figure adapted from Hanahan, D. & Weinberg, R.A. 2011 [114]

[115]. Some of the oncogenes identified as a target of *let-7* are in the RAS family of proteins. This discovery, made using *C. elegans,* showed that *let-7* targets the RAS homolog *let-60* [116]. Human have three RAS proteins, H-ras, K-ras, and N-ras, all of which contain binding sites for *let-7* in their 3'UTRs [116]; mutations in these binding sites are associated with increased lung and breast cancer [117, 118]. Bioinformatic software predicts that *let-7* can target over 1000 genes, and many of these predictions have not yet been validated [90]. A recent high-throughput study using a dual-luciferase reporter-based approach found that *let-7* targets several genes involved with cell cycle progression, DNA synthesis, and mitosis [11].

The previously described miRNAs act in a tumor-suppressing capacity, however, miRNAs can also act as oncogenes and are referred to as oncomiRs[109, 119]. For example, the miRNA *miR-10b* is a well-characterized oncomiR found overexpressed in over 18 cancer types and is correlated with increased invasion and metastasis, evading apoptosis, and proliferation [120]. Importantly upregulation of *miR-10b* is associated with poor prognosis and clinical progression of cancer [121]. *miR-10b* can target the retinoic acid pathway at multiple points and is a potential mechanism by which *miR-10b* regulates cell differentiation [94].

A less well-characterized miRNA, *miR-221* is experimentally validated to target tumor-suppressing genes such as ARHI, CDKN1C, Kip-1 (p-27), PTEN, and PUMA, but also the oncogenes c-Kit, MMP1, and SOD1 [122]. The overall role of *miR-221* seem to be oncogenic as the over-expression of this miRNA can be seen in nine different types of cancers, and its expression level is correlated with poor patient prognosis. Further

identifying targets for this miRNA can help with understanding the mechanisms leading to poor prognosis, and perhaps reveal avenues of therapeutic intervention [123].

Many studies have profiled miRNAs across cancer types and found them to be widely dysregulated [124, 125], however, the mechanisms which lead to this dysregulation of miRNAs are not clear. Several profiling studies have suggested the possibility of using miRNAs as biomarkers for detecting the onset, progression and treatment response of cancer [126-128]. In the case of chronic myelogenous leukemia, the expression of miRNAs *miR-10a*, *miR-17-92*, *miR-29a*, *miR-150* and *miR-203* have all been shown to change across disease progression [129]. Whether changes to miRNAs can act as drivers of cancer or are passenger changes induced by more dominant mutations is of great interest but has not yet been determined [130].

Initial studies targeting miRNAs as therapeutics in cancer and other diseases have been promising, with significant results in treating cancer in mice and hepatitis C virus in humans [131, 132]. However, a recent clinical trial targeting the *miR-34* family in primary liver cancer was terminated due to patients demonstrating immune-related adverse effects (ClinicalTrials.gov Identifier: NCT01829971) with the precise cause for these effects being currently unreported. There are still many outstanding questions about the basic biology of miRNAs, including the mechanisms by which miRNAs identify targets, the extent of miRNA-based regulation, biological consequences of targeting, and their tissue-specific roles, all of which present significant limitations for using miRNAs as therapeutics.

Hypothesis and specific aims

Given the well-established spatial and temporal regulation of miRNA expression and their potential to regulate gene pathways, it is conceivable that miRNAs play an essential role in producing and maintaining tissues. However, other than individual miRNA-gene interactions, the precise tissue-specific functions of miRNAs are not well understood. There is a need to identify the targets of miRNAs in high-throughput, especially in the context of diseases, where dysregulation of miRNAs is well characterized, but their contributions to diseased states are not well known.

I hypothesize that miRNAs have distinct tissue-specific roles that contribute to the gene expression differences seen across tissues. To address this hypothesis, I developed and used resources to study post-transcriptional regulation in high-throughput, optimized a system to identify miRNA targets in specific tissues of a living organism, and tested some of the consequences of tissue-specific miRNA targeting for gene expression. The approaches and techniques used to achieve each of these conditions are discussed below.

Aim #1: Develop the human 3'UTRome v 1 (h3'UTRome): A resources to rapidly identify miRNA targets in high-throughput (Chapter 2)

The first aim addresses the lack of high-throughput resources to study post-transcriptional gene regulation in humans. I have developed the h3'UTRome, a collection of cloned and sequenced human 3'UTRs, in a modular system that is convenient for downstream analysis. To better understand the principles that guide miRNA targeting, I

screened candidate 3'UTRs from the h3'UTRome for targeting by two miRNAs

implicated in cancer. The results from this study revealed potential mechanisms by which

miRNAs influence tissue identity and are perhaps dysregulated in cancer.

Aim #2: Identify tissue-specific targets of miRNA in the intestine and body muscle

tissues of *Caenorhabditis elegans*. (Chapter 3)

       Tissue-specific targets of miRNAs are not well characterized, and this aim seeks

to address this deficiency by profiling the miRNA targets in the tissues in the model

organism *C. elegans*. I optimized, validated and used this system to isolate, sequence and

annotate the miRNA targets from the intestine and body muscle of mixed stage *C.*

elegans. The identified miRNA targets showed that intestine and body muscle use this

form of regulation to different extents and that miRNA targets in each tissue correlate

with tissue-specific functions.

Aim #3: Explore the consequences of miRNA-based gene regulation to tissue-specific

gene expression patterns in *C. elegans*. (Chapter 4)

       Many of the tissue-specific targets of miRNAs in the intestine and body muscle

are RBPs. However, the functional relevance of miRNAs targeting these genes is not

known. In this aim, I utilized biochemical, genetic and genomic approaches to identify

the consequences of miRNAs targeting of specific RBPs and found evidence that

suggests an unexpected role for miRNAs in regulating the biogenesis of mRNAs.

CHAPTER 2

The h3'UTRome v 1 – A publicly available compendium of human 3'UTRs

Publication note

Overview

3'untranslated regions (3'UTRs) are the sequences located immediately downstream of the STOP codon of mature mRNAs. Although historical attention focused on protein coding sequences and upstream regions, 3'UTRs have recently become subject to intense study because they are targets of a variety of regulatory molecules, including RNA binding proteins (RBPs) and small non-coding RNAs (ncRNAs), that recognize small cis-elements present in the 3'UTRs. These cis-elements play critical roles in deciding the fate of the mRNA via various mechanisms. These include co-transcriptional processing, modulating protein translation, mRNA localization and trafficking, and mRNA degradation and stability [148]. Disruption of these processes is known to affect

diverse developmental and metabolic processes and contributes to the various diseases, including neurodegenerative diseases, diabetes, and cancer [130, 149-151].

RBPs play a role in every aspect of mRNA biogenesis, such as stability, localization, translation, and decay. The human transcriptome contains approximately ~400 proteins with distinct RNA binding domains [152], and their deregulation is linked to major neurodegenerative disorders, cancer, and muscular dystrophies. Compared to transcription factors, which generally bind highly specific linear DNA sequence elements, elements in 3'UTRs targeted by RBPs are generally more degenerate. Since RNA is a single-stranded molecule and RBP binding is mostly dictated by local folding and polarity, this can be challenging to identify bioinformatically [152]. Consequently, RBPs have the potential to bind to multiple elements in different 3'UTRs, leading to intricate, dynamic, and mostly unknown networks of RNA-protein interactions.

3'UTRs are also targeted by a class of post-transcriptional regulators known as microRNAs (miRNAs), which are short non-coding RNAs that bind to complementary sequences in the 3'UTRs of metazoans [148]. Once bound, based on the degree of complementarity, miRNAs can induce either translational repression or mRNA degradation [153]. MiRNAs canonically recognize targets in 3'UTRs via Watson-Crick base pairing, requiring complementarity with as few as six consecutive nucleotides between the 5'end of a mature miRNAs and the 3'UTR of a target transcript [153]. However, recent evidence suggests that miRNAs do not require perfect complementarity with target 3'UTRs to induce functional translational repression, and non-canonical interactions are frequent [109]. Because miRNA target elements are degenerate and

small, they are difficult to detect. Thus, a vast majority of biologically relevant miRNA targets are still unknown. Based on bioinformatic predictions of miRNA-binding sites in 3'UTRs, it has been proposed that each miRNA controls large networks of hundreds of mRNAs [154]. However, a recent analysis of the predictive performance of several of the most prominent prediction algorithms, such as TargetScan [110], PicTar [111] and DIANA-microT [112] report extremely high false negative rates [53, 109, 155]. While these algorithms are very useful for candidate gene approaches to identify miRNA targets, the extremely high error rates make high-throughput target detection challenging. When this challenge is paired with the absence of a publicly available and comprehensive 3'UTR library, the field currently lacks tools to systematically study miRNA targets, which is the gold standard in miRNA biology.

Several genomic resources are currently available to systematically study gene expression and its regulation in humans. The human ORFeome, for example, is a collection of over 12,000 human protein-coding genes cloned in modular vectors and optimized to study the dynamics of gene expression [156, 157]. The ORFeome has been used to characterize genome-wide protein-protein interaction networks, leading to important discoveries relevant to human disease [156]. High-throughput resources such as this can significantly advance our understanding of gene functions in multicellular organisms. Unfortunately, such a standardized HT tool to detect and study regulatory elements in 3'UTRs are not available since 3'UTR sequences are not present in the ORFeome. Some individual 3'UTR clones are available commercially, but these products

have sporadic coverage, are too expensive for HT studies, use only proprietary vectors and are not compatible with the ORFeome.

Furthermore, endogenous full-length 3'UTRs frequently undergo alternative processing in a tissue-specific fashion [158], which limits the biological relevance of experiments that use truncated or partial 3'UTRs. A recent study attempted to overcome the limited availability of 3'UTRs by using ~240,000 short RNA sequences, containing all possible 9-base nucleotide permutations immobilized on microarrays to study the binding requirements of 205 human RBPs [152]. Although this work and others highlight important binding properties of RBPs, they do not necessarily reflect natural settings, where accessory elements near binding sites that may cooperate with the RBPs targeting are not present.

Recently, our group experimented with the usage of a pilot human full-length 3'UTR library to detect miRNA targets in 3'UTRs using a scalable dual-luciferase assay named Luminescent Identification of Functional Elements in 3'UTRs (3'LIFE) [109, 159]. Although we cloned and screened only ~300 query 3'UTRs, the proof of principle 3'LIFE screen was highly effective at the rapid and efficient discovery of many novel targets for two cancer relevant miRNAs, *let-7c* and *miR-10b* [109]. This pilot screen demonstrates the value of such an unbiased HT approach and supports the need for the development of a publicly available genome-wide 3'UTR library.

Furthermore, there is a critical need in the field for a high-quality and standardized human 3'UTR resource, which could be widely used in the community to study miRNAs and RBPs using full-length 3'UTRs in unbiased and HT experiments.

To overcome these limitations, we have developed the first publicly available and high-quality human 3'UTR clone library, sequenced verified and cloned in modular vectors amenable to various downstream analyses. This resource enables the systematic study of 3'UTR biology, can be used to efficiently detect miRNA and RBP targets at high resolution, and study mRNA localization and dynamics. In the context of disease states, this library allows the study of crucial disease alterations in post-transcriptional processing, such as disease-specific: 1) mRNA mislocalization, 2) alternative polyadenylation, 3) altered miRNA expression, 4) mutation of RNA binding protein elements in 3'UTRs, and 5) more generally, the contribution of post-transcriptional gene regulation to gene output in disease initiation and progression.

Results

The human 3'UTRome v1 clone collection (h3'UTRome v1) consists of 1,461 unique, cloned and sequence-validated human 3'UTRs from transcription factors, kinases and other regulatory genes (Figure 2.1). This collection is contained in modular

# of 3'UTRs targeted	1,815
# of 3'UTRs cloned	1,461
% cloning success	80.1%

Figure 2.1 - Overview of 3'UTRs targeted for the h3'UTRome v1. We targeted a panel of 1,815 unique human 3'UTRs and successfully cloned, and sequence verified 1,461 unique 3'UTRs (80.1 % cloning success).

Gateway® compatible Entry vectors is amenable for large screens and is publicly

available to the community through the at the DNASU plasmid repository

(http://dnasu.asu.edu). (https://dnasu.org/DNASU/Home.do).

Primer design and genomic PCR

As a first release, we targeted and designed genomic primer pairs encompassing

the 3'UTR regions of 1,815 human protein-coding genes using the human genome

release 19 (GRCh37/hg19 Feb. 2009)[133]. The forward primers used for the genomic

PCR were designed to anneal within the last exon of the target gene, ending with the

gene-specific STOP codon in frame with the rest of the transcript (Figure. 2.2).

This expedient allowed us to increase the melting temperature of each forward

primer since the G/C content drops considerably after the STOP codon. In addition,

designing the forward primer within the open reading frame provides the 3'UTR with its

original gene-specific STOP codon at its 5'end, allows for convenient in-frame
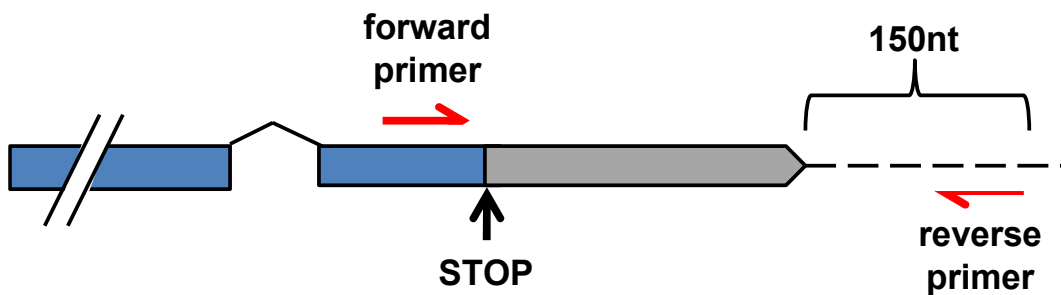
Figure 2.2 - Primer design used to target 3'UTRs and downstream processing elements. The forward primers used to amplify 3'UTR genomic loci were anchored within the last exon of each transcript, ending with the gene-specific STOP codons. The reverse primers bound 150 nucleotides downstream of the annotated transcript.

integration with the human ORFeome library, which instead lacks termination codons

[134] (Figure. 2.2). The melting temperatures of the primers ranged from 50 to 76 °C.

Given this wide range of temperature, we opted for a touchdown genomic PCR approach,

starting at 66 °C and decreasing by 1 °C each cycle [94]. The reverse primers were

designed to target a genomic site 150 nt downstream of the annotated transcript,

encompassing downstream elements that may play a role in mRNA 3'end formation

(Figure. 2.2). We added the Gateway® recombination elements attB2 (forward primers)

and attB3 (reverse primers) to the 5'ends of the genomic primers, to facilitate the cloning

into Gateway® compatible Entry vectors. A minimum of 200 ng of genomic DNA per

reaction was required to obtain an enriched PCR product while minimizing non-specific

amplicons, which is known to impact the recombinational cloning procedure. The

complete pipeline used in this study is shown in (Figure 2.3)
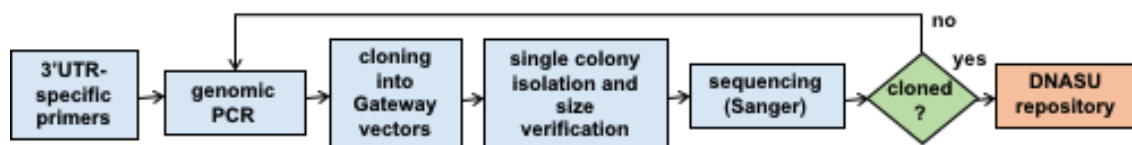
Figure 2.3 - Flow chart summarizing the cloning pipeline of the h3'UTRome v1.
Genomic PCR was performed using 3'UTR specific primers and the PCR products were
shuttled into Gateway® Entry vectors by recombinational cloning. Single cloned
colonies were isolated and screened based on the expected 3'UTR length using PCR
and gel electrophoresis. Bacterial colonies passing the screen were then re-arrayed, and
the cloned 3'UTRs were sequenced using the Sanger sequencing method. The sequence-
verified 3'UTRs were submitted to the DNASU plasmid repository for public
distribution. 3'UTRs that were not successfully cloned were subject to the second pass
of cloning.

Gateway® recombinational cloning

The full understanding of gene expression must consider both transcriptional and post-transcriptional regulation, requiring attention to the transcriptional promoter, the ORF and the regulatory sites within the 3'UTR. The human 3'UTRs in this collection were cloned into the pDONR P2r-P3 Gateway® Entry vector (Invitrogen) using BP recombinant cloning. This vector is part of the three-fragment Gateway® technology, which allows modular cloning of a given promoter, an ORFeome entry and, correspondent 3'UTR to be assembled in order into a single vector in the same reaction. This will enable investigators to combine these 3'UTRs with different ORFs (which are already available in the ORFeome collection) to create both natural and novel regulatory contexts. Current protein expression vectors typically rely on viral 3'UTRs, such as the SV40 polyA, which often do not reflect natural translational levels or post-transcriptional regulation. In addition, natural 3'UTRs may contribute to proper localization and stability. This technology is also compatible with the 3'LIFE assay system and has been previously used to screen for functional miRNA targeting in 3'UTRs [94]. Successfully cloned colonies were isolated and grown in LB and analyzed by colony PCR using primers specific to the pDONR P2r-P3 backbone. The PCR amplicons were analyzed by agarose gel electrophoresis and screened based on the expected lengths of the 3'UTRs (Figure 2.4). We observed an inverse correlation between the size of the inserted 3'UTR and the BP cloning success rate (Figure 2.5). A size bias during the BP cloning reaction has been previously reported [135], with a decreased efficiency for amplicons greater than 1,000nt and in agreement with our observations [135].

29

3'UTRs in the h3'UTRome v1 are enriched with longer 3'UTR isoforms and on average contain longer 3'UTRs than those within the human transcriptome (Figure 2.6). The nucleotide lengths of the human 3'UTR clones in this release span from 200nt to 2,500nt and have a median length of 1,159nt, as opposed to the median length of 3'UTRs within the human genome, which is 1,040nt (Figure 2.6, purple and red arrows).
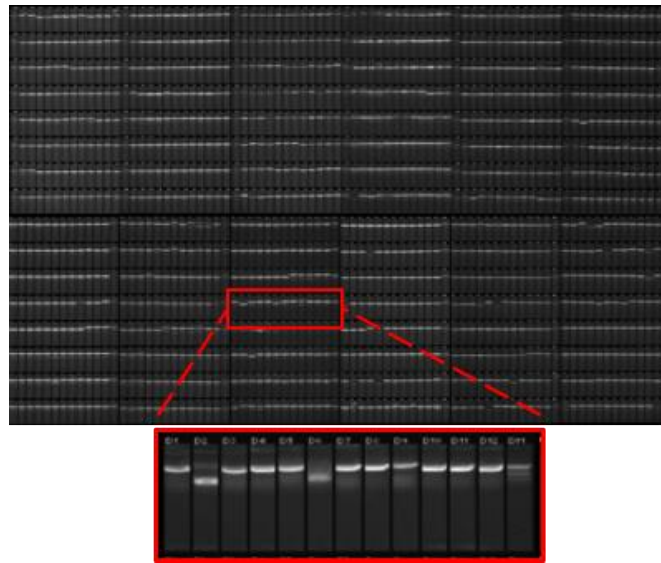
Figure 2.4 - Electrophoretic analysis of PCR products from the complete h3'UTRome v1. The sizes of 3'UTRs from 1,461 PCR reactions were analyzed on ethidium bromide stained gels.
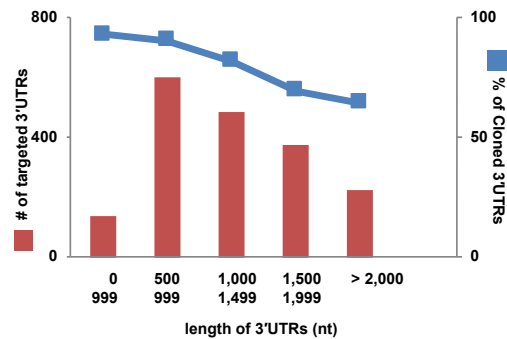
Figure 2.5 - Percentage of cloning success vs. 3'UTR length. The efficiency of 3'UTR cloning (blue line) decreases ~30 % beyond the length of 1,000 nucleotides in the cloned dataset (red). The maximum cloning efficiency for amplicons smaller than 499 nt was 92.8%, while the cloning efficiency of 3'UTRs greater than 2,000nts was 64.3%.

30

The first pass of cloning produced a yield of 1,410 bacterial colonies with PCR products of the expected size. We performed a second pass on all 405 missed 3'UTRs and gained an additional 172 3'UTRs, a 12% increase to the total number of size-verified clones (Figure 2.3).

Sanger sequencing

A total of 1,582 clones verified by size were subsequently sequenced using the Sanger method using a custom primer anchored within the P2rP3 plasmid backbone. We used Perl scripts to perform BLAT alignments [133] using the Sanger trace files obtained
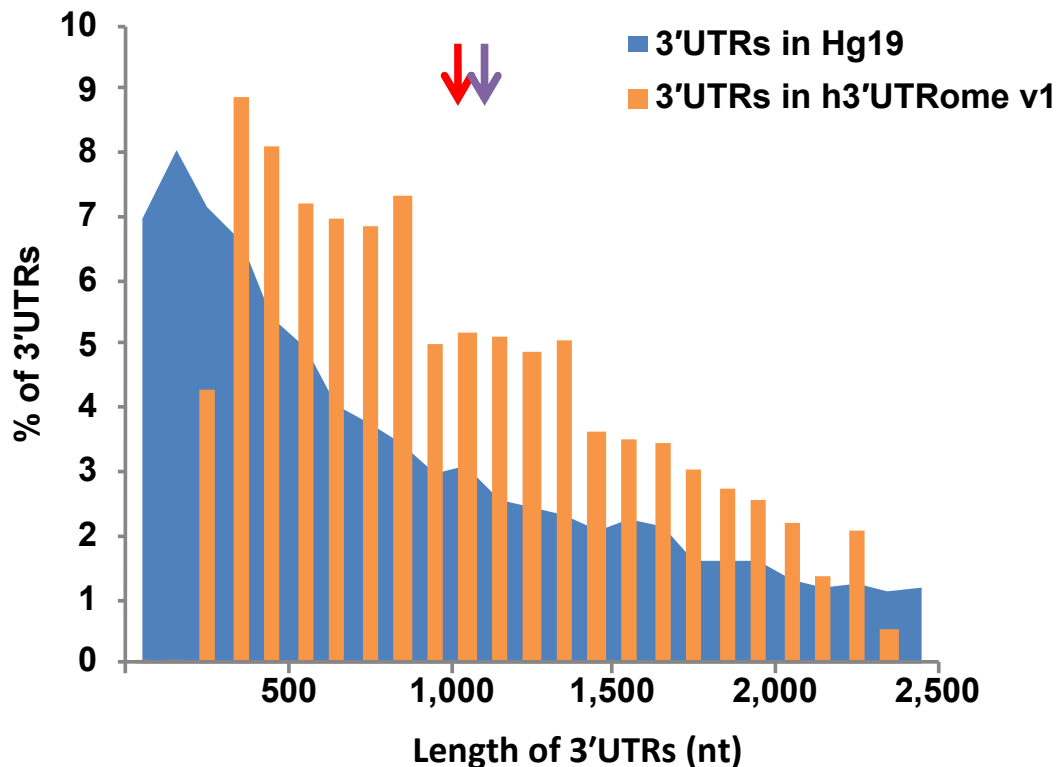
Figure 2.6 - Length of the 3'UTRs of genes in the human genome hg19 vs. the h3'UTRome v1. The 3'UTRs in the h3'UTRome v1 range between 200 nt to 2,500 nt, and are enriched for longer 3'UTR isoforms. The median length of 3'UTRs in the h3'UTRome v1 is 1,159 nt (purple arrow) while in the human genome hg19 it is 1,040 nt (red arrow).

31

during the sequencing. Our analysis revealed that out of the initially targeted 1,815 unique 3'UTRs, 1,461 were successfully sequence verified (~80% success rate from genomic PCR to sequence verification).

3'UTRome library overview

The human 3'UTR clones contained in the h3'UTRome v1 are unbiased towards any particular regions of the genome and correspond to ~6-10% of the total protein-coding genes present in each chromosome (Figure 2.7). The source of DNA used for the genomic PCR was GM12878, a lymphoblastoid cell line of female origin recommended as a Tier 1 cell line by the ENCODE project. Over 54% of the 3'UTRs in the h3'UTRome v1 overlap with genes present in the hORFeome V8.1 (Figure 2.8) [136].

We targeted 971 3'UTRs of genes already present in the ORFeome and successfully cloned 790 3'UTRs (Figure 2.8). For this first release, we targeted predominantly 3'UTRs of genes previously classified as transcription factors [137, 138], kinases [139], and RBPs [140] (Figure. 2.9). We targeted the 3'UTRs of this class of genes because they have widespread regulatory functions and have corresponding ORFeome clones. The h3'UTRome v1 release includes 3'UTRs for 985 transcription factors, 171 Kinases and 156 RBPs (Figure. 2.9).

Library distribution

The h3'UTRome v1 library is distributed by the DNA repository DNASU (http://dnasu.asu.edu), a public plasmid repository hosted at the Biodesign Institute at

Arizona State University, which already distributes over 180,000 individual plasmids and full genome collections, including the human ORFeome [141]. Users can either search for a given 3'UTR clone, a plate or order the complete dataset. Many researchers are not interested in HT screens nor have the resources for large screens in their departments, but want to detect miRNA targets, mutations, or truncation of regulatory elements in the 3'UTR of their gene of interest. These researchers will be able to accelerate their research significantly because they can now order the correct ORF, 3'UTR clones and the vectors they need for their analysis at a reduced cost. To simplify the ordering procedure, we have given a unique ID prefix 'HSU' to the human 3'UTRs available with this release.

chr.	ORFs	# cloned	%
1	1,987	182	9.2
2	1,214	117	9.6
3	1,052	82	7.8
4	747	54	7.2
5	860	79	9.2
6	1,025	104	10
7	876	88	10
8	660	58	8.8
9	759	74	9.7
10	732	68	9.3
11	1,270	75	5.9
12	1,010	102	10
13	314	32	10.2
14	596	62	10.4
15	560	47	8.4
16	807	80	9.9
17	1,142	102	8.9
18	266	28	10.5
19	1,378	227	16.5
20	525	63	12
21	215	16	7.4
22	427	33	7.7
X	813	70	8.6
Y	59	-	-

Figure 2.7 - The h3'UTRome v1 contains 3'UTRs for 6-10% of the ORFs within each chromosome. None of the 3'UTRs cloned in this release originates from the Y chromosome, as the source of genomic DNA used was of female origin.

33
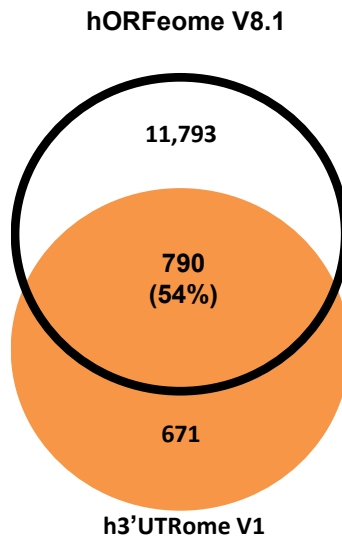
Figure 2.8 - The degree of overlap between the h3'UTRome v1 and the hORFeome V8.1. More than half of the 3'UTRs cloned in the h3'UTRome v1 contain corresponding clones present in the hORFeome V8.1.
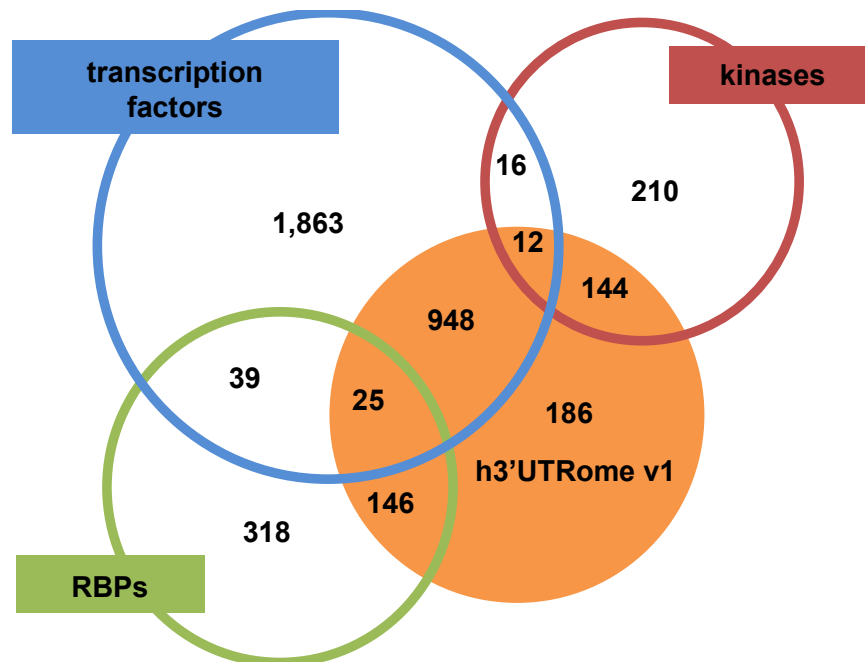
Figure 2.9 - The h3'UTRome v1 is enriched for the 3'UTRs of genes involved with gene regulation. Predominantly it contains 3'UTRs of transcription factors (33.9%), kinases (40.8%), and RNA binding proteins (32.4%).

3'LIFE validation screen

The 3'LIFE screen is a high throughput dual luciferase assay, previously shown to detect functional repression of test 3'UTRs by query miRNAs [94, 108]. The 3'LIFE screen utilizes Gateway® cloning technology and is fully compatible with the h3'UTRome v1. To demonstrate the usability and functionality of this library we have selected 87 human 3'UTRs from the h3'UTRome v1 library (Appendix A) and screened for miRNA targets of two disease-relevant miRNAs: *let-7c* and *miR-221* using the 3'LIFE assay [94, 108]. *let-7c* is a well-characterized tumor suppressor gene, is down-regulated in many cancers, and is known to target genes in the RAS pathway [116]. Conversely, *miR-221* is frequently overexpressed in breast cancer, hepatocellular carcinomas, glioblastoma and prostate cancer [142-145], and has been shown to target several tumor suppressor genes such as Kip-1(p-27), CDKN1B, CDKN1C, PTEN, ARHI and PUMA [143, 146, 147]. In addition, *miR-221* is known to be involved with muscle damage repair and atherosclerosis [134, 148]. One of the goals of this experiment was to use this 3'UTR library to rapidly identify bona fide miRNA targets from false targets predicted by miRNA targeting software. These programs, such as TargetScan [95], PicTar [96] and DIANA-microT [97] are known to have high false negative rates (~43%) [22, 94, 149] and false positive rates (~66%) [94, 150, 151], and cannot be used alone to definitively assign targets.

These 87 human-3'UTRs were enriched with *let-7c* and *miR-221* predicted and validated targets from all three prediction software algorithms (Appendix A) (9 predicted and 3 validated targets for *let-7c*, and 10 predicted and 9 validated targets for *miR-221*).

For the *let-7c* screens, we also included two genes that contain validated miRNA targets identified in a previous screen (Appendix A)[94]. In addition, since miRNAs preferentially target genes within the same regulatory pathways [94], and *let-7c* was previously shown to target the RAS family of genes [94, 116], we were interested to test if *let-7c* could also target an additional 25 members of this pathway, as defined by Gene Ontology [152] and KEGG databases (Appendix A)[153].

We shuttled 87 human 3'UTRs from the h3'UTRome v1 clone library into the 3'LIFE vector using LR recombination reactions. Using this custom library, we performed 435 fully automated transfections and dual luciferase experiments. The results of the screen are shown in (Figure 2.10). Using a cut-off for functionally repressed targets at a repression index of 0.8 and a p-value <0.05, we obtained 19 statistically significant hits for *let-7c*, and 13 for *miR-221* (Figure 2.10).

Our results validate 4 out of 9 of the *let-7c* targets predicted by prediction software algorithms [95-97]. Within the predicted hits, we detected all three previously validated targets (CDC25A, TRIM71, and BCL2L1) and an unvalidated, predicted target (RNF7) (Figure 2.10). Furthermore, we detected an additional 10 novel and unpredicted targets for *let-7c* (Figure 2.10). We found that one of these novel targets PAK3 was predicted by the prediction algorithm miRanda [99], which takes into account non-canonical seed interactions. Of note, 3 targets within this group (MCM2, BUB1B, and GMNN) were previously correlated indirectly with *let-7* expression [154].
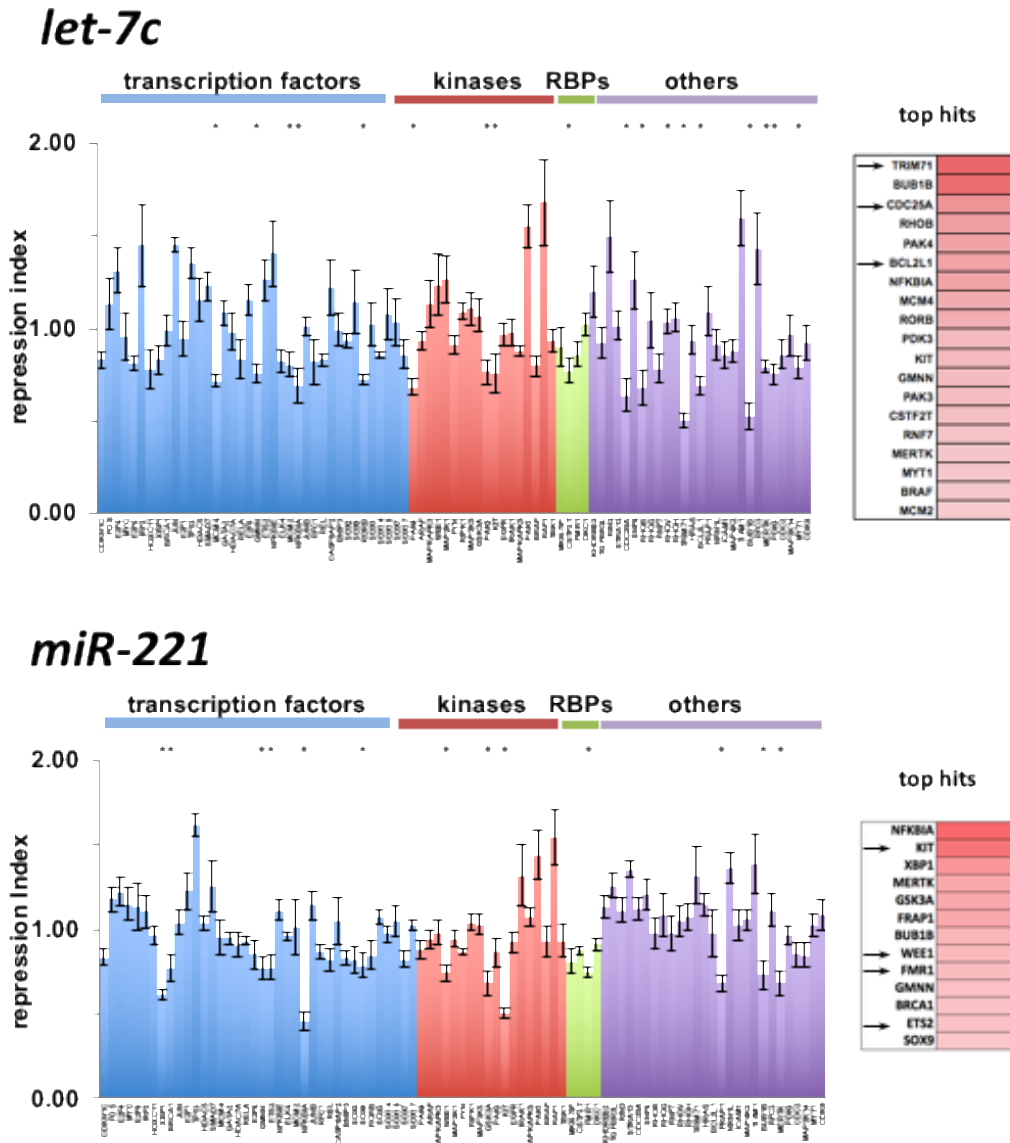
Figure 2.10 (next page) - The h3'UTRome v1 as a resource to detect miRNA targets in high throughput. 3'LIFE screen performed on 87 human 3'UTRs extracted from the h3'UTRome v1, queried against two miRNAs *let-7c* and *miR-221*. The repression index was calculated for each 3'UTR and indicates a normalized ratio between Firefly and Renilla luciferases. A minimum cutoff for repression was drawn at 0.8, and the asterisks denote statistically significant repression (p < 0.05). The top hits for each miRNA are displayed as a heat map on the right panels. Increased intensity of color indicates greater repression. 18 top hits were identified for *let-7c* of which 3 were previously validated (as denoted by the arrows). 13 top hits were identified for *miR-221* of which 4 were previously validated (as denoted by the arrows)

For *miR-221*, our results validated 4 out of 10 of the *miR-221* targets predicted by the prediction software [95-97], and an additional 4 out of 9 of the targets previously validated by others (WEE1, ETS2, FMR1, and KIT) [155-158]. Interestingly, we were unable to detect repression in 3'UTRs of 5 genes previously known to contain *miR-221* responsive elements (CDKN1C, FOS, IRF2, ICAM1, PAK1) [159-163]. Upon further review, we found that repression of all five targeted elements was demonstrated using truncated sections of the 3'UTR. Thus, the observation that the 3'LIFE screen did not detect these targets could be caused by the inability of these elements to recruit *miR-221* when expressed within their full-length endogenous 3'UTRs, or by the presence of alternative polyadenylation events that cause the loss of these elements [94]. Two targets of *miR-221* called by the prediction software [95-97] were also not detected as hits in our assay (KHDRBS2 and RORB). We also discovered 9 novel and unpredicted targets for *miR-221* not anticipated by major prediction software [95-97] or detected by others. Within this group, FRAP1 was the only gene predicted by miRanda [99]. Perfect complementarity within the seed region is considered the canonical indicator of miRNAs targeting. Interestingly, most of these novel targets do not always contain canonical seeds. Recent studies indicate that miRNAs are also capable of recognizing non-canonical elements in target mRNAs [94, 150, 164, 165], supporting our findings.

9 out of 18 bioinformatically predicted targets are validated by our experiments [95-97] (50% false positive rate), which is in accordance with the false positive rates of prediction algorithms reported in previous studies [22, 94, 149-151]. In previous studies, we used repression data from the 3'LIFE assay to identify and validate functional

miRNA binding sites [94]. With experimentally validated miRNA target sites, targeting signatures can be extrapolated to refine target predictions for specific miRNAs.

Interestingly, while the 3'LIFE assay is designed to detect repression of 3'UTRs by miRNAs, we identified several 3'UTRs that significantly enhanced the expression of the luciferase reporter gene in the presence of *let-7c* and *miR-221* (Figure 2.10). Perhaps these enhancements are caused by increase stability of a given 3'UTR due to direct or indirect interactions with the query miRNAs.

The ability to systematically screen large numbers of human 3'UTRs allowed in-depth analysis of high-confidence target genes regulated by different miRNAs and may reveal novel mechanisms that miRNAs use to regulate biological processes. For example, a gene ontology analysis of the *let-7c* top hits showed an enrichment for genes involved in cell cycle checkpoint regulation, while a similar analysis for *miR-221* revealed a relationship with genes involved in negative regulation of muscle differentiation.

In addition, out of the 25 genes involved in the RAS pathway, our screen identified 7 genes directly targeted by *let-7c* (RhoB, PAK1, PAK3, BRAF, NFKBIA, BCL2L1, KIT), suggesting a role for *let-7c* in regulating this pathway.

Discussion

The h3'UTRome is a publicly available tool for studying miRNA-based gene regulation

3'UTRs contain powerful regulatory elements that are critical in various biological processes, yet remain poorly characterized due to the absence of genomic tools

that allow their systematic study. In this work, we have prepared the first human 3'UTR clone collection named h3'UTRome v1, which is produced to a single high-quality standard. This library is compatible with the cloning technology used to produce the human ORFeome, expanding the potential of well-established operating methods for high throughput experimentation. The h3'UTRome v1 library is sequence verified, and readily available to the community with simple online access to information through the DNASU repository [166], at a low cost to all scientists with minimal restriction. The h3'UTRome v1 can be easily used in HT experiments to systematically study RBP targeting in 3'UTRs, mRNA localization and the role of small ncRNAs in post-transcriptional gene regulation.

In order to demonstrate its utility, we performed a screen with 87 human 3'UTRs cherry-picked from the h3'UTRome v1 and rapidly identified 27 miRNA targets for two disease-relevant miRNAs, *let-7c* and *miR-221*. Within this pool, we identified 18 novel targets for these two miRNAs, which were previously uncharacterized (67%). In addition, we were able to eliminate 9 out of 18 bioinformatically predicted targets (50% false positive), and rapidly associate miRNA activities to biological pathways using a rapid screening technology. We found that while *let-7c* targets predominantly oncogenes, *miR-221* targets a mix of both tumor suppressors and oncogenes, implying that *miR-221* may be involved with more context-dependent regulation of gene expression [122] (Figure 2.11).
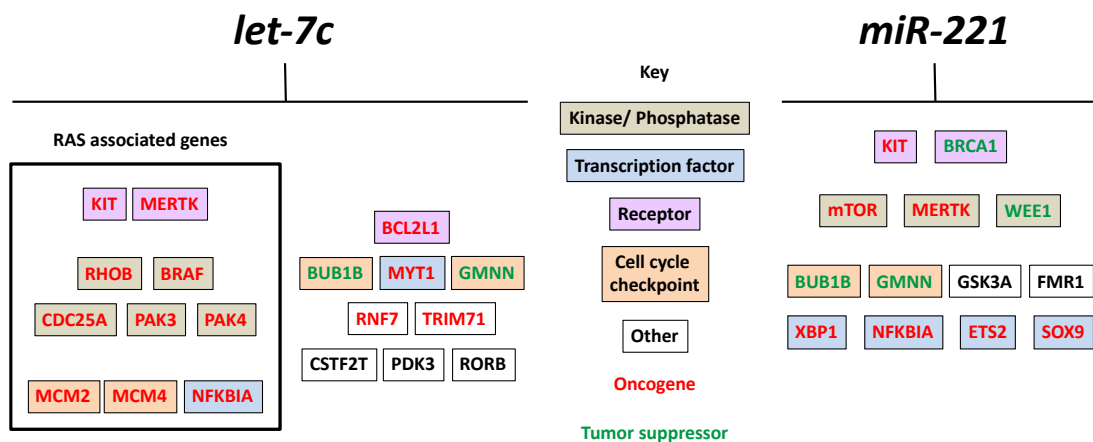
Figure 2.11 – Genes identified as targets of *miR-221* and *let-7c* using 3'UTRs from the h3'UTRome. *miR-221* targets a mix of oncogenes and tumor suppressors indicating that the role of this miRNA in cancer may be dependent on the cancer type and context. Conversely, almost all the targets of *let-7c* are oncogenes, further endorsing the tumor suppressing role of this miRNA. We found that *let-7c* targets the RAS pathway is targeted at multiple points, from receptor tyrosine kinases, downstream signaling kinases and final effectors of the pathway

Experimental

Primer design

DNA primer sequences were designed using custom Perl scripts using the annotated 3'UTR sequences in the Human Genome release 19 (GRCh37/hg19 Feb. 2009) [133]. The forward primers were anchored upstream of the last exon of each gene and included the gene specific endogenous STOP codons in frame with the ORFome library (Figure 2.2). The reverse primers were designed to target sites 150nt downstream of the longest annotated transcript, as per the RefSeq annotation, in order to include downstream 3'end processing elements (Figure 2.2). Forward and reverse primers were fused to the attB2 (5'-GGGGACAGCTTTCTTGTACAAAGTGGAG-3') and attB3 (5'-

GGGGACAACTTTGTATAATAAAGTTG-3') Gateway® sequences to allow modular cloning into pDONR P2rP3 Entry vectors. The full list of primers used is available through DNASU (https://dnasu.org/DNASU) [166]. The first release of the h3'UTRome V1 targeted a panel of 1,815 3'UTRs and is enriched for transcription factors, kinases, RNA binding protein and other regulatory genes (Figure 2.9). The length of the 3'UTRs cloned in this release ranges between 200 and 2,500nt in length, which is larger than the average size of human 3'UTRs (Figure 2.6).

Genomic DNA

We used the NA12878 DNA sample obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research (Camden, New Jersey). This genomic DNA was extracted from the GM12878, a B-lymphocyte cell line of a human female subject. Once received, the genomic DNA was diluted to a concentration of 200ng/µl, aliquoted in 96-well PCR reactions, and stored at -80oC until use.

Genomic touchdown PCR

The reactions were conducted using Platinum Taq polymerase (Invitrogen) in 96-well plates using 200ng of genomic DNA per reaction. The reaction conditions were maintained as per the manufacturers protocol with changes to the annealing temperature of the reaction. The PCR conditions included 16 cycles of touchdown PCR, where the temperature of the annealing phase decreased by 1°C per cycle, ending at a temperature of 50°C. The reaction proceeded for 15 more cycles at an annealing temperature of 55°C.

The resulting PCR products were visualized on by electrophoresis on 96-well agarose gels and screened by size to determine successful amplicons.

Gateway® BP recombination reaction and transformation

Site-specific DNA recombination was used to clone the human 3'UTR PCR amplicons into the Gateway® Entry vector pDONR P2r-P3 (Invitrogen), using BP Clonase II Enzyme Mix kit (Invitrogen, Carlsbad, CA) following the manufacturer's specifications. DH5α E. coli cells were transformed with 1μl of the resultant reaction mixture, and screened the following day for successful recombinants using Lysogeny Broth (LB) agar plates with Kanamycin (Kan) antibiotic.

3'UTR isolation and size screening

In order to isolate single clonal populations, unique bacterial colonies for each 3'UTR clone were picked from the LB plates, and grown overnight in 96 deep-well plates containing LB (500mL) with Kanamycin resistance (50μg/mL) (total colonies picked=1,824). The resultant bacterial growths were used as a template to perform colony PCR reactions using M13 DNA primer pairs. The amplicons were then analyzed in 96-well agarose gels and positive clones were initially screened based on their expected size (Fig. 1d). Up to three more colonies for genes that did not satisfy our quality control inspection were picked (total colonies picked=753), and rescreened by repeating the bacterial colony PCR step. Bacterial colonies that passed the initial screen were re-arrayed and stored in glycerol stocks, while the primer pairs of the remaining genes were

used in a second pass, starting at genomic PCR to capture any 3'UTR missed (Figure 2.3).

Sanger sequencing

PCR analysis with M13 DNA primer pairs was performed for each positive 3'UTR clone, using overnight bacterial growths as a template and Phusion® Taq polymerase (New England Biolabs), as per manufacturers protocol. These PCR amplicons were then sent for sequencing at the DNA Lab, School of Life Sciences, Arizona State University, using the sequencing primer 1FP2rP3 seq (5'-GCATATGTTGTGTTTTACAGTATTATGTAG-3') which binds ~100nt upstream of the recombination element in the P2rP3 plasmid. 1,461 3'UTR clones successfully sequence verified and passed this step. The trace files for each 3'UTR clone successfully screened are available through DNAsu website [166]. Using custom a BioPerl script with Blat integration [133] we mapped our sequencing results to 1,461 unique 3'UTRs in the human genome.

3'LIFE screen

The 3`LIFE assay was performed as previously described [108]. We re-arrayed bacterial colonies from a panel of human 3'UTRs from the h3'UTRome v1, and grew the plate over-night in a 96 deep-well format using 200mL of LB in presence of kanamycin (50μg/mL). We used 1μl of the resultant overnight culture to perform the colony PCR with Phusion Taq polymerase (Invitrogen) as per manufacturers protocol. The amplicons

from the PCR reaction were shuttled into the pLIFE-3'UTR vector (DNASU Plasmid ID: EvNO00601503) by LR recombination using LR Clonase enzyme (Invitrogen) as per manufacturers protocol. 1µl of the resultant LR reaction mixture was transformed in DH5α *E. coli* cells. The transformed cells were then plated on LB agar plates containing ampicillin (100µg/ml), and incubated overnight at 37°C. Single bacterial colonies were isolated and grown overnight in 500mL of LB containing ampicillin (100µg/mL). The resultant overnight bacterial growth was screened based on size using agarose gel electrophoresis. Bacterial colonies from wells passing the screen were frozen as glycerol stocks and also grown overnight for 96-well plasmid DNA extraction as previously described [94, 108]. In order to express *let-7c* miRNA we used the pLIFE-miR *let-7c* construct [94, 108].

The miRNA *miR-221* was extracted from human genomic DNA derived from GM12878 cells using DNA primers containing Gateway® recombination elements (forward primer – 5`-GGGGACAGCTTTCTTGTACAAAGTGGAGTTTCAACATGATGTCATGATTAAATG-3`; reverse primer- 5`-GGGGACAACTTTGTATAATAAAGTTGCACCTTATCTCTGGTTTACTAGGCTG-3`). The amplified PCR amplicon was cloned into pLIFE-miR (DNASU Plasmid ID: EvNO00601504) vector using LR Clonase II enzyme as per manufactures protocol (Invitrogen). We designed the positive and negative controls for *miR-221* targeting by introducing 22nt long complementary sequences for the 3p arm (positive control) and 5p arm (negative control) arms of *miR-221* into the SV40 3'UTR by site directed

mutagenesis (Quikchange®, Invitrogen), as per manufacturer protocol. We used the *let-7c* positive control as previously described [94]. Plasmid DNA was extracted as was previously described [94]. The 3'LIFE assay was performed as previously described [94, 108]. In brief, 87 queried human 3'UTRs + 3 controls were transfected into HEK293T cells using the 96-well Shuttle nucleofection system (Lonza). Transfected cells were cultured for 72 hours and then lysed, then used to perform the dual luciferase assay. The screen was performed five times (435 reactions), and the resulting data was analyzed as previously described [94, 108]. The top hits for each miRNA were distinguished by requiring a minimum repression index of 0.8 and a p-value<0.05.

CHAPTER 3

Identification of the tissue-specific targets of microRNAs using the model organism *Caenorhabditis elegans*

Publication note

The research reported in this chapter has been submitted for publication in *Genetics*.
Kasuen Kotagama, Anna L Schorr, Hannah S Steber, and Marco Mangone. All co-authors have granted permission for this work to be included in this dissertation.

Overview

Multicellular organisms have evolved complex forms of gene regulation achieved at different stages throughout development and equally executed at pre-, co-, and post-transcriptional stages. Post-transcriptional gene regulation events mostly occur through 3' Untranslated Regions (3'UTRs), which are portions of genes located between the STOP codon and the poly(A) tail of mature eukaryotic mRNAs. 3'UTRs have been recently subjected to intense study as they were found to be targeted by a variety of factors, which recognize small regulatory elements in these regions and can modulate the dosage of gene output at the post-transcriptional level [167-169]. While these regulatory mechanisms are still poorly characterized, and the majority of functional elements remain unknown, disorders in the 3' end processing of mRNAs have been found to play critical roles in the loss of tissue identity and the establishment of prevalent diseases, including neurodegenerative diseases, diabetes, and cancer [51, 170-172].

3'UTRs are frequently targeted by a class of repressive molecules named microRNAs (miRNAs). MiRNAs are short non-coding RNAs, ~22nt in length, that are incorporated into a large protein complex called the microRNA-induced silencing complex (miRISC), where they guide the interaction between the miRISC and the target mRNA by base pairing, primarily within the 3'UTR [173]. The outcome of miRNA targeting can be context-dependent. However, mRNAs targeted by the miRISC are typically held in translational repression before degradation of the transcript [15, 65]. Initial studies showed that although mismatches between miRNAs and their targets are common, many interactions make use of perfect complementarity at a small conserved hexametrical motif, located at position 2-7, at the 5'end of the miRNA (seed region), [15, 65]. Later findings showed that while important, the seed region may also contain one or more mismatches while pairing with its target mRNA and that this element alone is not a sufficient predictor of miRNA targeting [91-93, 174]. Compensatory base pairing at the 3' end of the miRNA (nucleotides 10-13) can also play a role in target recognition [150, 175], and have been implicated in conferring target specificity to miRNAs that share the same seed regions [11, 98].

MiRNAs and their 3'UTR targets are frequently conserved and play a variety of roles in modulating fundamental biological processes across metazoans. Bioinformatic algorithms, such as miRanda [101], TargetScan [95] and PicTar [96], use evolutionary conservation and thermodynamic principles to identify miRNA target sites and are the preferred tools for miRNA target identification. Based on these algorithms it was initially predicted that each miRNA controls hundreds of gene products [176]. Recent high-

throughput wet bench approaches, have validated and expanded on these initial predictions, and provide further evidence that miRNAs can indeed target hundreds of genes, and regulate molecular pathways throughout development and in diseases [11, 22, 94, 105, 177]. In the past few years, several groups produced tissue-specific miRNA localization data in mouse, rat, and human tissues [178, 179] and cancer [180].

C. elegans is a small roundworm that is a well-characterized model organism to study post-transcriptional gene regulation. C. elegans, colloquially known as the worm, has a fully-annotated and cloned 3'UTRome [40, 181], and a lot is known about its RBP [182] and miRNA-based gene regulation [60]. The worm has a well-known, invariable cell lineage making it ideal for studying tissue development. In addition, our group has pioneered the use of C. elegans to study tissue-specific gene expression systematically [69, 158]. In a previous study, we developed a method to isolate and sequence high-quality tissue-specific mRNA from worms and published several integrative analyses of gene expression in most of the C. elegans somatic tissues, including the intestine and body muscle [69, 158]. As a consequence, the worm is an unparalleled model for studying miRNA regulation where we have information concerning the tissue-specific genes, and 3'UTR isoforms expressed.

A previous low-throughput study has identified hundreds of miRNAs and their targets in the intestine and body muscle of C. elegans, which are mostly involved in the immune response to pathogens [183]. This study used a microarray-based approach, which unfortunately does not provide enough depth to understand miRNA function in a tissue-specific manner fully. In addition, this study identified only a subset of miRNA

targets, which rely on the scaffolding proteins AIN-1 and AIN-2, later found to be only present at specific developmental stages [183, 184].

In *C. elegans* there are three Argonaute proteins execute the miRNA pathway and are named *alg-1*, *alg-2,* and *alg-5*. Recent transcriptome analysis in strains deficient in each of these members show a remarkable difference in function, where *alg-1* and *alg-2* are mostly expressed in somatic tissue and are functionally redundant, while *alg-5* is expressed exclusively in the gonads, interacts with only a subset of miRNAs and is required for optimal fertility [177]. A more recent study used a novel methylation-dependent sequencing approach (mime-Seq) and identified high-quality tissue-specific miRNAs in the intestine and body muscle tissues [185]. These studies show that there are indeed distinct functional miRNA populations in tissues, which can potentially reshape transcriptomes and contribute to the acquisition and maintenance of cell identity. Since most miRNA targets are only predicted, it is still unclear how these events are initiated and maintained.

To better understand the tissue-specific contribution of miRNA-based regulation to tissue identity, we isolated and sequenced the tissue-specific targets of miRNAs, from two of its largest and most well-characterized tissues in *C. elegans*, the intestine and body muscle. The total number of genes regulated in each tissue correlates with transcriptome size. However, more of the intestinal transcriptome is regulated when compared to the body muscle, suggesting that the degree of regulation by miRNA in tissues is heterogeneous.

We find that tissue-specific miRNA targets include genes that are important for development and genes that are involved with normal tissue function. MiRNAs are known to reduce gene expression at the protein level, as opposed to silencing or knocking out expression, and as such, it is reasonable that miRNAs would target tissue-specific genes to modulate their expression and contribute to tissue homeostasis. Using a bioinformatic analysis of the 3'UTRs of the tissue-specific targets we identify that APA, and the tissue-specific expression patterns of miRNAs appear to be driving forces of tissue-specific miRNA regulation.

Results

A method for the identification of tissue-specific miRNA targets

To study the contribution of miRNA activity in producing and maintaining tissue identity, we performed RNA immunoprecipitations of miRNA target genes in two of the
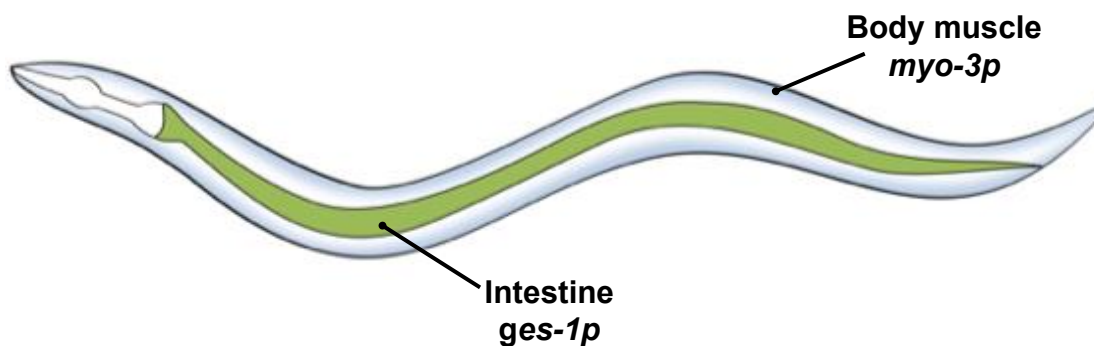
Figure 3.1 - The anatomical location of the two studied somatic tissues. We used the promoters of the intestine-specific gene *ges-1*(abnormal Gut ESterase) and the body muscle specific *myo-3*(MYOsin heavy chain structural genes) for the expression of tissue-specific constructs.

largest, morphologically different, and most well-characterized tissues in *C. elegans*: the

intestine [186] and body muscle [187] (Figure 3.1).

We took advantage of the ability of the Argonaute protein to bind miRNA target

genes and cloned *alg-1,* one of the worms orthologs of the human Argonaute 2 protein,

downstream of the green fluorescent protein (GFP). The expression of this construct was

then driven by the endogenous promoter (*alg-1p*), or restricted to the intestine (*ges-1p*) or

body muscle (*myo-3p*) using tissue-specific (TS) promoters (Figure 3.2). We produced
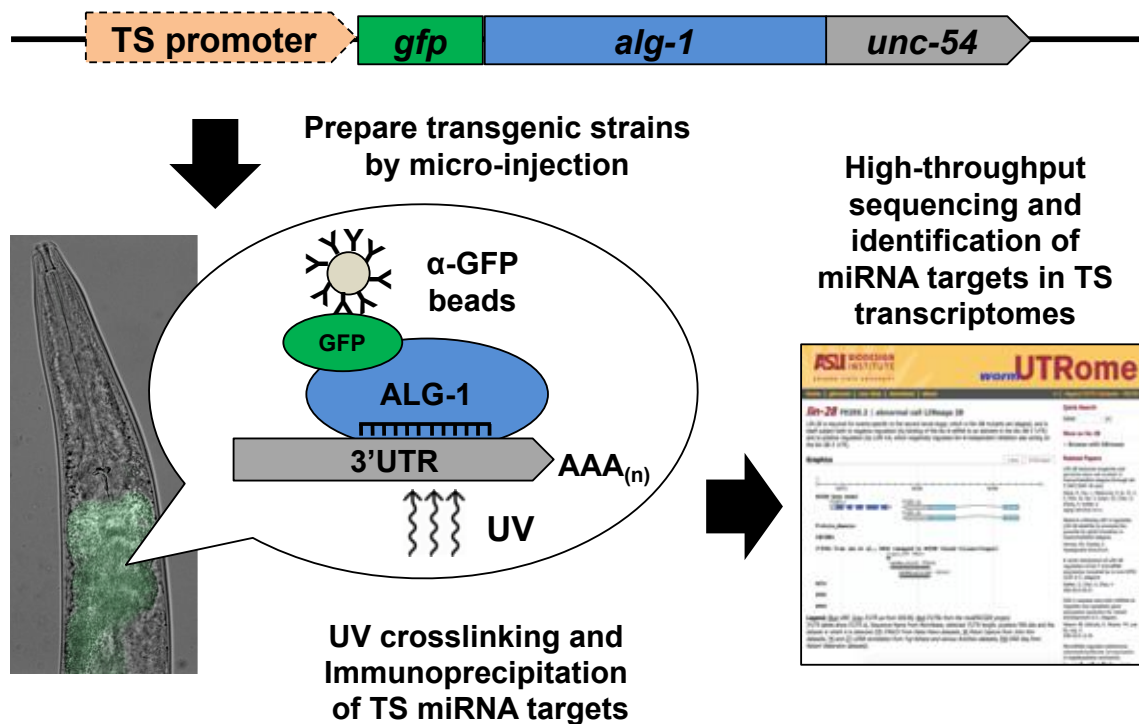
Figure 3.2 - Workflow for the identification of tissue-specific miRNA targets. We cloned the *C. elegans* Argonaute protein *alg-1* and fused it with the GFP fluorochrome and the unspecific *unc-54* 3'UTR. The expression of this cassette was driven in the intestine and body muscle by using tissue-specific (TS) promoters. These constructs were microinjected into MosSCI-compatible *C. elegans* strains to produce single-copy integrated transgenic animals. These strains were then subjected to UV crosslinking and lysed by sonication. The resulting lysate was subjected to RNA immunoprecipitations with α-GFP antibodies. The resultant tissue-specific miRNA target transcripts were purified, the cDNA libraries were made and sequenced using Illumina HiSeq.

transgenic strains for each construct (Figure 3.3) using single copy integration technology (MosSCI) [188, 189] to minimize the expression mosaics produced by repetitive extrachromosomal arrays. The strains were validated for integration using genomic PCRs and Western blots (Figure 3.4 and 3.5).

We then examined the functionality of our cloned *alg-1* in rescue experiments using the *alg-1*⁻/⁻ strain RF54(*gk214*). This strain has a decrease in fertility caused by the loss of functional *alg-1* [190], which was fully rescued by our cloned *alg-1* construct in a brood size assay (Figure 3.6 and 3.7), suggesting that our cloned *alg-1* is functional and able to mimic endogenous *alg-1* fully.
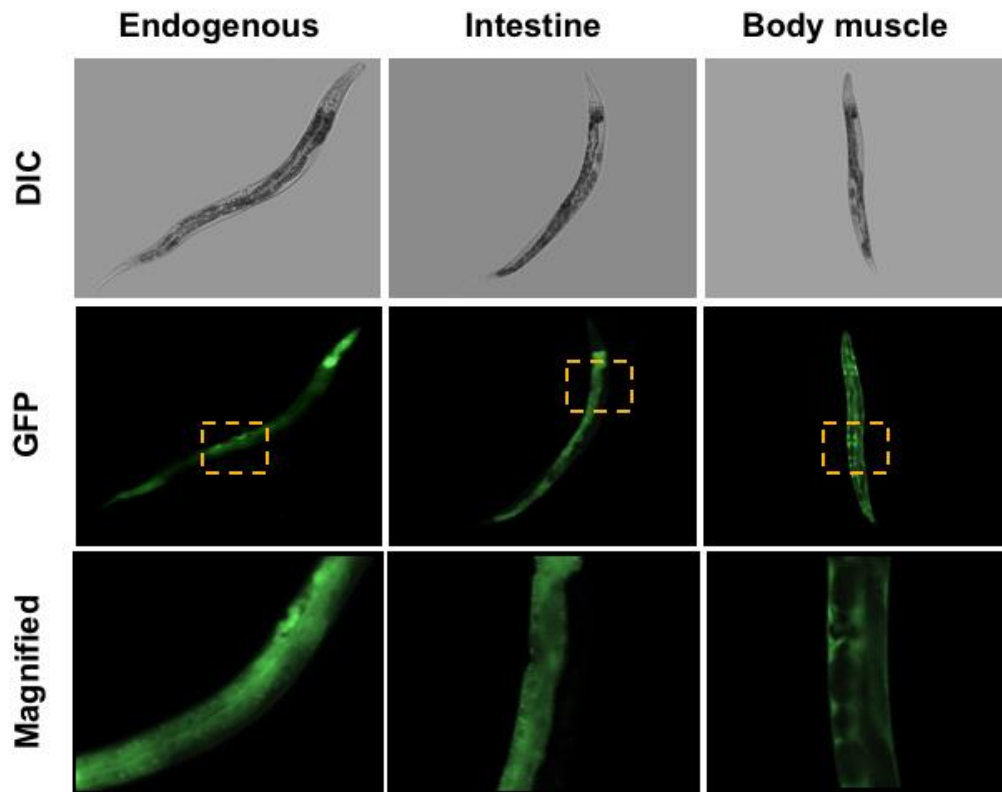
Figure 3.3 - Representative images of *C. elegans* single copy integrated strains. The expression of GFP tagged *alg-1* in endogenous (*alg-1p*), intestine (*ges-1p*), body muscle (*myo-3p*) tissues. The yellow box indicated magnified regions, yellow arrows mark intestine cells, and red arrows mark body muscle cells.
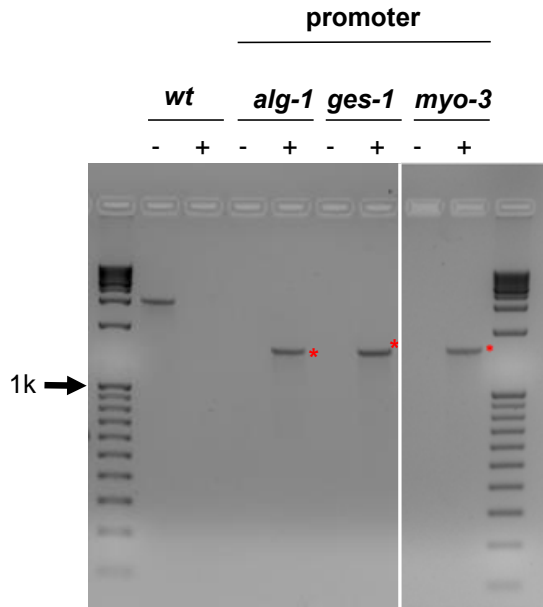
Figure 3.4 - Agarose gel electrophoresis of PCR products obtained from genomic DNA a using single-worm PCR approach from transgenic *C. elegans* strains prepared in this study. Each strain was tested using primers that detect either the presence (+) or the absence (-) of the integrated MosSCI transgene. Red asterisks mark expected bands confirming integration of our constructs.
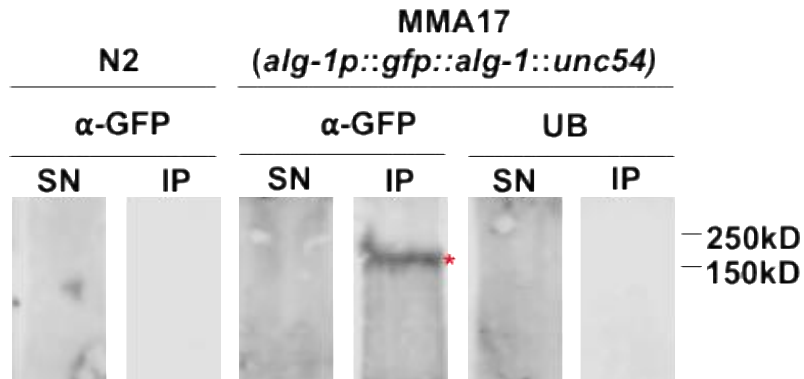
Figure 3.5 - Western blot experiments from wt N2 worms or strain expressing endogenous the ALG-1 pull-down construct. Immunoprecipitations were performed using either α-GFP antibodies or unconjugated beads (UB) on immunoprecipitated protein obtained from each strain. (IP) Immunoprecipitated protein, (SN) supernatant. Red asterisk marks the expected band.
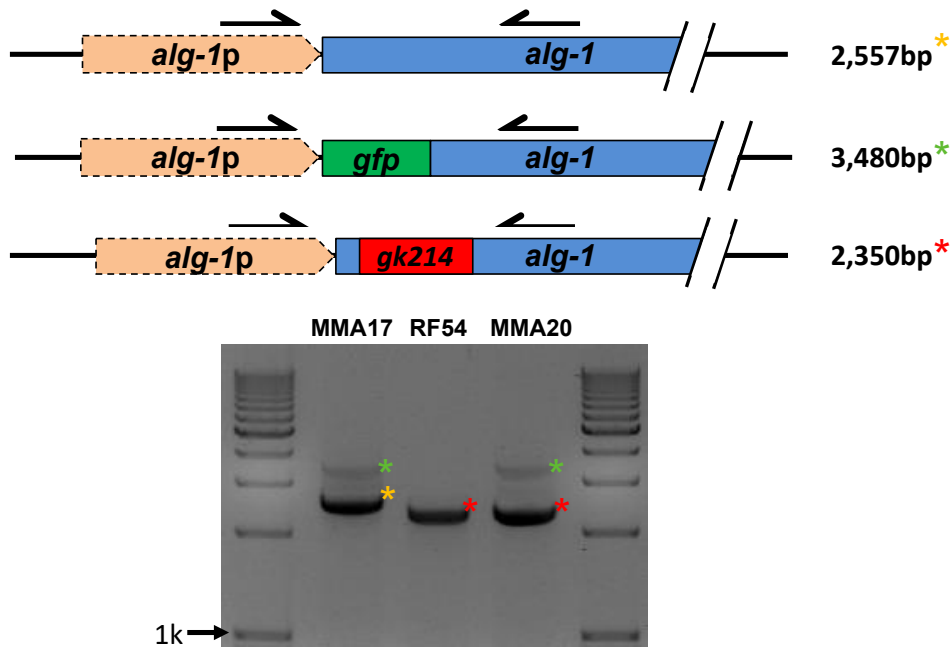
Figure 3.6 – PCR to genotype and validate worm crosses to rescue *alg-1* expression in RF54 with transgenic *gfp* tagged *alg*-1. An agarose gel used to visualize single worm genomic PCRs testing the presence of the *gk214* mutation before and after the cross**.** The yellow asterisks indicate the genomic locus with wt *alg-1* (2,557bp). The green asterisks indicate *gfp* tagged transgenic *alg-1*. The red asterisks indicate the presence of the *gk214* mutation (2,350bp). In the MMA20 strain, we re-introduced our cloned *gfp* tagged *alg-1* to the RF54 strain, which lacks of endogenous *alg-1*.
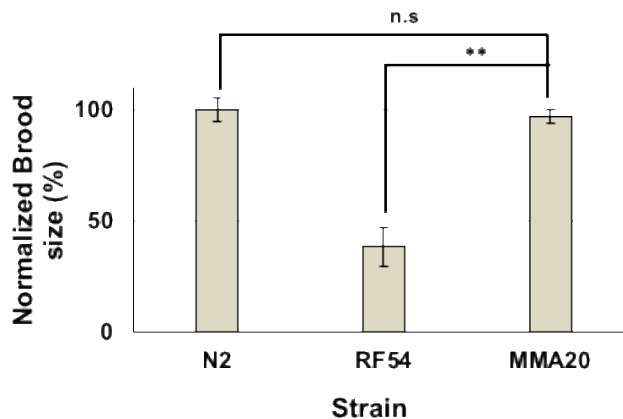
Figure 3.7 - Results from a brood size assay testing fertility. The loss of fertility seen in the alg-1 knockout strain RF54 is rescued in the MMA20 strain which expresses our GFP tagged transgenic ALG-1, suggesting that our ALG-1 pull-down construct is functional and mimic endogenous alg-1 activity. Student t-test. ** $p<0.01$. Error bars indicate the standard error of the mean.

We then used our strains to perform tissue-specific RNA immunoprecipitations. Each tissue-specific ALG-1 IP and control IPs were performed in duplicate using biological replicates (total 6 sequencing runs). We obtained ~25M reads on average for each tissue, of which ~80% were successfully mapped to the *C. elegans* genome (WS250) (Figure 3.8). To maximize our success, we used very stringent filters to determine gene presence, using only the top 25-50% of genes mapped in each dataset (Figure 3.9) (Materials and Methods) [56, 191]. Our analysis resulted in 3,681 different protein-coding genes targeted by the miRISC using the endogenous *alg-1* promoter or in the intestine or body muscle.

There are only 27 validated *C. elegans* miRNA-target interactions with strong evidence reported in the miRNA target repository miR-TarBase v7, and our study confirmed 16 of these interactions (59%), which is threefold enrichment when compared to a random dataset of similar size ($p<0.05$, chi-square test) (Figure 3.10).

Sample		Sequencing reads		
		Total (#)	Mapped (#)	Mapped (%)
	alg-1p	23,277,135	18,292,992	78.6
Intestine	Experiment (*rep 1*)	31,424,437	23,241,554	73.9
	Replicate 2 (*rep 2*)	23,522,056	17,861,605	75.9
Body muscle	Experiment (*rep 1*)	24,092,664	21,327,095	88.5
	Replicate 2 (*rep 2*)	25,347,564	22,062,749	87.0
	N2	21,894,026	19,378,018	88.5

Figure 3.8 - Overview of the sequencing reads obtained for each immunoprecipitation. Each sample was sequenced in duplicate with biological replicates. We obtained at least 15M reads each dataset mapped to the *C. elegans* genome (WS250).
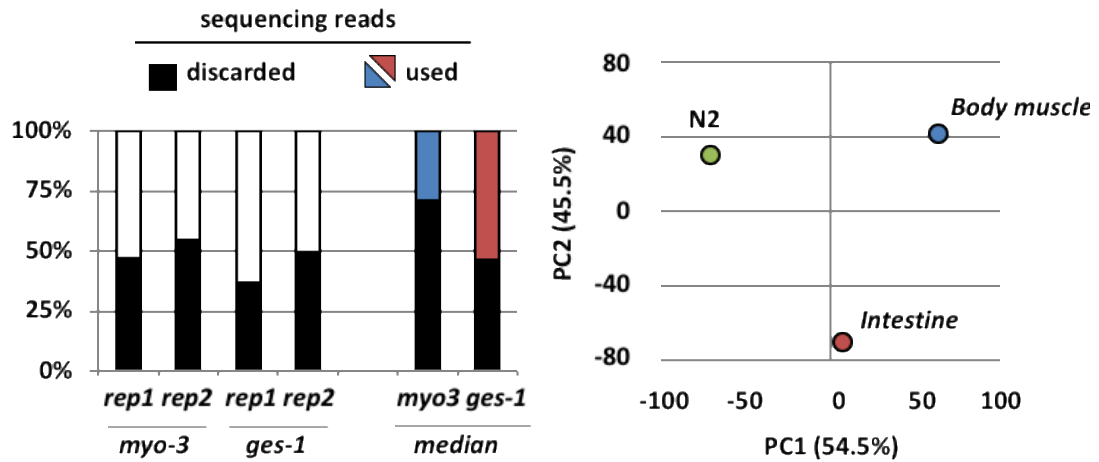
Figure 3.9 - In this study we have used only the top ~50-75% positive hits produced by the Cufflinks algorithm. Blue; Body muscle. Red; Intestine C) Principal Component Analysis (PCA) between the median FPKM values among replicates shows differences between tissues within our datasets. Each dot represents a median experiment and replicate.
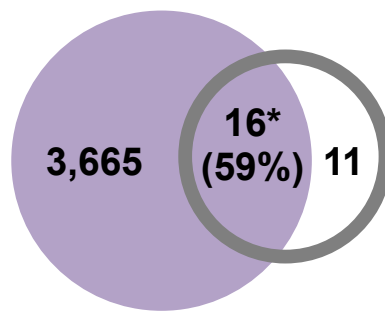
Figure 3.10 - Venn diagram showing the comparison of the genes identified in this study to miR-TarBase v7, a compendium of all experimentally validated miRNA targets. 59% of the genes in this database match those identified in this study. There is a 3-fold enrichment of miRNA targets, based on a Chi-squared test comparing to a random data-set of the same size *p<0.05.

When compared to genes present in the *C. elegans* intestine and body muscle transcriptomes [56], 81% of the intestine and 56% of the body muscle targets identified in this study match with their respective tissues (Figure 3.11). A comparison between our hits and a previously published ALG-1 IP dataset in all tissues also support our results (Figure 3.13) [106]. To further validate the quality of our hits, we used GFP-based approaches to confirm the tissue localization of a few tissue-specific genes identified in our study and found that except one, their observed localization matches the expected tissue (Figure 3.12). In addition, to further test the quality of our data, we compared our results with the intestine and body muscle-specific miRNA localization data from past studies [185] (Figure 3.13). We found that more than 84% of the genes identified in our study possess predicted binding sites in their 3'UTRs for miRNAs detected in each tissue, suggesting a strong correlation between our results and Alberti et al., 2018 [185] (Figure 3.14).
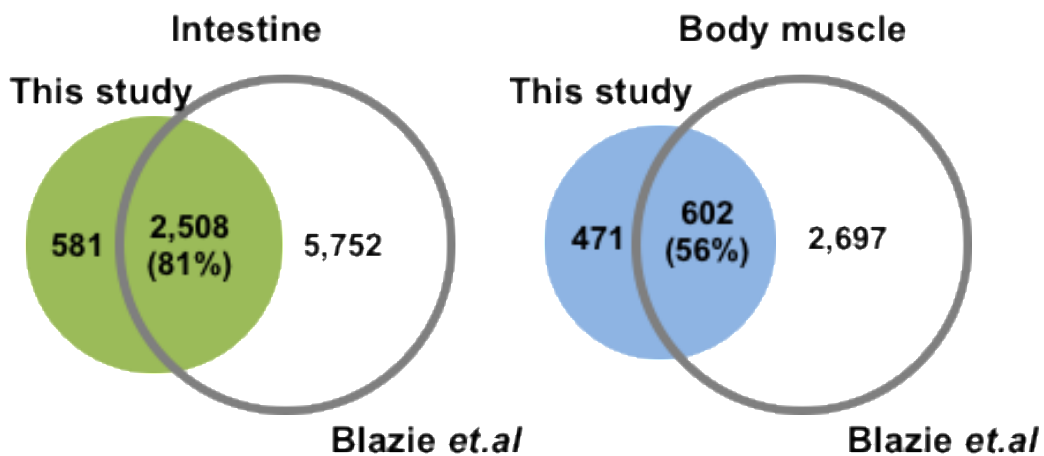
Figure 3.11 - Venn diagram showing the comparison of the genes identified in this study to previously published tissue-specific intestine and body muscle transcriptomes. The majority of the targets in both datasets were previously assigned to each tissue. Green – Intestine (*ges-1p*). Blue – Body muscle (*myo-3p*).
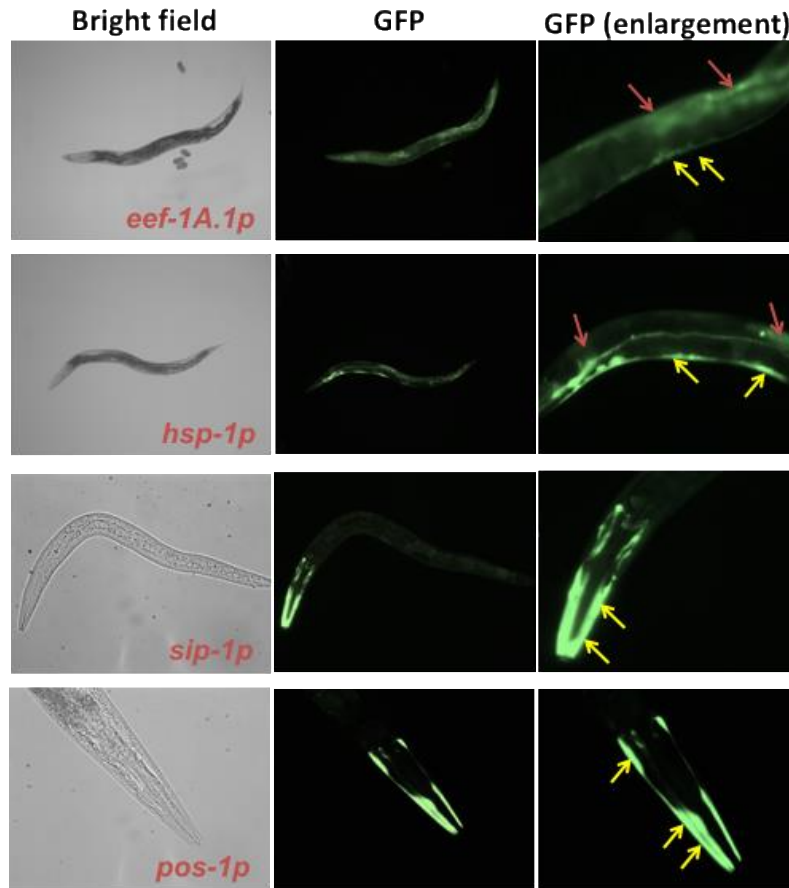
	Intestine		Body muscle	
Name	GFP Detected	Validated	GFP Detected	Validated
sip-1	**	yes	***	yes
eef-1A.1	***	yes	***	yes
pos-1	-	no	***	yes
hsp-1	***	yes	***	yes

Figure 3.12 - Validation of tissue localization selected hits detected in our tissue-specific ALG-1 pull-down. We have cloned selected promoters from genes identified in our tissue-specific ALG-1 pull-down and fused it to GFP reporters to validate the expression localization pattern. Except for one case, we detected strong GFP expression in the expected tissues. The red arrow marks intestine expression, the yellow arrow marks body muscle expression. Putative expression index:
 - not detected; * low expression; ** expressed; *** strong expression

59
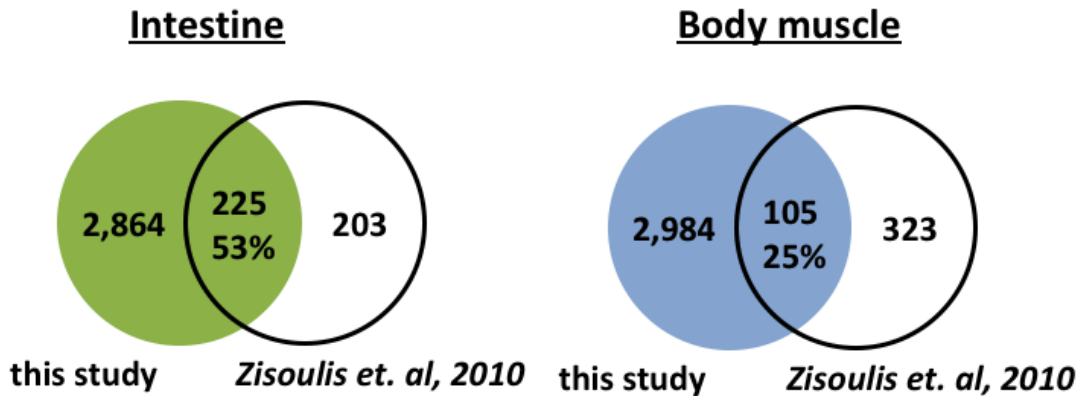
Figure 3.13 - Comparison of our tissue-specific ALG-1 pull-down to previously identified whole worm ALG-1 targets (Zisoulis et al., 2010). Venn diagram showing the comparison of hits between our datasets and a whole-worm RNA pull-down of ALG-1 targets.
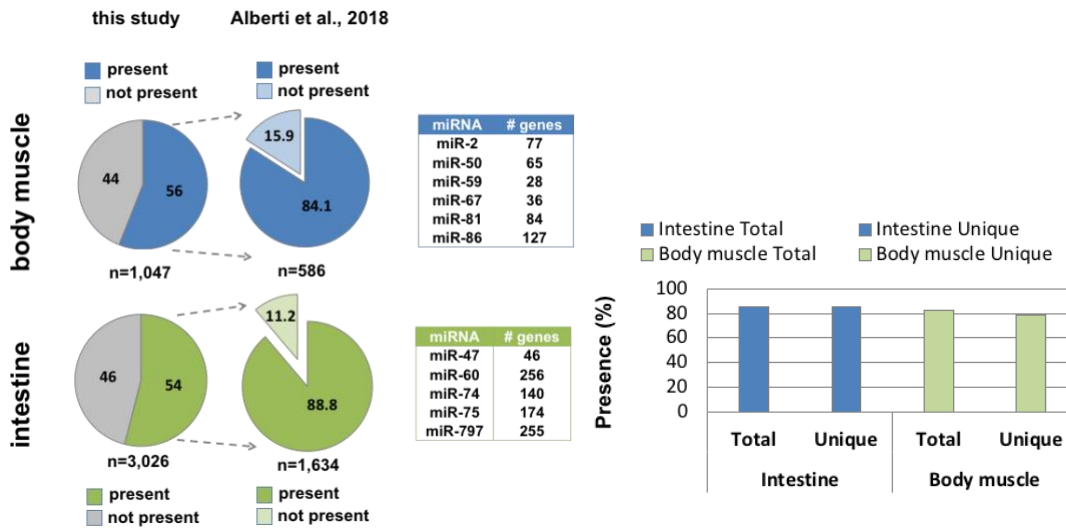
Figure 3.14 - Comparative analysis of our tissue-specific ALG-1 pull-down to previously identified intestine and body muscle miRNA (mime-seq) (Alberti et al., 2018). The left pie chart shows the percentage of genes identified in our study with predicted miRNA binding sites as for MiRanda algorithm. The right pie chart shows how many of those genes contain a seed region for intestine or body muscle miRNAs identified in Alberti et al., 2018. The tables show how many genes identified in our study are targeted by unique tissue-specific miRNAs identified in Alberti et al., 2018. The bar charts show the percentage of total and tissue-restricted (unique) intestine or body muscle ALG-1 targeted genes identified in our study with seed regions for miRNAs identified in Alberti et al., 2018

ALG-1 targets in the intestine regulate key metabolic enzymes

The *C. elegans* intestine is composed of 20 cells that begin differentiation early in embryogenesis and derive from a single blastomere at the 8-cell stage [186]. As the primary role of the intestine is to facilitate the digestion and the absorption of nutrients, many highly expressed genes in this tissue are digestive enzymes, ion transport channels and, regulators of vesicle transport [186].

In our intestinal ALG-1 pull-down, we identified 3,089 protein-coding genes targeted by miRNAs. 2,367 of these genes were uniquely targeted by miRNAs in this tissue (Figure 3.15). As expected, and consistent with the function of the intestine, we find many enzymes involved with glucose metabolism, such as *enol-1* an enolase, *ipgm-1* a phosphoglycerate mutase, and 3 out of 4 glyceraldehyde-3-phosphate dehydrogenases (*gpd-1*, *gpd-2,* and *gpd-4*). The human orthologue of the *C. elegans* gene *enol-1*, *ENO1* has been previously identified as a target of *miR-22* in the context of human gastric cancer [192]. In addition, some of our top hits are the fatty acid desaturase enzymes *fat-1*, *fat-2*, *fat-4* and *fat-6*, which are all involved with fatty acid metabolism, suggesting that these metabolic pathways are subjected to a high degree of regulation in the intestine. All of these genes contain seed elements in their 3'UTRs. Additionally, we find 5 out of 6 vitellogenin genes (*vit-1*, *vit-2*, *vit-3*, *vit-5* and *vit-6*) strongly targeted by miRNAs, with *vit-2* and *vit-6* being the most abundant transcripts in our immunoprecipitation. *vit-2* was shown to be targeted by ALG-1 in a previous study [183], and both possess MiRanda [99, 101] and/or PicTar [96] predicted

binding sites . These vitellogenin genes produce yolk proteins and are energy carrier molecules synthesized in the intestine. These yolk proteins are then transported to the gonads and into the oocytes to act as an energy source for the developing embryos [193]. Accordingly, we also find some RAB family proteins that are responsible for intracellular vesicular transport (*rab-1*, *rab-6.1*, *rab-7*, *rab-8*, *rab-21*, *rab-35* and *rab-39*).

Several transcription factors were also identified as miRNA targets in the intestine. *skn-1* is a bZip transcription factor that is initially required for the specification of cell identity in early embryogenesis, and then later plays a role in modulating insulin response in the intestine of adult worms [194]. This gene has already been found to be targeted by miRNA in many past studies [106, 183] and contains many predicted miRNA binding sites and seed regions from both MiRanda [99, 101] and PicTar [96] prediction software. A second transcription factor *pha-4* is expressed in the intestine, where it affects dietary restriction mediated longevity [195]. *pha-4* is a validated target of *let-7* in the intestine[196], and along with *skn-1*, is also targeted by *miR-228* [195]. Additionally, *pha-4* is targeted by *miR-71* [195].

We also find as a target of miRNA, *die-1* a gene which associated with the attachment of the intestine to the pharynx and the rectum [197], and the chromatin remodeling factor *lss-4* (let seven suppressor), which is able to prevent the lethal phenotype induced by knocking out the miRNA *let-7* [196]. Others have also validated these miRNAs as targets [196].

In *C. elegans* the intestine produces an innate immune response to pathogens. The genes *atf-7*, *pmk-1* and *sek-1* were all identified as targets of miRNAs in this tissue. These three genes act together to produce a transcriptional innate immune response where the transcription factor *atf-7* is activated through phosphorylation by kinases *pmk-1* and *sek-1*. Consistent with our findings, the role of miRNAs in regulating the innate immune response through the intestine and these genes has been reported in multiple studies [183, 198, 199].

Muscle ALG-1 targets modulate locomotion and cellular architecture

C. elegans possess 95 striated body wall muscle cells, which are essential for locomotion [187]. Its sarcomeres are composed of thick filaments containing myosin associated with an M-line, and thin filaments containing actin associated with the dense body. The pulling of actin filaments by myosin heads generates a force that produces locomotion [200].

Our ALG-1 pull-down identified 1,047 protein-coding genes targeted by miRNAs in the body muscle tissue (Figure 3.15). Within this group, 348 genes were not present in our intestine dataset, and are specifically restricted to the body muscle tissue. Our top hits include genes involved in locomotion, and general DNA maintenance (*grd-5*, *gcc-1*, *gop-2*, etc.) and several with unknown function. Consistent with muscle functions, we detected *mup-2*, which encodes the muscle contractile protein troponin T, *myo-3*, which encodes an isoform of the myosin heavy chain, *dlc-1*, which encodes dynein light chain 1 and F22B5.10, a poorly

characterized gene involved in striated muscle myosin thick filament assembly. *mup-2*, *myo-3* and *dlc-1* were all found to be targeted by ALG-1 in previous studies [106, 183]. Consistent with muscle function, a GO term analysis of this dataset highlights an enrichment of genes involved in locomotion, suggesting a potential role for miRNAs in this biological process.

We also identified numerous actin gene isoforms (*act-1*, *act-2*, *act-3* and *act-4*), which are required for maintenance of cellular architecture within the body wall muscle, and the Rho GTPase *rho-1*, which is required for regulation of actin filament-based processes including embryonic polarity, cell migration, cell shape changes, and muscle contraction. Small GTPase are a gene class heavily targeted by miRNAs [201, 202]. The human ortholog of *rho-1* is a known target for *miR-31*, *miR-133*, *miR-155* and *miR-185* [202].

Importantly, we also found several muscle-specific transcription factors including *mxl-3*, a basic helix-loop-helix transcription factor, and K08D12.3, an ortholog of the human gene *ZNF9*. These genes are known to regulate proper muscle formation and cell growth. *mxl-3* is targeted by *miR-34* in the context of stress response [203]. Both genes have been detected in past ALG-1 immunoprecipitation studies [106].

Our top hit in this tissue is the zinc finger CCCH-type antiviral gene *pos-1*, a maternally inherited gene necessary for proper fate specification of germ cells, intestine, pharynx, and hypodermis[204]. *pos-1* contains several predicted miRNA binding sites in its 3'UTR, and based on our GFP reporter validation study is strongly expressed in the

body muscle (Figure 3.13). We also find the KH domain containing protein *gld-1*, the

homolog of the human gene *QKI*, which is targeted by *miR-214* [205], *miR-200c* and

miR-375 [206].

miRNA targeting is more extensive in the intestine than it is in the body muscle

By comparing the percentage of tissue-specific miRNA targets identified in our

study to the previously published intestine and body muscle transcriptomes [56, 191], we

found that the hits in the intestine are almost twice the number of hits we obtained in the

body muscle tissue (30.3% vs 18.2%) (Figure 3.15).

The length of the 3`UTRs of genes identified as miRNA targets in the intestine

and the body muscle tissues are similar when comparing the two tissues, but are on

average longer and have more predicted miRNA binding sites than the overall *C. elegans*

transcriptome (Figure 3.16). Our results indicate that despite the similarity in average

3'UTR length in tissues. In this specific case, we found that the intestine utilizes

miRNA-based gene regulation to a greater extent when compared with the body muscle.

Biological Process	FE	p-value
cytoskeleton organization	11.54	6.98E-07
chromatin organization	6.92	9.11E-04
organelle organization	5.28	7.09E-08

Biological Process	FE	p-value
cell growth	46.75	6.14E-05
locomotion	14.07	3.86E-05
cellular component movement	7.69	1.59E-04
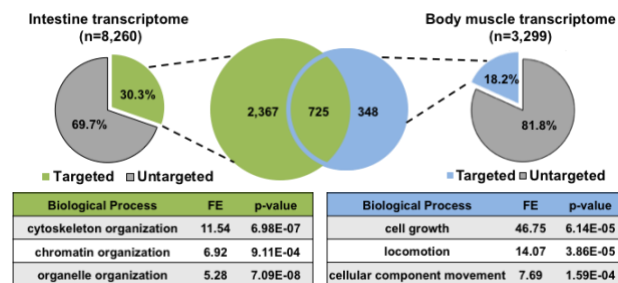
Figure 3.15 – Unique miRNA targets in each tissue correspond to tissue-specific functions. Pie chart showing the proportion of miRNA targets detected in this study compared to tissue-specific transcriptomes previously characterized by Blazie et.al 2017. The Venn diagram in the center shows the number of protein-coding genes identified in this study as miRNA targets between the intestine and body muscle. The tables show a Gene Ontology analysis for pathway enrichment using the top 100 genes from each dataset used in this study. Green – Intestine (*ges-1p*). Blue – Body muscle (*myo-3p*).

65
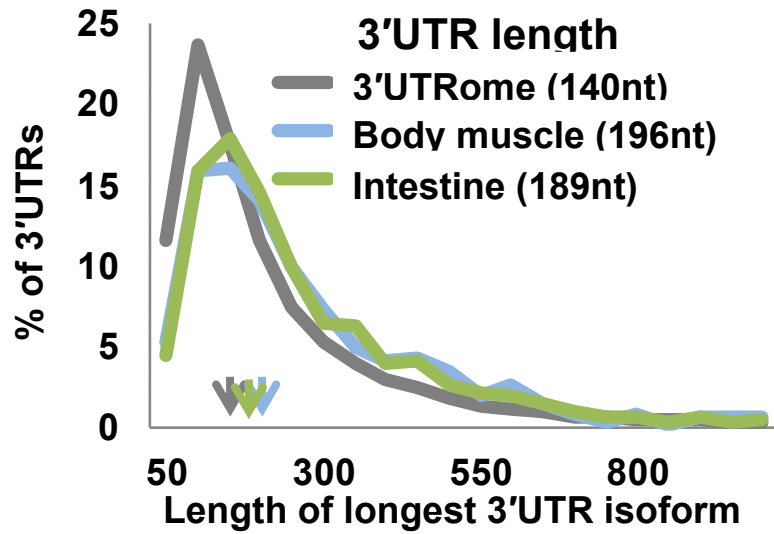
Figure 3.16 - The length of 3'UTRs from protein-coding genes as from the 3'UTRome v1 (MANGONE et al. 2010) compared to the intestine and body muscle targets identified in this study. The arrows indicate the median 3`UTR length. Genes targeted in the intestine and body muscle have longer 3`UTRs on average than those published in the *C. elegans* 3`UTRome v1.
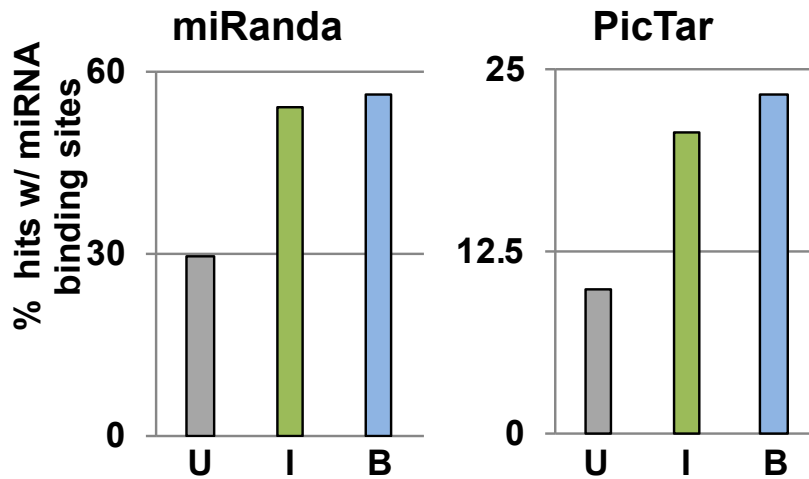
**Figure 3.17 - Proportion of 3`UTRs with predicted miRNA binding sites** miRanda and PicTar (LALL et al. 2006; BETEL et al. 2008) in the 3`UTRome (gray), when compared to the 3'UTRs of genes identified in this study. Green – Intestine; Blue – Body muscle. The genes identified in this study in the intestine and body muscle are enriched for predicted miRNA binding sites.

MiRNA targets in the intestine and body muscle are enriched for miR-355 and miR-85 binding sites

Bioinformatic analysis of the longest 3`UTR isoforms of the targeted genes showed there was no specific requirement for the seed regions in either tissue (Figure 3.18). However, the use of predictive software showed that in addition to others, there is an intestine-specific bias for *miR-355* targets (Figure 3.18). This miRNA is involved in the insulin signaling and innate immunity [207], which in *C. elegans* are both mediated through the intestine.

In contrast, we observed an enrichment of targets for the poorly characterized *miR-85* in the body muscle dataset (Figure 3.18). These two miRNAs are uniquely expressed in the respective tissues [208]. While *miR-85* and *miR-355* were the most abundant and tissue-restricted miRNAs identified in this study, several other miRNAs, including *miR-71, miR-86, miR-785* and *miR-792* were also found highly expressed but less spatially restricted.
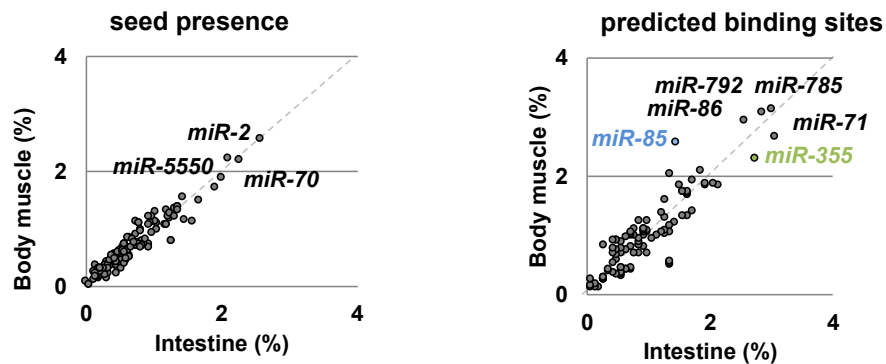
Figure 3.18 - Analysis of miRNA target sites identified in this study. The two-axis show the proportion of 3'UTRs with perfect seeds or with predicted target sites (miRanda) (BETEL et al. 2008), normalized to the total number of genes targeted in each tissue for each miRNA. miRNAs that target more than 2% of the genes are listed. The blue mark denotes *miR-85*, a body muscle-specific miRNA. The green mark denotes *miR-355*, an intestine-specific miRNA.

Discussion

We have developed tools to identify in vivo tissue-specific targets of miRNAs in C. elegans

In this study, we have developed tools and techniques to identify tissue-specific miRNA targets and applied them to uniquely define the genes targeted by miRNAs in the *C. elegans* intestine and body muscle. We validated previous findings and mapped ~3,000 novel tissue-specific interactions (Figure 3.10-3.12).

To perform these experiments, we have prepared worm strains expressing ALG-1 fused to GFP and expressed this cassette in the intestine and body muscle using tissue-specific promoters. We validated the ALG-1 expression (Figure 3.3-3.5), and the viability of our ALG-1 construct in *in vivo* studies (Figure 3.6 and 3.7). We have then performed ALG-1 immunoprecipitations in duplicate, separated the miRNA complex from their targets, and sequenced the resultant RNA using Illumina sequencing (Figure 3.2 and 3.8). To confirm our results, we validated a few selected hits with expression localization studies in both tissues (Figure 3.13). Importantly, our ALG-1 pull-down results agree with previous studies (Figure 3.10-3.12, and 3.14), and are significantly enriched with predicted miRNA targets (Figure 3.17).

We have uploaded our intestine and body muscle miRNA target datasets into the 3'UTRome database (www.UTRome.org), which is the publicly available resource for the *C. elegans* community interested in 3'UTR biology [40, 181]. In

order to provide a more comprehensive overview, we have also manually curated

and included results from several available datasets including PicTar [96] and

TargetScan [95] miRNA target predictions, experimentally validated ALG-1

interaction [106, 183], tissue-specific gene expression and expanded 3'UTR

isoform annotation data [56, 75, 191].

The level of miRNA-based gene regulation is heterogeneous across tissues.

      The genes identified in this study overall match with the intestine

transcriptome previously published by our group (81%) [56, 191]. Of note, only

56% of genes identified as miRNA targets in the body muscle match the body

muscle transcriptome (Figure 3.15). Perhaps, the remaining targets are genes

strongly down-regulated by miRNAs in this tissue, leading to rapid deadenylation

and mRNA degradation that make them undetectable using our PAB-1-based

pull-down approach. Given the fact the body muscle transcriptome is significantly

smaller than the intestine transcriptome, it may also be subjected to less regulation

through miRNA. However, if we normalize the number of genes expressed in

each tissue and study the proportion of the transcriptome targeted by miRNA, we

still find significantly more regulation in the intestine (Figure 3.15), suggesting

that this tissue may indeed employ miRNA-based gene regulation to a greater

extent.

MiRNAs regulate tissue-specific targets to contribute to homeostasis.

The majority of the targeted genes in our intestine pull-down are unique to the intestine and share only a handful of genes with our body muscle dataset (725 genes, 23% of the total intestine dataset) (Figure 3.15). Many of the genes uniquely targeted by miRNAs in each tissue correspond to the functions of that tissue (Figure 3.15). The intestine of the worm is involved with a variety of functions throughout its life. From hatching, the intestine is responsible for metabolism, vesicle transport, and mounting innate immune response but upon reaching the adult stage, the intestine also plays an important role in regulating fertility. Accordingly, we find genes uniquely targeted by miRNAs in the intestine that are a part of each of these pathways. In the case of the body muscle, very few genes are uniquely targeted, but yet show an enrichment for genes involved with locomotion and cellular architecture. The importance of miRNAs in development is well established, however this pattern of targeting genes that are important for normal tissue function suggests that miRNAs also contributes to the maintenance of these tissues. The small pool of shared genes between both tissues includes housekeeping genes that are most likely regulated similarly in both tissues. Of note, this minimal overlap between our tissue-specific datasets indicates that our ALG-1 pull-down is indeed tissue-specific with marginal cross-contamination.

Precise tissue-specific miRNA targeting relies on principles outside of target site availability

We hypothesized that the binding sites for tissue-specific miRNAs would be more abundant in the 3'UTRs of tissue-specific targets. To test this hypothesis, we performed a bioinformatic analysis of predicted miRNA binding sites and seed availability in the longest 3'UTR isoforms of genes in our ALG-1 IPs. We found binding sites for tissue-specific miRNAs in the 3'UTRs of our targets, which is in agreement with miRNA localization datasets [185] (Figure 3.14). However, the binding sites for these miRNAs are not more abundant than the binding sites of other miRNAs.

A majority of miRNAs are predicted to target less than 2% of the 3'UTRs identified in our IPs (Figure 3.18), which is intriguing as it suggests that the availability of predicted miRNA binding sites and seed regions are poor indicators of targeting by miRNAs, and indicates that mechanisms such as APA and tissue-specific miRNA expression patterns modulate precise miRNA-based gene regulation.

The exceptions to this observation are the miRNAs *miR-71, miR-86, miR-785, miR-792, miR-85* and *miR-355,* which have predicted binding sites in more than 2% of the 3'UTRs of the targets isolated from each tissue. Very little is known about most of these miRNAs. However, *miR-355* is uniquely expressed in the intestine and *miR-85* is uniquely expressed in the body muscle and both

miRNAs show a slight enrichment of binding sites of the targets from the corresponding tissue.

Our experimental approach was designed for tissue-specific mRNA target identification and unfortunately did not provide miRNA data. We assign tissue-specific targets to miRNAs relying primarily on prediction software and correlation to past-published datasets. These comparative approaches required conversion between genomic releases and data consolidation across different developmental stages and conditions, which may have added unwanted variability to our comparative analysis. Overall, the results from this study highlight 1) A method for the *in vivo* identification of tissue-specific targets of miRNA. 2) The tissue-specific roles of miRNAs in regulating gene expression, especially in regards to tissue development and homeostasis. 3) Provides insights on the mechanisms that influence tissue-specific miRNA targeting.

Experimental

Preparing MosSCI vectors for generating GFP::ALG-1 strains

The strains used for the ALG-1 pull-down were prepared using a modified version of the previously published polyA-pull construct [56, 191]. We produced a second-position Entry Gateway vector containing the genomic sequence of *alg-1* tagged at its N-terminus with the GFP fluorochrome. Briefly, we designed primers flanking the coding sequence of *alg-1* and performed a Polymerase Chain

72

Reaction (PCR) amplification to clone the *alg-1* locus from genomic DNA extracted from N2 *wt* worms (primer 1 and 2 in Table 3.1). The resulting PCR product was analyzed on a 1% agarose gel, which displayed a unique expected band at ~3,500 nucleotides. This band was then isolated using the QIAquick Gel Extraction Kit (QIAGEN, cat. 28704) according to the manufacturer's protocol. Upon recovery, we digested the purified PCR product with the restriction enzymes SacI and BamHI and then cloned it into the modified polyA-pull construct [56, 191], replacing the gene *pab-1*. The ligation reaction was performed using the NEB Quick Ligation Kit (cat. MS2200S) according to the manufacturer's protocol. We used the QuikChange II Site-Directed Mutagenesis Kit (Agilent, cat. 200523) to remove the unnecessary C-terminal 3xFLAG tag from the polyA-pull vector (primers 3 and 4 in Table 3.1). We then cloned the previously described endogenous *alg-1* promoter [209] by designing primers to add Gateway BP cloning elements and then performed PCR using N2 *wt* genomic DNA as a template (primers 5 and 6 in Table 3.1). Using the resulting PCR product, we performed a Gateway BP cloning reaction into the pDONR P4P1R vector (Invitrogen) according to the manufacturer's protocol. To assemble the final injection clones, we performed several Gateway LR Clonase II plus reactions (Invitrogen, cat. 12538-013) using the destination vector CFJ150 [188], the tissue-specific or endogenous promoters (*alg-1* for endogenous, *ges-1* for the intestine and *myo-3* for the body muscle), the *gfp* tagged *alg-1* coding sequence, and the *unc-54* 3'UTR as previously published[56].

Microinjections and screening of transgenic C. elegans strains

To prepare single copy integrated transgenic strains we used the *C. elegans* strain Eg6699 [ttTi5605 II; unc-119(ed3) III; oxEx1578][188], which is designed for MosI mediated single copy integration (MosSCI) insertion, using standard injection techniques. These strains were synchronized by bleaching[210], then grown at 20ºC for 3 days to produce young adult (YA) worms. YA worms were then picked and subjected to microinjection using a plasmid mix containing; pCFJ601 (50ng/µl), pMA122 (10ng/µl), pGH8 (10ng/µl), pCFJ90 (2.5ng/µl), pCFJ104 (5ng/µl) and the transgene (22.5ng/µl)[211]. Three injected worms were isolated and individually placed into single small nematode growth media (NGM) plates (USA Scientific, cat 8609-0160) seeded with OP50-1 and were allowed to grow and produce progeny until the worms had exhausted their food supply. The plates were then screened for progenies that exhibited wild type movement and proper GFP expression, and single worms exhibiting both markers were picked and placed onto separate plates to lay eggs overnight. In order to select for single copy integrated worms, an additional screen was performed to select for worms that lost the mCherry fluorochrome expression (extrachromosomal injection markers).

Genotyping of transgenic C. elegans strains

Single adult worms were isolated and allowed to lay eggs overnight and then genotyped for single copy integration of the transgene by single worm PCR

as previously described [98] (primers 7-9 in Table 3.1). Progeny from worms that contained the single copy integrations were propagated and used for this study. A complete list of worm strains produced in this study is shown in Table 3.2.

Validating expression of the transgenic construct

To validate the expression of our transgenic construct, and to evaluate our ability to immunoprecipitate GFP tagged ALG-1, we performed immunoprecipitation (as described below) followed by a western blot. For the western blot we used a primary anti-GFP antibody (Novus, NB600-303) (1:2000) and a fluorescent secondary antibody (LI-COR, 925-32211)(1:5000), followed by imaging using the ODYSSEY CLX system (LI-COR Biosciences, NE) (Figure 3.5).

In vivo validation of GFP::ALG-1 functionality by brood size assay

In order to validate the *in vivo* functionality of our transgenic GFP tagged ALG-1, we used a genetic approach. It was previously shown that the knock out *alg-1* strain RF54 [*alg-1(*gk214) X] lead to a decrease in fertility [190]. We rescued this decrease in fertility in the *alg-1* knockout strain RF54[*alg-1(*gk214) X] by crossing it into our strain MMA17 (Table 3.2), which expresses our GFP tagged transgenic ALG-1, driven by the endogenous *alg-1* promoter. The resulting strain MMA20 [*alg-1*(*gk214*)X; *alg-1p*::*gfp*::*alg-1*::*unc-54* II] only expresses our cloned *alg-1* gene tagged with the GFP fluorochrome. We validated the genotype of MMA20 using single worm PCRs as previously described[98]

75

(primers 10 and 11 in Table 3.1 and Figure 3.6). The brood size assay was used to evaluate the ability of our transgenic GFP tagged ALG-1 construct to rescue the loss in fertility seen in the *alg-1* knockout strain (RF54). The brood size assay was performed by first synchronizing N2 (*wt*), RF54 and MMA20 strains to arrested L1 larvae, through bleaching followed by starvation overnight in M9 solution. We then plated the L1 arrested worms on NGM plates seeded with OP50-1 and allowed the worms to develop to the adult stage for 48 hours after which single worms were isolated onto OP50-1 seeded plates. The adult worms were left to lay eggs overnight (16 hours) after which the adult worms were removed. The eggs were allowed to hatch and develop for 24 hours, and the number of larvae in each plate was counted.

Sample preparation and crosslinking

0.5ml of mixed stage *C. elegans* of each strain was grown on five large 20 cm plates (USA Scientific, cat 8609-0215) and harvested by centrifugation at 400rcf for 3 minutes. The pellets were initially washed in 15ml dH_2O water and spun down at 400 rcf for 3 minutes and then resuspended in 10ml of M9+0.1%Tween20 and then cross-linked 3 times on ice, with energy setting: 3000 x 100 $\mu J/cm^2$ (3kJ/m²) (Stratalinker 2400, Stratagene)[212]. After the crosslinking, each *C. elegans* strain was recovered by centrifugation at 400 rcf for 3 minutes, and resuspended in two volumes (1ml) of lysis buffer (150mM NaCl, 25mM HEPES(NaOH) pH 7.4, 0.2mM DTT, 10% Glycerol, 25 units/ml of

RNasin® Ribonuclease Inhibitor (Promega, cat N2611), 1% Triton X-100 and 1 tablet of protease inhibitor for every 10ml of lysis buffer (Roche cOmplete ULTRA Tablets, Sigma, cat 5892791001). The lysed samples were subjected to sonication using the following settings: amplitude 40%; 5x with 10sec pulses; 50sec rest between pulses (Q55 Sonicator, Qsonica). After the sonication, the cell lysate was cleared of debris by centrifugation at 21,000rcf at 4°C for 15 min and the supernatants were then transferred to new tubes.

GFP-TRAP bead preparation and immunoprecipitation

25µl of GFP-TRAP beads (Chromotek, gtma-10) (total binding capacity 7.5µg) per immunoprecipitation were resuspended by gently vortexing for 30 seconds, and washed three times with 500µl of cold Dilution/Wash buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA). The beads were then resuspended in 100µl/per IP of Dilution/Wash buffer. 100µl of resuspended beads were then incubated with 0.5ml of lysate for 1 hour on the rotisserie at 4°C. At the completion of the incubation step, the beads were collected using magnets. The unbound lysate was saved for PAGE analysis. The beads containing the immunoprecipitatred *alg-1* associated to the target mRNAs were then washed three times in 200µl of Dilution/Wash buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA), and then the RNA/protein complex was eluted using 200µl of Trizol (Invitrogen, cat 15596026) and incubated for 10 minutes at room temperature.

Trizol/Driectzol RNA purification

The RNA purification was performed using the RNA MiniPrep kit (Zymo Research, cat ZR2070) as per the manufacturer's protocol. All centrifugation steps were performed at 21,000g for 30 seconds. We added an equal volume of ethanol (95-100%) to each sample in Trizol and mixed thoroughly by vortexing (5 seconds, level 10). The samples were then centrifuged, recovered using a magnet, and the supernatant was transferred into a Zymo-Spin IIC Column in a Collection Tube and centrifuged. The columns were then transferred into a new collection tube and the flow through were discarded. 400 µl of RNA wash buffer was added into each column and centrifuged. In a separate RNase-free tube, we added 5 µl DNase I (6 U/µl) and 75 µl DNA Digestion Buffer, mixed and incubated at room temperature (20-30°C) for 15 minutes. 400 µl of Direct-zol RNA PreWash (Zymo Research, cat ZR2070) was added to each sample and centrifuged twice. The flow-through was discarded in each step. 700 µl of RNA wash buffer was then added to each column and centrifuged for 2 minutes to ensure complete removal of the wash buffer. The columns were then transferred into RNase-free tubes, and the RNAs were eluted with 30 µl of DNase/RNase-Free Water added directly to the column matrix and centrifuging.

cDNA library preparation and sequencing

Each cDNA library was prepared using a minimum of 500pg of immunoprecipitated RNA from each tissue. The total RNA was reverse transcribed using

the IntegenX's (Pleasanton, CA) automated Apollo 324 robotic preparation system using

previously optimized conditions[191]. The cDNA synthesis was performed using a SPIA

(Single Primer Isothermal Amplification) kit (IntegenX and NuGEN, San Carlos,

CA)[213]. The cDNA was then sheared to approximately 300 bp fragments using the

Covaris S220 system (Covaris, Woburn, MA). We used the Agilent 4200 TapeStation

instrument (Agilent, Santa Clara, CA) to quantify the abundance of cDNAs and calculate

the appropriate amount of cDNA necessary for library construction. Tissue-specific

barcodes were then added to each cDNA library, and the finalized samples were pooled

and sequenced using the HiSeq platform (Illumina, San Diego, CA) with a 1x75bp HiSeq

run.

Data analysis

We obtained ~15M unique reads per sample (~130M reads total). The

software Bowtie 2 [214] run using default parameters was used to perform the

alignments to the *C. elegans* genome WS250. We used custom Perl scripts and

Cufflinks [215] algorithm to study the differential gene expression between our

samples. A summary of the results is shown in (Figure 3.8-3.9). Mapped reads

were further converted into a bam format and sorted using SAMtools software run

with generic parameters [216], and used to calculate Fragments Per Kilobase

Million (FPKM) values, as an estimate of the abundance of each gene per sample.

We used an FPKM $\geq$ 1 on the median from each replicate as a threshold for

identifying positive hits. This stringent approach discarded ~50-75% of mapped

reads for each sample (Figure 3.9). The quality of our finalized list of target genes

was tested using a principal component analysis versus our N2 *wt* negative control

(Figure 3.9).

Molecular cloning and assembly of the expression constructs

   The promoters of candidate genes were extracted from genomic DNA

using genomic PCR and cloned into Gateway-compatible entry vectors

(Invitrogen). We designed Gateway-compatible primers (primers 12-19 in Table

3.1) targeting 2,000 bp upstream of a given transcription start site, or up to the

closest gene. Using these DNA primers, we performed PCRs on *C. elegans*

genomic DNA, amplified these regions, and analyzed the PCR products by gel

electrophoresis. Successful DNA amplicons were then recombined into the

Gateway entry vector pDONR P4P1R using Gateway BP Clonase reactions

(Invitrogen). The reporter construct pAPAreg has been previously described in

Blazie et al., 2017 [56]. The coding region of this construct was prepared by

joining the coding sequence of the mCherry fluorochrome to the SL2 splicing

element found between the *gpd-2* and *gpd-3* genes, and to the coding sequence of

the GFP gene. The entire cassette was then PCR amplified with Gateway-

compatible primers and cloned into pDONR P221 by Gateway BP Clonase

reactions (Invitrogen).

Bioinformatic analysis of tissue-specific miRNA targeting biases

The tissue-specific miRNA studies were performed in two steps. First, we utilized custom-made Perl scripts to scan across the longest 3'UTR of each *C. elegans* protein-coding gene (WS250) in our datasets, searching for perfect sequence complementarity to the seed regions of all *C. elegans* miRNAs present in the miRBase database (release 21) [63, 218-221]. This result was then used to calculate the percentage of seed presence in the intestine and body muscle datasets. To calculate the percentage of predicted targets, we extracted both predicted target genes, and their target miRNA name from the miRanda database [101] and compared the results with our study.

Comparison with other datasets

We extracted the WormBase IDs of genes in the intestine and body muscle transcriptomes previously published by our group [56], and most abundant miRNA targets (transcript names) identified by Kudlow et al., 2012 in these tissues [183]. We then translated the transcript names from Kudlow et al., 2012 into WormBase IDs using custom Perl scripts, and compared how many genes in each of these groups overlap with our ALG-1 pull-downs. The results are shown in Figure 3.13. For the analysis shown in Figure 3.14 we extracted the names of the miRNAs previously identified by Alberti et al., 2018 in the *C. elegans* intestine and body muscle tissues [185]. We then used custom Perl scripts to search for the presence of the seed regions of these miRNAs in the 3'UTRs of the genes identified in this study (Figure 3.1

Table 3.1 – Primers used in Chapter 3.

Primer #	Primer Name	Primer sequence
1	alg1Ext_fwd	AACGAGCTCTCCGGCGGGCCGCAATATTTGCCAGGAGTCATGAATTC
2	alg1Ext_rev	AACGGATCCAGCAAAGTACATGACGTTGTTGGCATCCGGATGAACC
3	3xFLAG del CALF F	GCCAACAACGTCATGTACTTTGCTTAACATGTGAATTCGTATAGATCTCACCCAGCTTTC
4	3xFLAG del CALF R	GAAAGCTGGGTGAGATCTATACGAATTCACATGTTAAGCAAAGTACATGACGTTGTTGGC
5	alg-1p_Fwd	GGGGACAACTTTGTATAGAAAAGTTG GCTAATTATTTCTAGGTACTATATATTTG
6	alg-1p_Rev	GGGGACTGCTTTTTTGTACAAACTTG TGGGTCGTTTGAGGCGACG
7	F_Universal	GAGGCAGAATGTGAACAAGACTCGAGC
8	R_Yes_N	CTTTGAGCCAATTCATCCCGG
9	R_No	CCACGCCCAGGAGAACACGTTAG
10	alg-1 start F	ATGTCCGGCGGGCCG
11	alg-1 exon2 R	CCAGGTAATTGTCCCGGTGC
12	attB4_Peef-1A.1_F	GGGGACAACTTTGTATAGAAAAGTTGCTTCCATTGTTTCTGTTAAATTAATGAAT
13	attB1r_Peef-1A.1_R	GGGGACTGCTTTTTTGTACAAACTTGTGGCTGCTACGGAGTGAGC
14	attB4_Phsp-1_F	GGGGACAACTTTGTATAGAAAAGTTGAAATTTCATTTCTTTTTTATAAATACTTCGGC
15	attB1r_Phsp-1_R	GGGGACTGCTTTTTTGTACAAACTTGTTTTTACTGTAAAAAATAATTTAAAAATCAAGAAAT
16	attB4_Ppos-1_F	GGGGACAACTTTGTATAGAAAAGTTGCCCAAATACCATTTTGAAACTGAATC
17	attB1r_Ppos-1_R	GGGGACTGCTTTTTTGTACAAACTTGTTTTGAATCTAGAAAAATAAATTTACAAACTG
18	attB4_Psip-1_F	GGGGACAACTTTGTATAGAAAAGTTGATTTTGCGTGCAATGCATTT
19	attB1r_Psip-1_R	GGGGACTGCTTTTTTGTACAAACTTGTAGTTGAGTGAATAGGGTTAAGAATGAGG

Table 3.2 – Worm strains generated in Chapter 3.

Strain Name	Strain Background	Injected transgenic construct
MMA17	EG6699 [*ttTi5605* II; *unc-119*(*ed3*) III; oxEx1578]	CFJ150 *alg-1p*::*gfp*::*alg-1*::*unc-54* 3'UTR
MMA18	EG6699 [*ttTi5605* II; *unc-119*(*ed3*) III; oxEx1578]	CFJ150 *ges-1p*::*gfp*::*alg-1*::*unc-54* 3'UTR
MMA19	EG6699 [*ttTi5605* II; *unc-119*(*ed3*) III; oxEx1578]	CFJ150 *myo-3p*::*gfp*::*alg-1*::*unc-54* 3'UTR
MMA20	RF54 [*alg-1*(*gk214*) X] + MMA17 [CFJ150 *alg-1p*::*gfp*::*alg-1*::*unc-54* 3'UTR II]	

Chapter 4

A potential novel role for miRNA in regulating tissue-specific alternative splicing

Publication note

The research reported in this chapter has been submitted for publication in

Genetics. Kasuen Kotagama, Anna L Schorr, Hannah S Steber, and Marco Mangone. All

co-authors have granted permission for this work to be included in this dissertation.

Overview

Splicing is a critical mechanism which determines the exons available for

translation and is catalyzed by a large RNA protein complex known as the spliceosome.

In higher eukaryotes, the major splicing complex consists of the U1, U2, and U4/U6

small nuclear ribonucleoproteins (snRNPs), which contain small nuclear RNAs

(snRNAs) to identify splice sites in mRNAs [222]. The functional spliceosome in cells

can include upwards of 80 proteins that interact with each other [223]. One category of

RBPs that associate with the spliceosome are known as splicing factors, which enhance

or decrease the usage of specific splice sites and contribute to a mechanism known as

alternative splicing (Figure 4.1) [222]. Alternative splicing leads to the production of

different protein isoforms using single mRNA precursors, fine-tune regulatory networks

and contribute to the acquisition of tissue identity and function. Alternative splicing occurs in >90% of multi-exon genes in humans [224, 225], and is one of the mechanisms by which the human genome with only ~20k genes can produce ~30k proteins [226]. This mechanism is required to ensure that each tissue possesses the correct gene expression pattern needed to thrive [227] and many aberrant alternative splicing events are linked to diseases [228, 229].
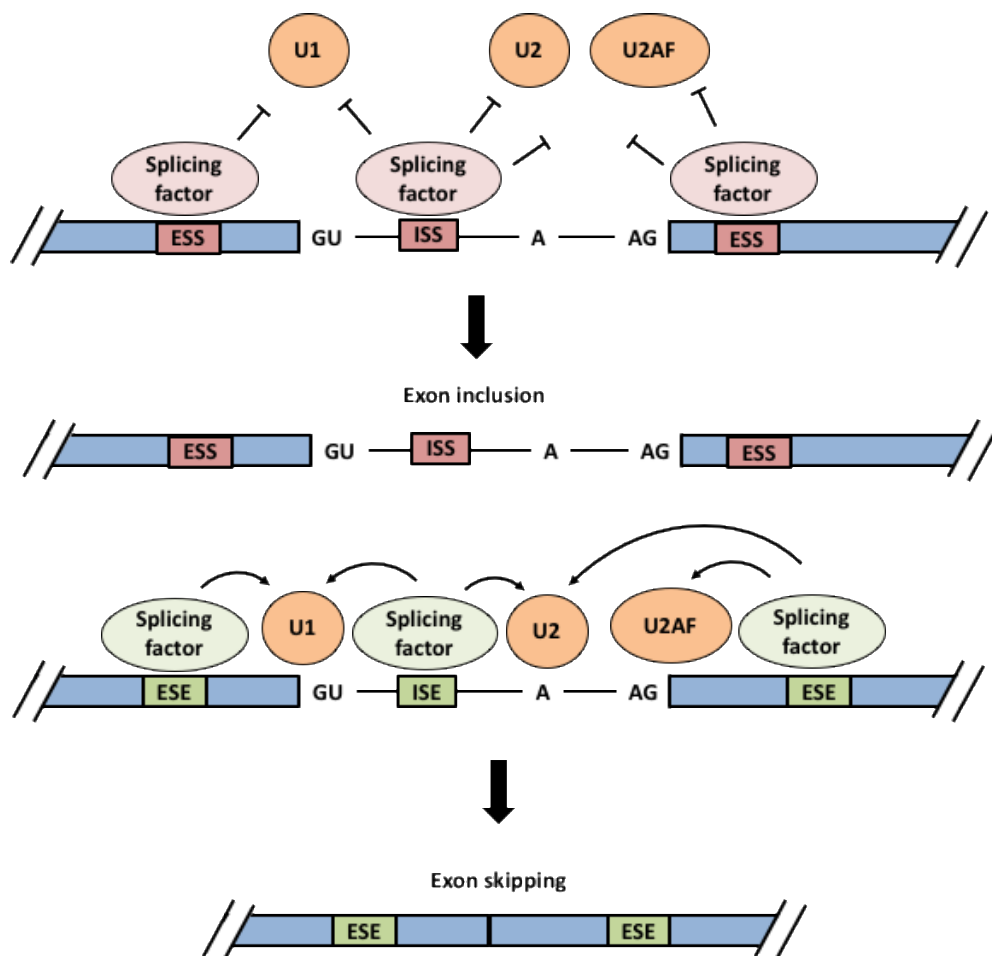
Figure 4.1 – Alternative splicing can be directed by splicing factors. Splicing factors bind to sequence elements in the mRNA transcripts (ESS – Exonic splice silencer, ISS – Intronic splice silencer, ISE – Intronic splice enhancer, Exonic splice enhancer) and promote or repress splice site usage leading to exon inclusion or exon skipping.

Tissue-specific alternative splicing has been observed in many multicellular organisms [227, 230-234]. However, many splicing factors that produce alternative splicing are transcribed with promoters that are expressed ubiquitously across tissues. It is believed that alternative splicing is regulated in a combinatorial manner by multiple interactions between splicing factors [235-237]. Splicing factors act by bind specific *cis*-acting elements located within exon/intron junctions in a combinatorial manner, promoting or inhibiting splicing. One class of splicing factor, Serine Arginine (SR) proteins recognize exon splicing enhancers (ESEs) and are important in promoting constitutive and alternative pre-mRNA splicing. While heterogeneous nuclear ribonucleoproteins (hnRNPs) are a large class of nuclear RBPs that bind exon splicing silencers (ESSs) and usually promote exon retention [238]. The relative expression levels of members from these two classes of splicing factors vary between tissues, and this imbalance is believed to promote the outcome of tissue-specific alternative splicing events [239, 240].

A study performed by Cáceres et al. in 1994 showed that the regulation of tissue-specific alternative splicing occurs through the dosage of splicing factors [239]. In this particular study, the overexpression of a single splicing factor, SF1, forced the expression of a muscle-specific isoform of tropomyosin in mouse embryonic fibroblasts. The study also provided evidence that the relative expression level of a second splicing factor hnRNP A1 had an antagonistic effect on splice site selection. These results suggest that the corresponding dosage of splicing factors can have an impact on tissue-specific alternative splicing. A study by Boutz et.al in 2007, showed that the expression of an

alternative splicing factor, polypyrimidine tract-binding protein (PTBP1), can be regulated by the miRNA *miR-133* during muscle development [241]. When taken in context with the results from Cáceres et al. in 1994, which showed that alternative splicing is regulated in a dosage dependent manner [239], we hypothesized that miRNAs regulate the dosage of splicing factors to influence tissue-specific alternative splicing.

In fact, an analysis of our tissue-specific miRNA target data sets from the intestine and body muscle reveled a large number of RBPs targeted in both tissues, with a specific enrichment of splicing factors targeted in the intestine. We validated the regulation of candidate splicing factors, *hrp-2*, *asd-2* and *smu-2,* and find that they are indeed regulated through their 3'UTRs in the intestine. We tested the implications of miRNA targeting to the alternative splicing patterns of genes regulated by *asd-2* and *hrp-2* (*unc-60*, *lin-10*, *unc-52* and *ret-1*), and find that the alternative splicing of all four genes are impacted by miRNAs. To explore the extent to which miRNAs contribute to alternative splicing, we used publicly available transcriptome data from worm strains deficient in the miRNA pathway and found genome wide dysregulation of splicing in the absence of miRNAs. The results presented in this chapter suggest a potential novel role for miRNAs in mRNA biogenesis, where regulation of splicing factors by miRNAs produces tissue-specific alternative splicing.

Results

Intestine and body muscle miRNAs target RNA binding proteins

Upon further analysis, we observed an unexpected enrichment of genes containing RNA binding domains in both the intestine and body muscle datasets. RNA binding proteins are known to play an important role in producing tissue-specific gene regulation by controlling gene expression at both the co- and post-transcriptional levels [242], and out of the ~887 RNA binding proteins (RBPs) defined in *C. elegans* [242], we identified almost half as targets of miRNAs across both tissues (45%).
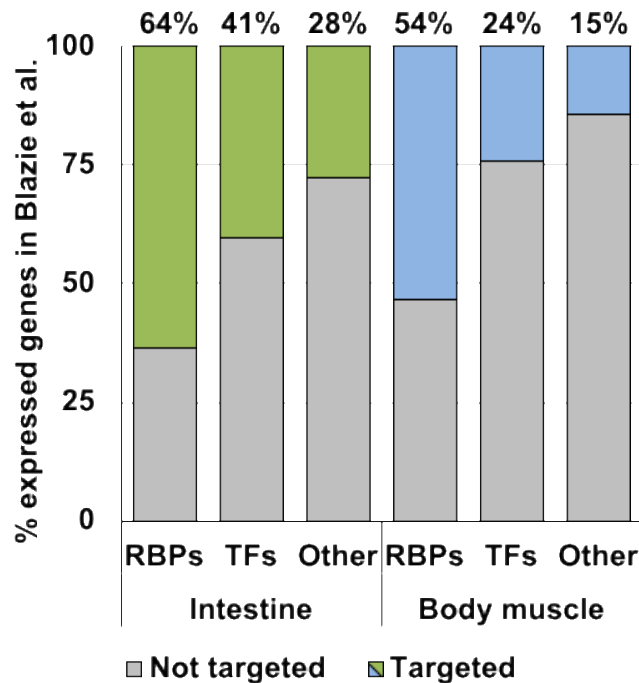
Figure 4.2 - Proportion of RBPs targeted by miRNAs in each tissue. There is an enrichment of RBPs targeted in the intestine (green 64%) and the body muscle (blue 54%). 'TFs' represents genes annotated as transcription factors while 'Other' represents protein-coding genes that are not RBPs.

We found that out of the 599 known RBPs present in the intestine

transcriptome [56, 191], 380 (64%) were present in our intestine dataset as targets

of miRNAs (Figure 4.2). This is a notable enrichment when compared to

transcription factors and non-RBP genes found in these tissues by Blazie et al.

2017 [56], of which only a fraction were identified in our study as miRNA targets

(Figure 4.2)

A similar trend is also present in the body muscle, with 170 (54%) of RBPs

identified as miRNA targets (Figure 4.2). Importantly, the largest pool of targeted RBPs

in both tissues was composed of general factors (GF), such as translation factors, tRNA

interacting proteins, ribosomal proteins, and ribonucleases (Figure 4.3), suggesting

extensive miRNA regulatory networks are in place in this tissue.
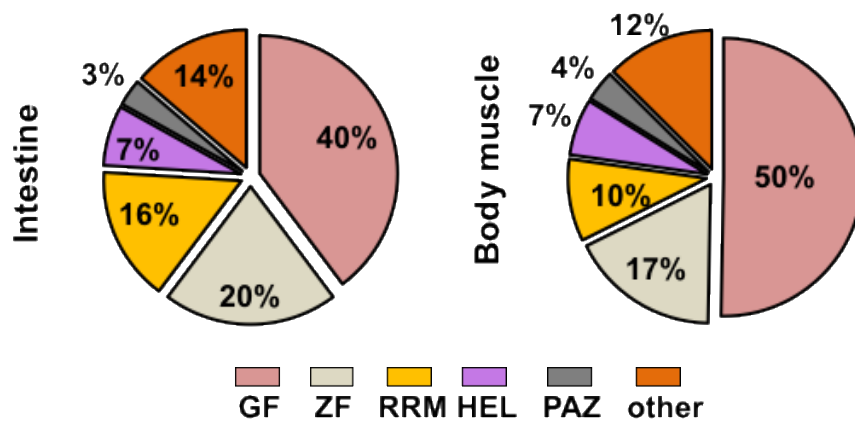
**Figure 4.3 - Subtypes of RBPs targeted by miRNAs in the intestine and
body muscle.** GF - General Factors, including translation factors, tRNA
proteins, ribosomal proteins, and ribonucleases; ZF - Zinc finger; RRM - RNA
recognition motif; HEL - RNA Helicase; PAZ - PIWI PAZ, PIWI, Argonautes.
The majority of the targeted RBPs are general and zinc finger-containing
factors.

89

Further analysis of the RNA binding proteins targeted in each tissue revealed that one of the most abundant classes of RBPs detected in our ALG-1 pull-down in the intestine and body muscle datasets were RNA splicing factors (Figure 4.4). The *C. elegans* transcriptome contains at least 78 known RNA splicing factors involved in both constitutive and alternative splicing [242]. 64 RNA splicing factors (82%) have been previously assigned by our group in the intestine [56, 191] and presumably are responsible for tissue-specific RNA splicing. 31 RNA splicing factors (40%) were also previously assigned by our group to the body muscle tissue [56, 191].

#	Splicing factor	Human homologue
1	*prp-17*	CDC40
2	M03F8.3	CRNKL1
3	*fubp-3.2*	FUBP3
4	*hrpf-1*	HNRNPF
5	*hrpf-2*	HNRNPH3
6	*hrp-2*	HNRNPR
7	*hrp-1*	HNRPA0
8	*smu-2*	IK
9	F53B7.3	ISY1
10	*prp-19*	PRPF19
11	*prp-3*	PRPF3
12	*prp-31*	PRPF31
13	*prp-4*	PRPF4
14	*rnp-6*	PUF60
15	*asd-2*	QKI
16	*fox-1*	RBFOX

#	Splicing factor	Human homologue
17	*sfa-1*	SF1
18	T13H5.4	SF3A3
19	T08A11.2	SF3B1
20	W03F9.10	SF3B2
21	*sap-49*	SF3B4
22	*rsp-2*	SFRS4
23	*swp-1*	SFSWAP
24	K07C5.6	SLU7
25	*snr-3*	SNRPD1
26	*snr-4*	SNRPD2
27	*snr-6*	SNRPE
28	*snr-5*	SNRPF
29	*rsp-7*	SREK1
30	*spk-1*	SRPK1
31	*rsp-3*	SRSF1
32	*rsp-4*	SRSF2

#	Splicing factor	Human homologue
33	*rsp-5*	SRSF4
34	*rsp-1*	SRSF6
35	*rsp-6*	SRSF7
36	*rsr-2*	U2SURP
37	F56G4.4	WBP4

#	Splicing factor	Human homologue
1	F56G4.4	WBP4
2	*prp-3*	PRPF3
3	*rsp-2*	SFRS4
4	*rsp-3*	SRSF1
5	*rsp-5*	SRSF4
6	*swp-1*	SFSWAP
7	*prp-38*	PRPF38A
8	*rsr-1*	SRRM1
9	T10C6.5	CWC15

Figure 4.4 - RNA Splicing factors identified as miRNA targets in the intestine and body muscle tissues. Intestine – green. Body muscle – blue**.** More than half of the RNA splicing factors examined are targeted by miRNAs in the intestine as compared to body

Our tissue-specific ALG-1 pull-down identified 37 RNA splicing factors as miRNA targets in the intestine (~47%) (Figure 4.4), and 34 of these were also previously identified by our group as being expressed in this tissue [56, 191]. In contrast, we have detected only nine RNA splicing factors targeted by miRNAs in our body muscle tissue ALG-1 pull-down, five of which previously assigned by our group in the body muscle transcriptome [56, 191] (Figure 4.4).

The difference in RNA splicing factors targeted by miRNA in these two tissues is significant as with the intestine contains three orders of magnitude more miRNA targeted RNA splicing factors than the body muscle. Of note, many different sub-types of RNA splicing factors identified in this study have human homologs, such as well-known snRNPs, hnRNPs and SR proteins (Figure 4.4).

Expression of the RNA splicing factors asd-2, hrp-2 and smu-2 is modulated through their 3'UTRs

In order to validate that RNA splicing factors found in our ALG-1 pull-down IPs are targeted by miRNAs in the intestine, we used the pAPAreg dual fluorochrome vector we developed in a past study [56] (Figure 4.5). This vector uses a single promoter to drive the transcription of a polycistronic pre-mRNA where the coding sequence of the mCherry fluorochrome is separated from the coding sequence of GFP by a SL2 trans-splicing element (SE) [56]. The test 3'UTR is cloned downstream of the GFP gene. Since the mCherry transcript is trans-spliced, it reports transcription activation. The GFP gene instead reports translational activity; since its expression is dictated by the downstream tested 3'UTR. If a given miRNA targets the test 3'UTR, the GFP intensity decreases

when compared with an untargeted 3'UTRs (*ges-1*). By comparing the ratio of the

mCherry (indicating transcription) to the GFP (indicating translation) fluorochromes, we

are able to define the occurrence of post-transcriptional silencing triggered by the tested

3'UTR (Figure 4.5) [56].

We selected three representative RNA splicing factors identified in our

study in the intestine (*asd-2*, *hrp-2 and smu*-2) and prepared transgenic strains to

validate their expression and regulation (Figure 4.6). We used the *ges-1* 3'UTR as

a negative control for miRNA targeting, as it is strongly transcribed and translated

in the intestine, with no predicted miRNA binding sites (PicTar), and poorly

conserved seed regions (TargetScan), suggesting minimal post-transcriptional

gene regulation [243, 244]. *ges-1* was not significantly abundant in our intestine

ALG-1 pull-down. The presence of the *ges-1* 3'UTR in the pAPAreg vector led to

the expression of both mCherry and GFP fluorochromes, indicating robust

transcription and translation of the construct as expected (Figure 4.6).

We then cloned *asd-2*, *hrp-2* and *smu-2* 3'UTRs downstream of the GFP

fluorochrome in our pAPAreg vector, prepared transgenic worms expressing these
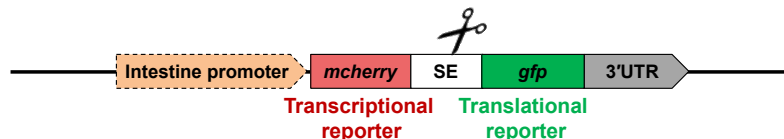
Figure 4.5 - Diagram of the construct used in these experiments (pAPAreg). An intestine-specific promoter drives the expression of a bi-cistronic dual fluorochrome vector in the intestine. The mCherry fluorochrome reports transcription activity of the construct, while the GFP reports post-transcriptional activity through the test 3'UTR cloned downstream of the GFP reporter sequence. If the test 3'UTR is targeted by repressive regulatory factors, such as miRNAs, the GFP fluorochrome lowers in its expression. SE: trans-splicing element extracted from the intergenic region located between the genes *gpd-2* and *gpd-3*.

constructs, and studied the fluctuation of the expression level of the GFP fluorochrome in these transgenic strains. All three 3'UTRs were able to significantly lower GFP expression when compared to the control strain with the *ges-1* 3`UTR, with ~40% repression, while the mCherry signal was similar in all strains (Figure 4.6). These results suggest that these three RNA binding proteins contain regulatory binding sites within their 3'UTRs potentially able to repress their expression.

MiRNAs target intestine RNA splicing factors promoting tissue-specific alternative splicing

We then tested changes to tissue-specific alternative splicing in the intestine caused by the STAR protein family member *asd-2*, which regulates the alternative splicing pattern of the gene *unc-60*. *unc-60* is expressed as two alternatively spliced isoforms in a tissue-specific manner [245] (Figure 4.7); *unc-60a* is expressed predominantly in the body muscle while *unc-60b* is expressed in many other tissues including the intestine [245].

We first tested the *unc-60* RNA isoform ratio in *wt* N2 worms. We extracted total RNA from N2 worms in triplicate and performed RT-PCR experiments using primers flanking the two *unc-60* isoforms (Figure 4.7). As expected, we found that the *unc-60a* longer isoform was more abundantly expressed in *wt* worms (62%) (Figure 4.8 and 4.9).
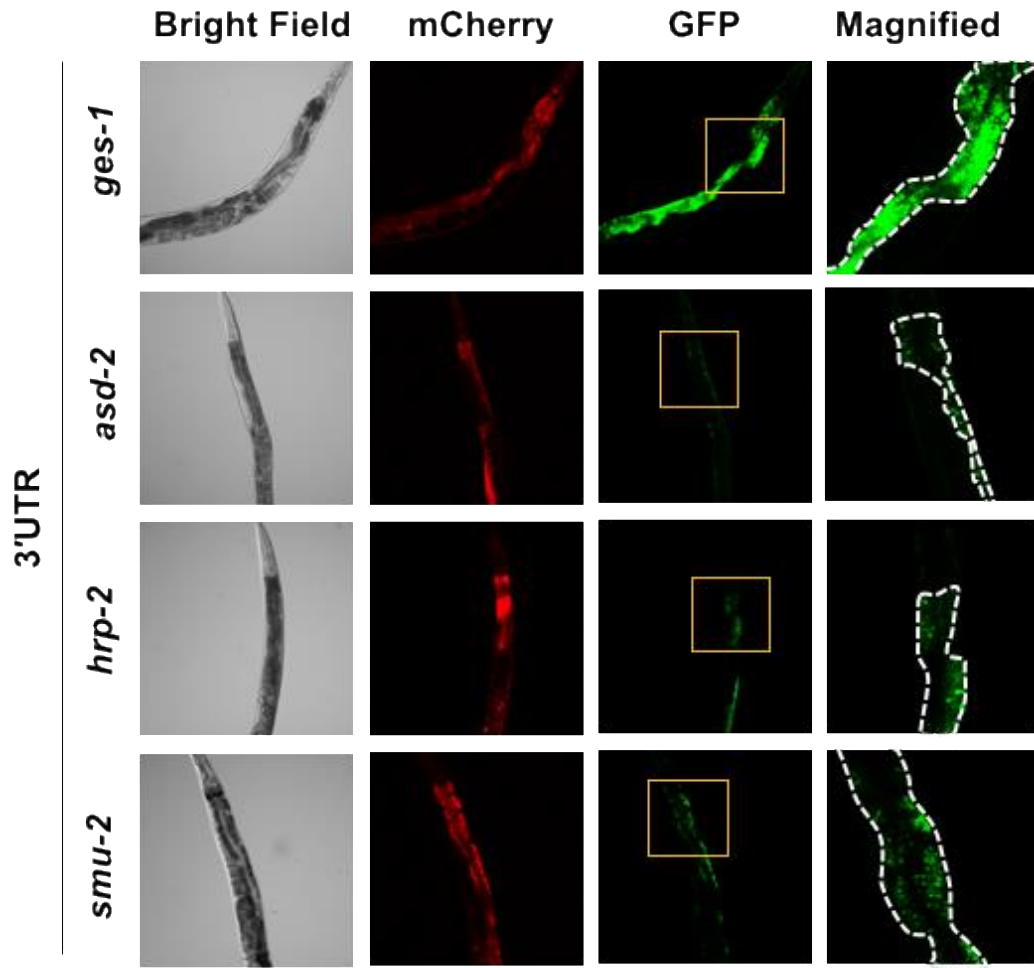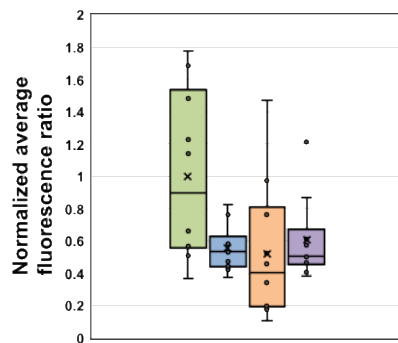
Figure 4.6 - Representative images of *C. elegans* strains generated with pAPAreg constructs expressing one of the following 3'UTRs: *ges-1, asd-2, hrp-2* or *smu-2* downstream of the GFP fluorochrome. Yellow boxes indicate magnified regions. White dotted lines indicate the intestine. C) The bar graphs show the quantified and normalized mean fluorescence ratio between the GFP and the mCherry fluorochromes. The mean fluorescence ratio is calculated from 10 worms per strain. Student t-test *p<0.05. We observed ~40% reduction in normalized GFP intensity modulated by *asd-2*, *hrp-2*, and *smu-2* 3'UTRs.

Gene: *unc-60*
Splicing factor: *asd-2*
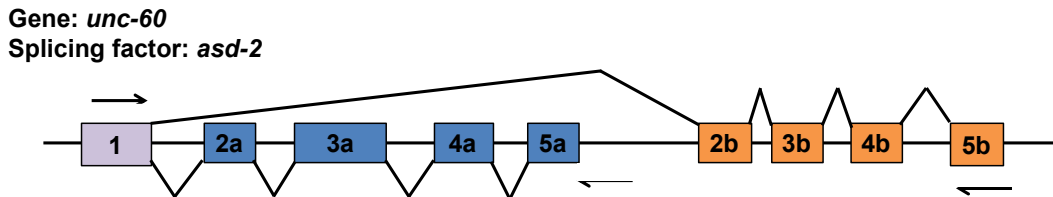
Figure 4.7 - Schematic of the genomic locus of *unc-60*. This gene is expressed as a longer *unc-60a* isoform, and a shorter *unc-60b* isoform. Arrows mark the binding sites of the primers used to detect the two isoforms.
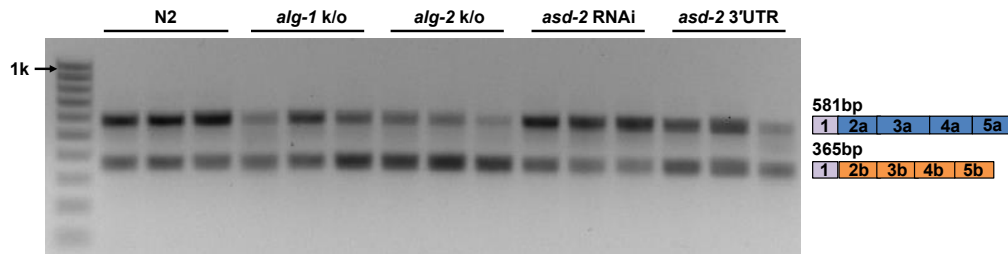
Figure 4.8 - RT-PCR performed from total RNA extracted from biological replicates in triplicate and visualized in 1% agarose gel. 1) N2: *wt* worms. 2) *alg-1* k/o: RF54[*alg-1(gk214)* X], 3) *alg-2* k/o: WM53[*alg-2(ok304)* II], 4) *asd-2* RNAi: N2 worms subjected to *asd-2* RNAi, 5) Over expression of *asd-2* 3'UTR in the intestine.
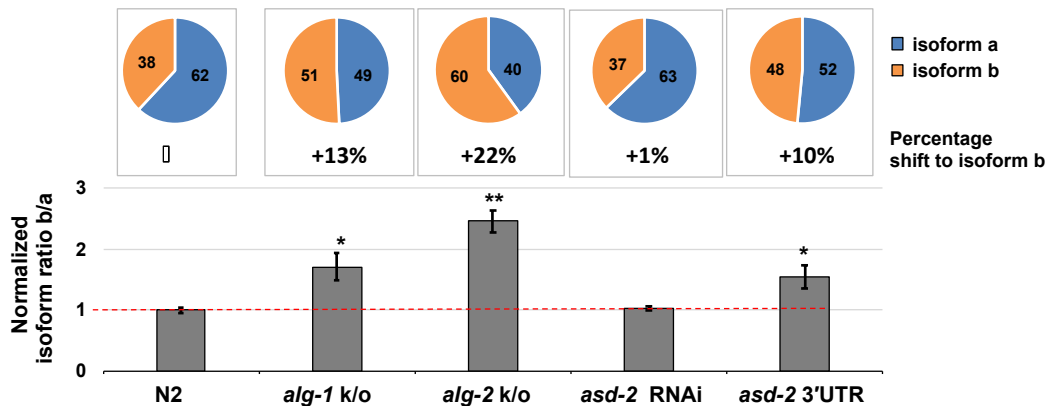
Figure 4.9 – The splicing pattern of *unc-60* is modulated by miRNA activity in the intestine. The pie charts below each gel show quantification of each of the occurrence of the two isoforms. The percentage below the pie chart is the increase in *unc-60* isoform b abundance when compared to N2(*wt*). The bar chart shows the change in isoform ratio between strains. The y-axis shows the abundance ratio (shorter isoform/longer isoform) of the two alternatively spliced isoforms examined. Exon skipping increases in *alg-1* and *alg-2* k/o strains, and in *asd-2* 3'UTR overexpression strains. Error bars indicate standard error of the mean. Student t-test *$p<0.05$ **$p<0.01$.

We then investigated if the miRNA pathway has a role in regulating these splicing events, by testing changes in *unc-60* isoform abundance in the *alg-1* and *alg-2* knockout strains (RF54 (*alg-1(gk214)* X) and WM53(*alg-2(ok304)* II). These strains are deficient in miRNA-based gene regulation. We found that loss of these miRNA effectors leads to a 10-20% shift in the expression of the two *unc-60* isoforms (Figure 4.5 and 4.9), indicating the importance of the miRNA pathway in regulating alternative splicing of this gene.

We then used a genetic approach to test the alternative splicing of this gene in the context of miRNA regulation. We reasoned that if ALG-1 targets the *asd-2* 3'UTR in the intestine lowering the expression of *asd-2,* which in turn causes *unc-60* alternative splicing pattern, we should be able to interfere with this
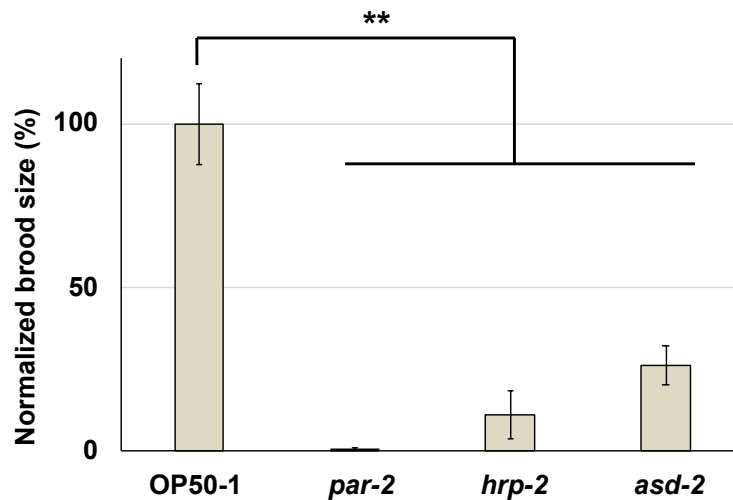
Figure 4.10 - Brood size assay to determine the efficiency of RNAi targeting *hrp-2* and *asd-2*. RNAi performed against the gene *par-2* (positive control) produced almost no hatched larvae as expected. A decrease in brood size was observed when RNAi was performed for the splicing factors *hrp-2* and *asd-2* indicating the knockdown of these genes. **p<0.01 n=5

96

mechanism by overexpressing the *asd-2* 3'UTR in this tissue and in turn test the role of the miRNA pathway in this process.

As expected, the overexpression of the *asd-2* 3'UTR in the intestine led to changes in the *unc-60* alternative splicing pattern, indicating that post-transcriptional regulation of *asd-2* through its 3'UTR is important for the alternative splicing pattern of *unc-60* in the intestine (Figure 4.8 and 4.9). Conversely, *asd-2* RNAi did not induce changes in *unc-60* alternative splicing pattern (Figure 4.8 and 4.9).

We validated the efficiency of our RNAi experiments by performing a brood size assay, which indicated strong RNAi activity (Figure 4.10). Similar results were observed by testing a second splicing factor (*hrp-2*) known to direct alternative splicing of the genes *ret-1*, *lin-10* and *unc-52* [246, 247] (Figure 4.11).

Loss of miRNA function lead to dispersed changes in splice junction usage

Since our data support a role for the miRNA pathway in modulating mRNA biogenesis, we were interested in testing the extent of these effects at the transcriptome level. We decided to download and mapped splicing junctions in genes from *alg-1* and *alg-2* knockout strains previously published by Brown et al., 2017 [177]. These worm strains are viable but are severely impaired. We reasoned that if the miRNA pathway contributes at some level to mRNA biogenesis, we should be able to see widespread changes in the usage of splice junctions in these datasets. To test this hypothesis, we downloaded the *alg-1* and *alg-2* datasets (three replicates for each strain plus *wt* N2 control), and extracted splice junction information.
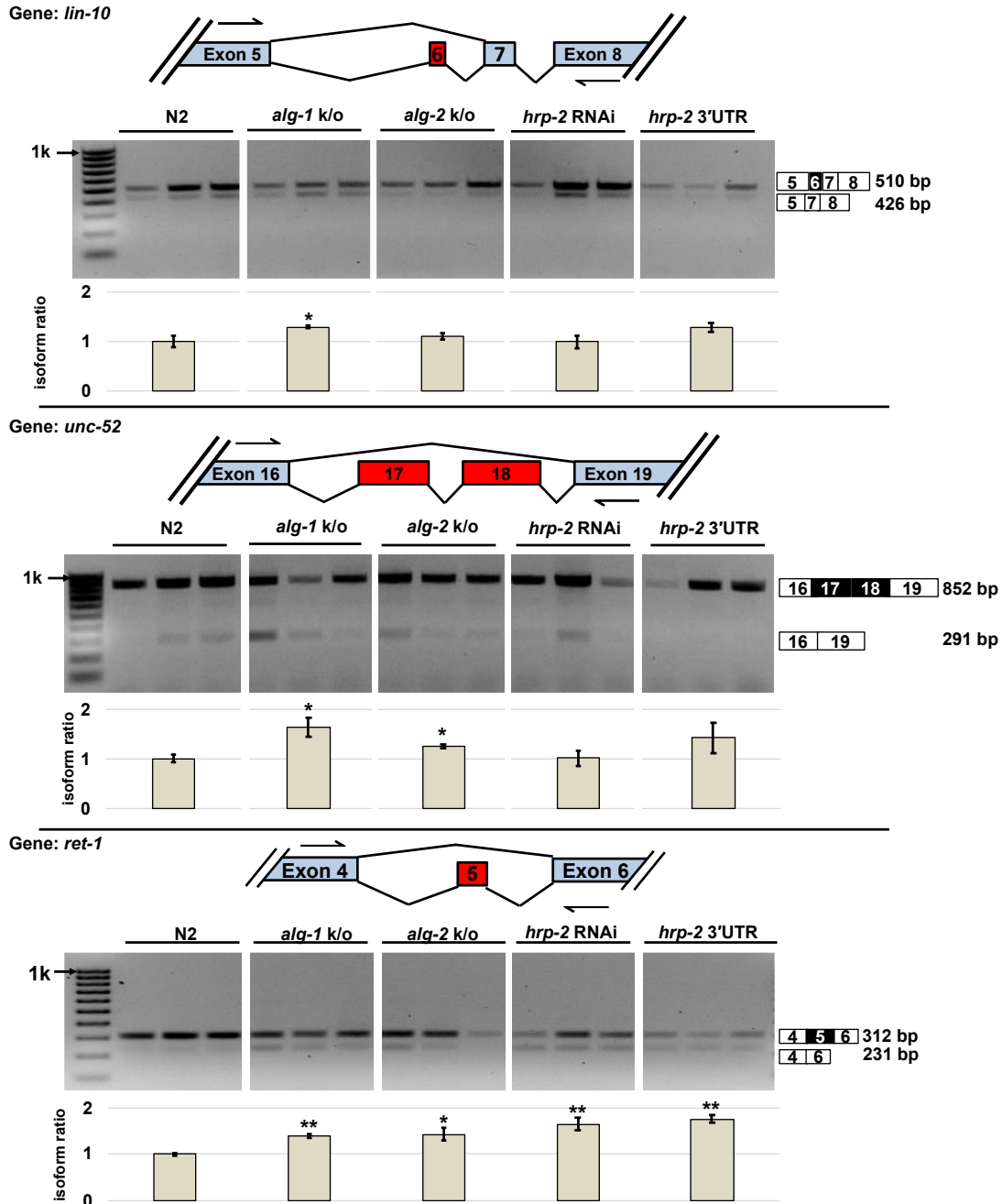
Figure 4.11 - *lin-10*, *unc-52* and *ret-1* exon skipping events in the intestine (1) Biochemical evidence. Top panels: The *lin-10*, *unc-52* and *ret-1* gene models are shown to illustrate the alternatively spliced exons (red) and the primer pairs used in the RT-PCR experiments to detect intestine specific alternative splicing modulated by miRNAs. Bottom Panels: Agarose gels visualizing RT-PCRs performed in triplicate on independent RNA extractions to test changes of alternative splicing of individual genes. *p < 0.05; **p <0 .01

We first tested if the effects we observed in *unc-60* with our biochemical and genetic approaches (Figure 4.8 and 4.9) could also be detected in these datasets. In the case of *unc-60*, there is a 6-10% change in splice junction usage between isoforms consistently in all re-annotated replicates, in both *alg-1* and *alg-2* knockout strains (Figure 4.12). This result is in line with our analysis in Figure 4.8 and 4.9. A similar and more striking aberrant splice junction usage is observed in the case of *lin-10*, and *unc-52*, and with a less pronounced effect in *ret-1* (Figure 4.13). These results are also in agreement with our study in Figure 4.11.

We then expanded this analysis to all splicing junctions we were able to map using these transcriptomes. From a total of 30,115 high quality known splice junctions
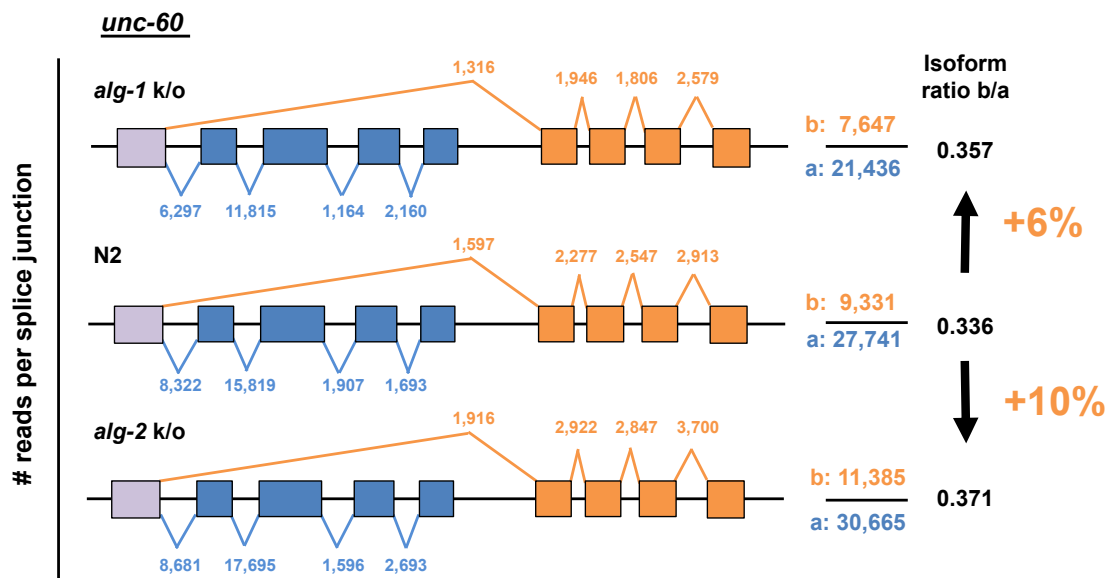
Figure 4.12 (next page) - Comparison of the splice junction usage in *unc-60* as observed in transcriptome data for *alg-1* and *alg-2* knockout strains (BROWN et al. 2017). The numbers above each slice junction indicates the number of reads mapped to that splice junction. The total reads for each isoform are indicated next to the gene model. The isoform ratios indicated next to the gene models are calculated by dividing the total reads for each isoform. There is a ~6-10% increase in the expression of the shorter *unc-60b* isoform in the miRNA deficient strains. Blue: reads corresponding to *unc-60a*. Orange: reads corresponding to *unc-60b*.

present in all three datasets, we identified ~3,946 of them in ~2,915 protein-coding genes

that were affected by more than 2-fold change in usage in both *alg-1* and *alg-2* knockout

datasets (~13.2% of the total mapped splice junctions) (Figure 4.13). In addition, we

detect several cases of exon inclusion, skipping and aberrant splicing events that occur

exclusively in the *alg-1* and/or *alg-2* mutant strains (Figure 4.14).
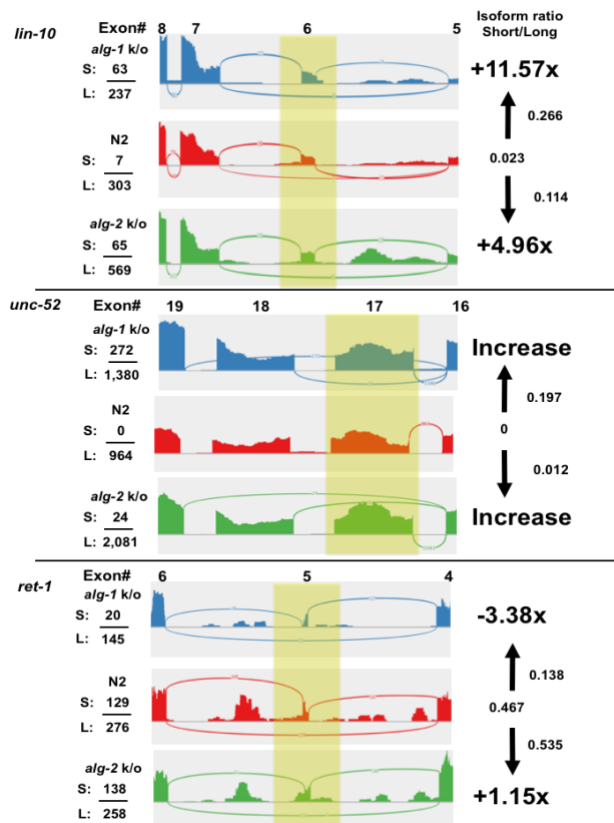
Figure 4.13 - *lin-10, unc-52* **and** *ret-1* **exon skipping events in the intestine (2) Sashimi plots generated by re-analyzing RNA-seq data from Brown et.al 2017**. Replicates for each strain were combined and analyzed with the TopHat bioinformatic package. The plots illustrate the number of reads mapped for each splice junction. Top and middle panels: The splicing events for *lin-10* and *unc-52* showed an increase of exon skipping occurring in both miRNA deficient strains. Lower panel: *ret-1* exon 5 skipping increases slightly in *alg-2* but shows increased exon inclusion in *alg-1* knockout worms. Yellow boxes highlight alternatively spliced exons analyzed in this figure S: total reads for the short isoform; L: total reads for the long isoform.

100
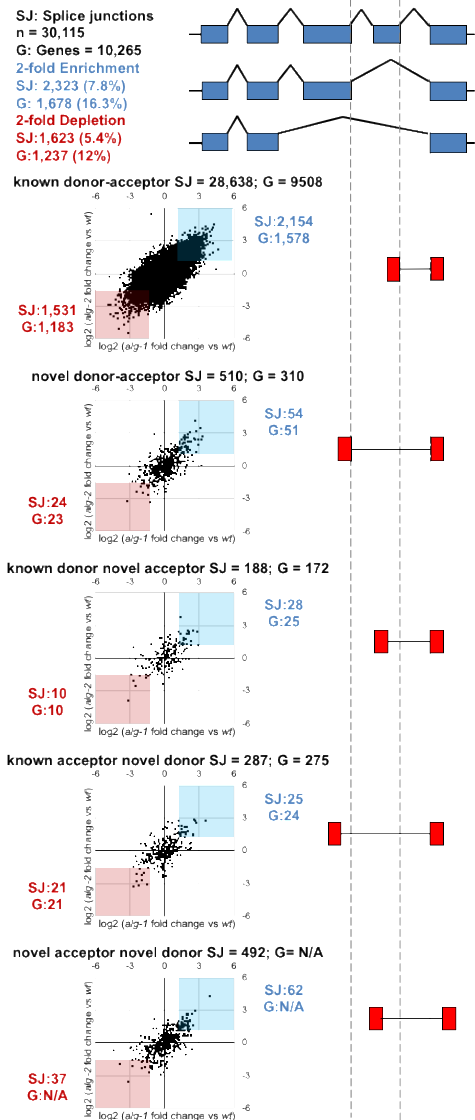
Figure 4.14 - Genome-wide changes in splice junction usage in *C. elegans* strains deficient in the miRNA pathway. Analysis of splice junction (SJ) usage in miRNA deficient strains (*alg-1*(*gk214*) or *alg-2*(*ok304*)) re-annotated from Brown et al. 2017 (BROWN et al. 2017). The graphs illustrate the changes in splice junction abundance for different types of splicing events. The x-axis represents the fold-change of the normalized number of reads for each splice junction, comparing the *alg-1*(*gk214*) strain to *wt*, while the y-axis represents the fold change obtained when comparing *alg-2*(*ok304*) to *wt*. Splice junctions with more than 2-fold enrichment in both strains are highlighted in blue, while splice junctions with 2-fold depletion in both strains are highlighted in red. The number of genes (G) with enriched or depleted splice junctions are indicated next to the graphs. There is a 2-fold change in ~13% of all the splicing events mapped in the knockout strains, affecting 3,301 genes when compared to the N2 *wt* control.

101

# of novel splice junction		
exon		aberrant
inclusion	skipping	
alg-1	1,622	1,840	409
alg-2	1,897	1,231	456
both	2,526	1,766	153
either	6,045	4,837	1,018

Figure 4.15 - The table summarizes the number of novel splicing events seen in *alg-1* and *alg-2* datasets. These are splicing events not observed in the N2 *wt* control, and indicate an increase in novel and aberrant splicing events in miRNA deficient strains.

Discussion

RBPs are enriched in the tissue-specific miRNA target datasets in both the intestine and body muscle

We found that miRNAs target RBPs in a tissue-specific manner, including several mRNA splicing factors such as hnRNPs and SR proteins (Figure 4.4). 64% of the RBPs expressed in the intestine were found in our intestinal ALG-1 pull-down, while 54% of the muscle RBPs were in our muscle ALG-1 pull-down. This result was unexpected given the small number of RNA binding proteins

previously identified in the *C. elegans* genome (n = 887) [242], which amounted to only 4% of the total *C. elegans* protein-coding genes. However, previous studies have hinted at a strong regulatory network between miRNAs and RBPs, as the 3'UTRs of RBPs were found to contain on average more predicted miRNA binding sites than other gene classes [242].

RNA binding domain-containing proteins are involved in many biological processes, and their role is not limited to RNA biogenesis [242]. RBPs can bind single or double strand RNAs, and associate with proteins forming ribonucleoprotein complexes (RNPs). Longevity, fat metabolism, and development are all processes controlled by RNPs [182, 248, 249], and in the context of miRNA regulation, the ability of miRNAs to control RBPs abundance and function allows for increased control of fundamental cellular core processes. 234 RBPs uniquely detected as miRNA targets in the intestine, while 147 RBPs are shared between both datasets.

RNA Splicing factors are regulated by miRNAs in a tissue-specific manner and influence alternative splicing

Within this intestinal dataset, we mapped a surprising number of RBPs involved in RNA splicing (Figure 4.4). We performed a literature search for known RNA splicing factors in *C. elegans*; out of the 72 total protein identified, 37 of them were enriched in our intestine ALG-1 pull-down. In contrast, we do

not observe this level of complexity in the body muscle, with only 9 RNA

splicing factors identified in this dataset (Figure 4.4).

asd-2 and *smu-2* are well-known RNA splicing factors that induce exon

retention in a dosage-dependent manner [245, 250], while *hrp-2* abundance leads

to exon skipping [246]. Here we show that all three RNA splicing factors possess

regulatory targets within their 3'UTRs (Figure 4.5 and 4.6) that amount to ~40%

silencing activity in the intestine (Figure 4.5). Although we do not know which

miRNAs target the *asd-2* and *hrp-2* 3'UTRs, in Figures 4.7 – 4.9 and Figure 4.11

we show that the miRNA pathway influences splice junction usage by regulating

these genes. The depletion of miRNAs which target these RNA splicing factors

by using sponge approaches led to defects in the alternative splicing pattern of

downstream genes regulated by *asd-2* and *hrp-2*.

Interestingly, the miRNAs predicted to target most splicing factors were

not found highly expressed in this study. *miR-85* and *miR-355*, the most abundant

and tissue-restricted miRNAs identified, are only predicted to target less than 10%

of all the RBPs found. This suggests that since miRNAs are highly reactive, the

abundance of those involved in regulating RNA alternative splicing may be

tightly regulated in tissues, to make sure splicing events are properly executed.

Our genome-wide splice junction mapping effort in miRNA deficient

strains shows similar trends of aberrant splicing of *unc-60, unc-52, lin-10* and *ret-*

1 (Figure 4.12 and 4.13) and display an overall disruption of splicing events

(~13.2% of all splice junctions mapped) (Figure 4.14 and 4.15). Most of these

defects are in known donor-acceptor splicing events, perhaps because RNA surveillance mechanisms may hide more severe disruptions.

Unfortunately, our *in vivo* approach does not reach the resolution needed to conclusively pinpoint the extent of the miRNA pathway in this process. To perform *in vivo* experiments, we used total RNA extracted from transgenic worms and studied the change in exon abundance occurring in a single tissue within a whole animal, which prevented us from reaching the same resolution obtainable with *in vitro* splicing experiments and mini-genes. In addition, the effects we observe are ameliorated by the presence of at least one functional Argonaute protein, which can compensate for the loss of the other. Knockout of the entire miRNA pathway is lethal in *C. elegans*, and while aberrant splicing may play a role in producing this phenotype, these activities are challenging to detect *in vivo*.

Taken together, our results support a role for miRNAs in regulating alternative splicing in the intestine, where their presence in a tissue-specific manner may lead to alteration of the dosage balance of RNA splicing factor, leading to tissue-specific alternative splicing. MiRNAs are known to alter gene expression dosage, rather than induce a complete loss of protein function [11, 15]. On the other hand, many RNA splicing factors involved with constitutive and alternative splicing are ubiquitously expressed [251] but are somehow able to induce tissue-specific alternative splicing in a dosage-dependent manner. In this context, it is feasible that miRNAs may alter the dosage of RNA splicing factors, leading to tissue-specific alternative splicing.

Experimental

Cloning of expression constructs and preparation of transgenic strains

The 3'UTRs of the genes in this study were cloned by anchoring the Gateway-compatible primers at the translation STOP codon of each gene, to the longest annotated 3`UTR (Primers 1-6 Table 4). We have included 50 base pairs downstream of the annotated PAS site to include 3'end processing elements. The PCR products were analyzed using gel electrophoresis analysis and used to perform Gateway BP Clonase reactions (Invitrogen, cat. 11789020) into pDONR P2RP3 as per the manufacturer's protocol. The *unc-54* 3'UTR used in this study was previously described in Blazie et al., 2017. The constructs injected were assembled by performing Gateway LR reactions (Invitrogen) with each promoter, reporter, and 3'UTR construct per the manufacturer's protocol into the MosSCI compatible destination vector CFJ150. We then microinjected each reporter construct (100ng/μl) with CFJ601 (100ng/μl) into MosSCI compatible *C. elegans* strains using standard microinjection techniques [217].

Fluorescent imaging and analysis of nematodes

Confocal images used in Figure 4.6 were acquired in the Biodesign Imaging Core, Division of the Arizona State University Bioimaging Facility. Transgenic strains were grown at room temperature on NGM plates seeded with OP50-1. The mixed stage worms were washed twice with M9 and resuspended in 1mM of levamisole before imaging using a Nikon C1 Ti-E microscope with 488

nm and 561 nm lasers, 0.75 numerical aperture, 90 µM pinhole microscope with a 40x magnification objective lens. We acquired 10 images for each transgenic strain (total of 40 images) using the same microscope settings. The fluorescence of GFP and mCherry fluorochromes from the acquired images were individually quantified using the integrated density (ID) function of the ImageJ software [252]. Fluorescence ratios were then calculated for each worm (n=10, total 40 images) by dividing the ID for GFP by the ID for mCherry. The finalized result for each strain is the averaged fluorescence ratio calculated across all 10 imaged worms. We performed a two-tailed student t-test to compare the mean fluorescence ratios for each strain with a p-value cut off <0.05 to establish the presence of post-transcriptional gene regulation.

Re-annotation of alg-1 and alg-2 knockout transcriptome datasets and splice junction identification

We downloaded from the GEO database the following transcriptome datasets published by Brown et al., 2017 [177]: Project number GSE98935, Wild type Rep 1-3 (GSM2628055, GSM2628056, GSM2628057); *alg-1*(*gk214*) Rep 1-3 (GSM2628061, GSM2628062, GSM2628063); *alg-2*(*ok304*) Rep 1-3 (GSM2628064, GSM2628065, GSM2628066). We used in-house Perl scripts to prepare the reads for mapping, and then these reads as input to the TopHat algorithm [253] to map splice junctions in all nine datasets independently. The TopHat algorithm mapped between 30-56M reads to splice junctions in each sample. *wt*_rep1; 43,721,355 mapped reads (64% of total input reads),

*wt*_rep2; 44,440,441 (64%), *wt*_rep3; 37,248,408(62.7%), *alg-1*_rep1; 30,808,645

(62.3%), *alg-1*_rep2; 35,914,514 (63.2%), *alg-1*_rep3; 43,721,355(63.9%), *alg-2*_rep1;

54,471,761(63.2%), *alg-2*_rep2; 56,000,173 (66.8%), *alg-2*_rep3; 46,638,369 (63.9%).

We then combined the mapped reads obtained in the three replicates for each strain and

used the open source software regtools (Griffith Lab, McDonnell Genome Institute) to

annotate these splice junctions using the following command 'regtools junctions annotate

junctions.bed WS250.fa WS250.gtf'. The software produced ~41.8k splice junctions

supported by at least 10 reads for the combined N2 *wt* dataset, ~42.3k splice junctions for

the *alg-1* dataset and 46.3k for the *alg-2* dataset. We analyzed the three resulting

cumulative datasets normalized by dividing each score by the total number of mapped

reads within each sample. This approach produced 36.7k high-quality splice junction for

the combine N2 *wt* dataset, ~37k for the *alg-1* combined dataset and ~38k for the

combined *alg-2* dataset. The analysis in Figures 4.14 and 4.15 were performed using

splice junctions that are present in all three datasets (30,115 total). To calculate the fold-

change for each splice junction, we divided the normalized scores of each splice junction

in the *alg-1* and *alg-2* combined datasets by the corresponding scores in the wild type

combined dataset. The fold change of each splice junction was then plotted on a $\log_2$

scale shown in Figure 4.14.

RNAi experiments

The RNAi experiments shown in Figure 4.8 and Figure 4.11 were performed as

follows. N2 worms were synchronized by bleaching and starving overnight in M9 buffer

until they reached the L1/dauer stage and then transferred to agar plates containing OP50-1 bacteria, HT115 bacteria with pL4440 *hrp-2* RNAi or pL4440 *asd-2* RNAi [254]. We used *par-2* RNAi as a positive control for the experiments, which results in 100% embryonic lethality. To measure the brood size, individual synchronized young adult worms were left overnight (16 hours) to lay eggs. Hatched larvae were counted 24 hours later. Total RNA was extracted from N2 worms treated with either *hrp-2* or *asd-2* RNAi at the adult stage in triplicates.

RNA extraction for detection of intestine-specific splicing variants

We extracted total RNA using the Direct-zol™ RNA MiniPrep Plus kit (Zymo Research, cat ZR2070) from (1) N2 *wt* worms, (2) RF54 (*alg-1(gk214)* X) strain, (3) WM53 (*alg-2(ok304)* II) strain, (4) N2 strain subjected to RNAi as previously described[255] for *asd-2* and *hrp-2*, and (5) transgenic worms overexpressing the *asd-2* 3'UTR or the *hrp-2* 3'UTR under control of an intestinal promoter (*ges-1p*::pAPAreg::3'UTR). Each strain was synchronized by growing in M9 media to L1/dauer stage then transferred to plates containing HT115. We extracted RNA 48 hours later from adult worms in triplicate for each condition.

cDNA preparation, image acquisition and splicing isoform analysis

cDNA was synthesized from each RNA sample using SuperScript III RT (Life Technologies, cat 18080093) according to the manufacturer's protocol. Briefly, 200ng of each RNA sample was incubated with 1 µL of 50mM poly dT anchor, 1 µL of 10mM

dNTP mix and brought to a total volume of 14 µL with nuclease-free H_2O and incubated for 5 minutes at 60°C then iced for 1 minute. 4 µL of 5x first strand buffer, 1 µL of 0.1M DTT and 1 µL (200 units) of SuperScript III reverse transcriptase were added to each sample and incubated at 50°C for 60 minutes then heat inactivated at 70°C for 15 minutes. 200ng of cDNA from each sample was used in PCRs consisting of 34 cycles using HiFi Taq Polymerase (Invitrogen, cat 11304011) according to manufacturer protocols. Primers used to test alternative splicing of *unc-60, unc-52, lin-10,* and *ret-1* were designed to flank the alternatively spliced exons and were adapted from previous studies (Table 4.1 primers 7-15) [245-247]. We then acquired images of the PCR amplicons (5 µL) separated by agarose gel electrophoresis. We used the integrated density function of ImageJ [252], by defining equally sized regions of interest around each band in the images and compared the integrated density values by normalizing the smaller bands to the larger bands. The resulting isoform ratios are displayed in Figures 4.9 and 4.11. Each strain was quantified in triplicate and subjected to a two-tailed student t-test. Statistical significance was assigned for p-values <0.05.

Table 4.1 - Primers used in Chapter 4

Primer #	Primer Name	Primer sequence
1	atB2r_asd-2_F	GGGGACAGCTTTCTTGTACAAAGTGGAG TAACCAACCAACACCACCAAC
2	attB3_asd-2_R	GGGGACAACTTTGTATAATAAAGTTG ATTGAATTCATAAAACTGTTTTCTTGC
3	attB2r_hrp-2_F	GGGGACAGCTTTCTTGTACAAAGTGGAG TAAATTGCACCCCCGGCAGA
4	attB3_hrp-2_R	GGGGACAACTTTGTATAATAAAGTTG CAAATTTGGGCAAACTTCCG
5	attB2r_smu-2_F	GGGGACAGCTTTCTTGTACAAAGTGGAG TGATTTTTTTAATGATTTTTATGAATTC
6	attB3_smu-2_R	GGGGACAACTTTGTATAATAAAGTTG AATAATTCGATTTTTATCATTTTCAG
7	lin-10_exon5_F	AAGAATCAGCGGGAGGAGAG
8	lin-10_exon8_R	GATGTGGTTGCATCATCTGG
9	ret-1_exon4_F	CATCCGCTGAAGGATCCATAG
10	ret-1_exon6_R	GAGCTTCCTCAGCAATCGGAG
11	unc-52_exon16_F	CGGACATCCAAGTGTTCAGC
12	unc-52_exon19_R	AACTGATGTCGCTCTCCTGG
13	unc-60_exon1_F	GAAACTCAACTTGATTCTATTTCCCC
14	unc-60_exon5A_R	CGACGACGACTTCGGAAGAGAC
15	unc-60_exon5B_R	GACTGGGGCGTTGTCTGGGC

Chapter 5

Conclusion

MicroRNA targetting principles uncovered using the h3'UTRome

The h3'UTRome (Chapter 2) is the first publicly available compendium of human 3'UTRs, and has enabled the study of human miRNA targets at the genomic level. Identifying miRNA targets at this scale has enabled us to observe targeting principles that were not apparent in low-throughput studies. Our results indicate that miRNAs target gene pathways by regulating multiple genes in the same pathway, as seen in the case of *let-7c* targeting multiple members of the RAS pathway. Expanding on these initial observations we have used the h3'UTRome and the 3'LIFE assay, to further identify the targets of *let-7c* and *miR-10b* [11]. We found further evidence that these miRNAs regulate gene pathways at multiple points, eluding to a general mechanism by which miRNAs act in cells [11]. This general targeting principle presents a possible explanation as to how miRNAs, that rarely silence genes entirely, are able to have significant effects on phenotypes.

Using the h3'UTRome, we identified regions of miRNAs outside the seed that impact target identification [11]. In this study we screened the h3'UTRome against multiple members of the *miR-10* family, which shares the same seed sequence but has differences at the 3'end of the miRNA. We found that these miRNA family members, that were believed to be functionally redundant, share many of the same targets but are able to target unique genes based on differences outside the seed. These observations

112

were made possible by studying miRNA targets at the genomic level using the h3'UTRome. Currently, the h3'UTRome captures only ~1.5% of the annotated human 3'UTRs, which while useful, does not allow us to fully identify all the targets of any miRNA. Expanding this library would allow us to better appreciate the contributions of miRNAs to global gene regulation.

MicroRNAs contribute to tissue-specific gene expression

Previous studies have identified specific cases where miRNAs are able to regulate gene expression in a tissue-specific manner, however the targets of miRNAs had not been systematically studied at the tissue-specific level. To study the tissue-specific roles of miRNAs, we optimized a crosslinking and immunoprecipitation-based approach to isolate and sequence miRNA targets from the intestine and body muscle of *C. elegans* (Chapter 3).

We found that at the tissue-specific level, miRNAs target gene pathways, and that each pathway was targeted at multiple points (Figure 5.1). The pathways regulated by miRNAs are important for tissue-specific function, suggesting that miRNAs fine tune gene expression to maintain tissue homeostasis. Many RBPs are targeted by miRNAs in a tissue-specific manner, and within this group of RBPs we found an enrichment of splicing factors targeted by miRNAs. Our data indicates a novel tissue-specific role for miRNAs in regulating the biogenesis of mRNA and contributing to alternative splicing by adjusting the dosage of splicing factors (Figure 5.2).

Intestine					Body muscle	
Metabolism	Vesical transport	Vitellogenesis	Transcription factors		Cellular architecture	Locomotion
enol-1	rab-1	vit-1	atf-7	pos-1	act-1	mup-2
ipgm-1	rab-6.1	vit-2	pha-4	mxl-3	act-2	myo-3
gpd-1	rab-7	vit-3	die-1		act-3	dlc-1
gpd-2	rab-8	vit-5	lss-4		act-4	
gpd-4	rab-21	vit-6			rho-1	
fat-1	rab-35					
fat-2	rab-39					

Figure 5.1 – The tissue-specific targets of miRNAs correspond to the known functions of each tissue. The intestine in the worm is responsible for fertility in addition to its role in metabolism, while the predominant function of the body muscle is to enable locomotion and these tissue-specific roles are reflected in the miRNA targets.
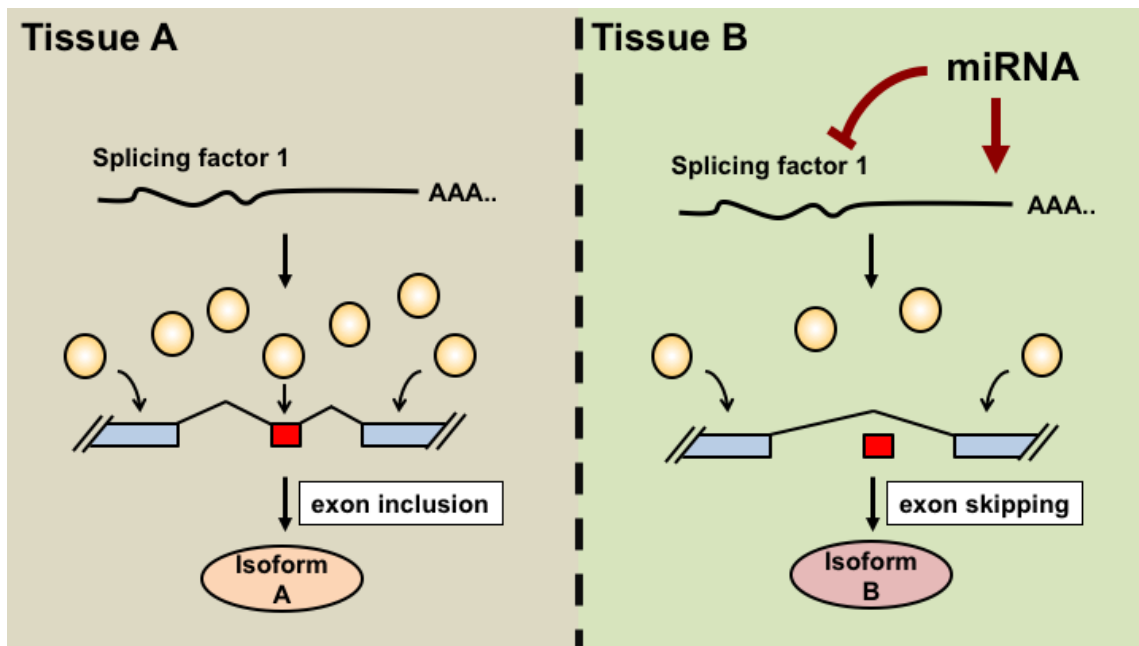
Figure 5.2 - A proposed role for miRNAs in the modulation of tissue-specific alternative splicing. The abundance of RNA splicing factors (yellow circles) dictates the splicing events in a given tissue A. The presence of a miRNA in Tissue B may lower the dosage of splicing factors resulting in tissue-specific alternative splicing.

114

Complex RNA-based regulatory networks

An elaborate program of gene regulation is required to develop and maintain the different cell types in a multicellular organism. Recently, the contributions of RNA to these regulatory networks have come to light. MiRNAs impact gene regulation throughout expression, from epigenetic control of DNA to post-translational modification of proteins (Figure 5.3) [241, 256, 257].

MiRNAs and their target sites in 3'UTRs are regulated in a tissue-specific manner, indicating that miRNAs have the potential to dictate tissue-specific gene expression. I hypothesized that miRNAs regulate tissue-specific gene expression. To test
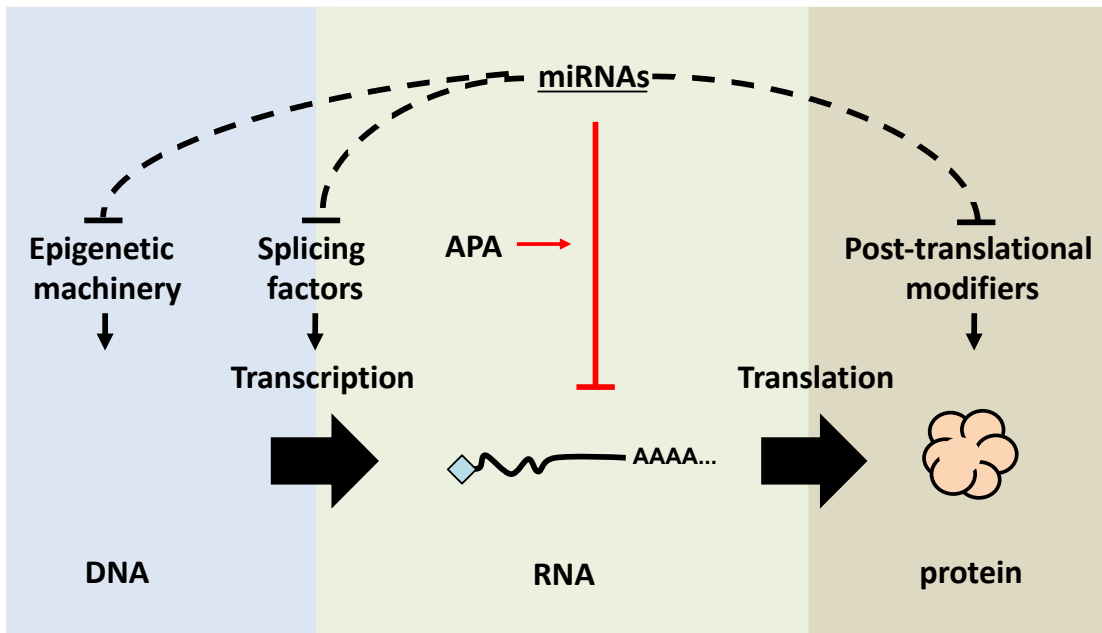
Figure 5.3 – An overview of miRNAs in complex RNA-based regulatory networks. MiRNAs directly regulate gene expression post-transcriptionally by targeting the 3'UTRs of mRNAs (red bar). APA modulates miRNA targeting by producing 3'UTR isoforms (red arrow). Due to their wide array of targets, miRNAs are able to impact gene regulation at every point of gene expression.

this hypothesis, we developed and utilized tools and techniques to identify miRNA targets using human cell lines, and *in vivo* in the model organism *C. elegans*. The research presented in this thesis provides compelling evidence for tissue-specific roles of miRNAs. We find that while miRNAs act at the post-transcriptional level, the effects of miRNAs can impact gene expression at multiple points leading to complex RNA-based regulatory networks that exist at the tissue-specific level (Figure 5.3).

REFERENCES

1.   Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.

2.   Sender, R., S. Fuchs, and R. Milo, *Revised Estimates for the Number of Human and Bacteria Cells in the Body.* PLoS Biol, 2016. **14**(8): p. e1002533.

3.   Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

4.   Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.

5.   International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.

6.   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.

7.   Shaye, D.D. and I. Greenwald, *OrthoList: a compendium of C. elegans genes with human orthologs.* PLoS One, 2011. **6**(5): p. e20085.

8.   Hillier, L.W., et al., *Genomics in C. elegans: so many genes, such a little worm.* Genome Res, 2005. **15**(12): p. 1651-60.

9.   Gilbert, W., *Origin of life: The RNA world.* Nature, 1986. **319**(6055): p. 618-618.

10.   Chen, C.Z., et al., *MicroRNAs modulate hematopoietic lineage differentiation.* Science, 2004. **303**(5654): p. 83-6.

11.   Wolter, J.M., et al., *Evolutionary patterns of metazoan microRNAs reveal targeting principles in the let-7 and miR-10 families.* Genome Res, 2017. **27**(1): p. 53-63.

12.   Kretz, M., et al., *Control of somatic tissue differentiation by the long non-coding RNA TINCR.* Nature, 2013. **493**(7431): p. 231-5.

13.   Bayne, E.H. and R.C. Allshire, *RNA-directed transcriptional gene silencing in mammals.* Trends Genet, 2005. **21**(7): p. 370-3.

14.	Maniatis, T. and B. Tasic, *Alternative pre-mRNA splicing and proteome expansion in metazoans.* Nature, 2002. **418**(6894): p. 236-43.

15.	Bartel, D.P., *Metazoan MicroRNAs.* Cell, 2018. **173**(1): p. 20-51.

16.	Xue, S. and M. Barna, *Specialized ribosomes: a new frontier in gene regulation and organismal biology.* Nat Rev Mol Cell Biol, 2012. **13**(6): p. 355-69.

17.	Payne, S.H., *The utility of protein and mRNA correlation.* Trends Biochem Sci, 2015. **40**(1): p. 1-3.

18.	Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast.* Mol Cell Biol, 1999. **19**(3): p. 1720-30.

19.	Ghazalpour, A., et al., *Comparative analysis of proteome and transcriptome variation in mouse.* PLoS Genet, 2011. **7**(6): p. e1001393.

20.	Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nat Rev Genet, 2012. **13**(4): p. 227-32.

21.	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.

22.	Selbach, M., et al., *Widespread changes in protein synthesis induced by microRNAs.* Nature, 2008. **455**(7209): p. 58-63.

23.	Jens, M. and N. Rajewsky, *Competition between target sites of regulators shapes post-transcriptional gene regulation.* Nat Rev Genet, 2015. **16**(2): p. 113-26.

24.	Siepel, A., et al., *Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes.* Genome Res, 2005. **15**(8): p. 1034-50.

25.	de Moor, C.H., H. Meijer, and S. Lissenden, *Mechanisms of translational control by the 3' UTR in development and differentiation.* Semin Cell Dev Biol, 2005. **16**(1): p. 49-58.

26.	Fabian, M.R., N. Sonenberg, and W. Filipowicz, *Regulation of mRNA translation and stability by microRNAs.* Annu Rev Biochem, 2010. **79**: p. 351-79.

27.	Mayr, C., *Evolution and Biological Roles of Alternative 3'UTRs.* Trends Cell Biol, 2016. **26**(3): p. 227-237.

28.   Yaffe, D., et al., *Highly conserved sequences in the 3' untranslated region of mRNAs coding for homologous proteins in distantly related species.* Nucleic Acids Res, 1985. **13**(10): p. 3723-37.

29.   Duret, L., F. Dorkeld, and C. Gautier, *Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression.* Nucleic Acids Res, 1993. **21**(10): p. 2315-22.

30.   Bashirullah, A., et al., *Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster.* EMBO J, 1999. **18**(9): p. 2610-20.

31.   Evans, T.C. and C.P. Hunter, *Translational control of maternal RNAs*, in *WormBook*, T.C.e.R. Community, Editor. November 10, 2005, WormBook.

32.   Zaucker, A., P. Kumari, and K. Sampath, *Zebrafish embryogenesis - A framework to study regulatory RNA elements in development and disease.* Dev Biol, 2019.

33.   Tadros, W. and H.D. Lipshitz, *The maternal-to-zygotic transition: a play in two acts.* Development, 2009. **136**(18): p. 3033-42.

34.   Lobbardi, R., et al., *Fine-tuning of Hh signaling by the RNA-binding protein Quaking to control muscle development.* Development, 2011. **138**(9): p. 1783-94.

35.   Hilgers, V., et al., *Neural-specific elongation of 3' UTRs during Drosophila development.* Proc Natl Acad Sci U S A, 2011. **108**(38): p. 15864-9.

36.   Ji, Z., et al., *Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 7028-33.

37.   Merritt, C., et al., *3' UTRs are the primary regulators of gene expression in the C. elegans germline.* Curr Biol, 2008. **18**(19): p. 1476-82.

38.   Higgs, D.R., et al., *Alpha-thalassaemia caused by a polyadenylation signal mutation.* Nature, 1983. **306**(5941): p. 398-400.

39.   Tian, B. and J.L. Manley, *Alternative cleavage and polyadenylation: the long and short of it.* Trends Biochem Sci, 2013. **38**(6): p. 312-20.

40.   Mangone, M., et al., *The landscape of C. elegans 3'UTRs.* Science, 2010. **329**(5990): p. 432-5.

41. 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.

42. Derti, A., et al., *A quantitative atlas of polyadenylation in five mammals.* Genome Res, 2012. **22**(6): p. 1173-83.

43. Ulitsky, I., et al., *Extensive alternative polyadenylation during zebrafish development.* Genome Res, 2012. **22**(10): p. 2054-66.

44. Lau, A.G., et al., *Distinct 3'UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF).* Proc Natl Acad Sci U S A, 2010. **107**(36): p. 15945-50.

45. Ji, Z. and B. Tian, *Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types.* PLoS One, 2009. **4**(12): p. e8419.

46. Shepard, P.J., et al., *Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq.* RNA, 2011. **17**(4): p. 761-72.

47. 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.

48. Orengo, J.P., et al., *Expanded CTG repeats within the DMPK 3' UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy.* Proc Natl Acad Sci U S A, 2008. **105**(7): p. 2646-51.

49. Park, J.Y., et al., *Comparative analysis of mRNA isoform expression in cardiac hypertrophy and development reveals multiple post-transcriptional regulatory modules.* PLoS One, 2011. **6**(7): p. e22391.

50. Creemers, E.E., et al., *Genome-Wide Polyadenylation Maps Reveal Dynamic mRNA 3'-End Formation in the Failing Human Heart.* Circ Res, 2016. **118**(3): p. 433-8.

51. 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.

52. Shi, Y., et al., *Molecular architecture of the human pre-mRNA 3' processing complex.* Mol Cell, 2009. **33**(3): p. 365-76.

53.   Tian, B. and J.L. Manley, *Alternative polyadenylation of mRNA precursors.* Nat Rev Mol Cell Biol, 2017. **18**(1): p. 18-30.

54.   Martin, G., et al., *Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length.* Cell Rep, 2012. **1**(6): p. 753-63.

55.   Neve, J., et al., *Cleavage and polyadenylation: Ending the message expands gene regulation.* RNA Biol, 2017. **14**(7): p. 865-890.

56.   Blazie, S.M., et al., *Alternative Polyadenylation Directs Tissue-Specific miRNA Targeting in Caenorhabditis elegans Somatic Tissues.* Genetics, 2017. **206**(2): p. 757-774.

57.   Lianoglou, S., et al., *Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression.* Genes Dev, 2013. **27**(21): p. 2380-96.

58.   Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-54.

59.   Wightman, B., I. Ha, and G. Ruvkun, *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.* Cell, 1993. **75**(5): p. 855-62.

60.   Vella, M.C. and F.J. Slack, *C. elegans microRNAs*, in *WormBook*, T.C.e.R. Community, Editor. September 21, 2005, WormBook.

61.   Pasquinelli, A.E., et al., *Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA.* Nature, 2000. **408**(6808): p. 86-9.

62.   Bohmert, K., et al., *AGO1 defines a novel locus of Arabidopsis controlling leaf development.* EMBO J, 1998. **17**(1): p. 170-80.

63.   Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data.* Nucleic Acids Res, 2011. **39**(Database issue): p. D152-7.

64.   Ha, M. and V.N. Kim, *Regulation of microRNA biogenesis.* Nat Rev Mol Cell Biol, 2014. **15**(8): p. 509-24.

65.   Ambros, V. and G. Ruvkun, *Recent Molecular Genetic Explorations of Caenorhabditis elegans MicroRNAs.* Genetics, 2018. **209**(3): p. 651-673.

66.   Lee, Y., et al., *MicroRNA genes are transcribed by RNA polymerase II.* EMBO J, 2004. **23**(20): p. 4051-60.

67.   Schwarz, D.S., et al., *Asymmetry in the assembly of the RNAi enzyme complex.* Cell, 2003. **115**(2): p. 199-208.

68.   Schirle, N.T. and I.J. MacRae, *The crystal structure of human Argonaute2.* Science, 2012. **336**(6084): p. 1037-40.

69.   Barca-Mayo, O. and Q.R. Lu, *Fine-Tuning Oligodendrocyte Development by microRNAs.* Front Neurosci, 2012. **6**: p. 13.

70.   Rogers, K. and X. Chen, *Biogenesis, turnover, and mode of action of plant microRNAs.* Plant Cell, 2013. **25**(7): p. 2383-99.

71.   Jin, H.Y. and C. Xiao, *MicroRNA Mechanisms of Action: What have We Learned from Mice?* Front Genet, 2015. **6**: p. 328.

72.   Eulalio, A., E. Huntzinger, and E. Izaurralde, *Getting to the root of miRNA-mediated gene silencing.* Cell, 2008. **132**(1): p. 9-14.

73.   Mathonnet, G., et al., *MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F.* Science, 2007. **317**(5845): p. 1764-7.

74.   Wakiyama, M., et al., *Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system.* Genes Dev, 2007. **21**(15): p. 1857-62.

75.   Chekulaeva, M., et al., *miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs.* Nat Struct Mol Biol, 2011. **18**(11): p. 1218-26.

76.   Giraldez, A.J., et al., *Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs.* Science, 2006. **312**(5770): p. 75-9.

77.   Bagga, S., et al., *Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation.* Cell, 2005. **122**(4): p. 553-63.

78. Chan, S.P. and F.J. Slack, *microRNA-mediated silencing inside P-bodies.* RNA Biol, 2006. **3**(3): p. 97-100.

79. Eichhorn, S.W., et al., *mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues.* Mol Cell, 2014. **56**(1): p. 104-15.

80. Tops, B.B., R.H. Plasterk, and R.F. Ketting, *The Caenorhabditis elegans Argonautes ALG-1 and ALG-2: almost identical yet different.* Cold Spring Harb Symp Quant Biol, 2006. **71**: p. 189-94.

81. Grishok, A., et al., *Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing.* Cell, 2001. **106**(1): p. 23-34.

82. Morita, S., et al., *One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation.* Genomics, 2007. **89**(6): p. 687-96.

83. Lu, K., et al., *Slicer Endonuclease Argonaute 2 Is a Negative Regulator of Hematopoietic Stem Cell Quiescence.* Stem Cells, 2016. **34**(5): p. 1343-53.

84. Ngondo, R.P., et al., *Argonaute 2 Is Required for Extra-embryonic Endoderm Differentiation of Mouse Embryonic Stem Cells.* Stem Cell Reports, 2018. **10**(2): p. 461-476.

85. Ludwig, N., et al., *Distribution of miRNA expression across human tissues.* Nucleic Acids Res, 2016. **44**(8): p. 3865-77.

86. Liang, Y., et al., *Characterization of microRNA expression profiles in normal human tissues.* BMC Genomics, 2007. **8**: p. 166.

87. Friedman, R.C., et al., *Most mammalian mRNAs are conserved targets of microRNAs.* Genome Res, 2009. **19**(1): p. 92-105.

88. Akhtar, M.M., et al., *Bioinformatic tools for microRNA dissection.* Nucleic Acids Res, 2016. **44**(1): p. 24-44.

89. Krek, A., et al., *Combinatorial microRNA target predictions.* Nat Genet, 2005. **37**(5): p. 495-500.

90.   Agarwal, V., et al., *Predicting effective microRNA target sites in mammalian mRNAs.* Elife, 2015. **4**.

91.   Ha, I., B. Wightman, and G. Ruvkun, *A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation.* Genes Dev, 1996. **10**(23): p. 3041-50.

92.   Didiano, D. and O. Hobert, *Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions.* Nat Struct Mol Biol, 2006. **13**(9): p. 849-51.

93.   Grimson, A., et al., *MicroRNA targeting specificity in mammals: determinants beyond seed pairing.* Mol Cell, 2007. **27**(1): p. 91-105.

94.   Wolter, J.M., et al., *3'LIFE: a functional assay to detect miRNA targets in high-throughput.* Nucleic Acids Res, 2014. **42**(17): p. e132.

95.   Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.

96.   Lall, S., et al., *A genome-wide map of conserved microRNA targets in C. elegans.* Curr Biol, 2006. **16**(5): p. 460-71.

97.   Paraskevopoulou, M.D., et al., *DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W169-73.

98.   Broughton, J.P., et al., *Pairing beyond the Seed Supports MicroRNA Targeting Specificity.* Mol Cell, 2016. **64**(2): p. 320-333.

99.   Betel, D., et al., *Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites.* Genome Biol, 2010. **11**(8): p. R90.

100.   Miranda, K.C., et al., *A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes.* Cell, 2006. **126**(6): p. 1203-17.

101.   Betel, D., et al., *The microRNA.org resource: targets and expression.* Nucleic Acids Res, 2008. **36**(Database issue): p. D149-53.

102.   Steinkraus, B.R., M. Toegel, and T.A. Fulga, *Tiny giants of gene regulation: experimental strategies for microRNA functional studies.* Wiley Interdiscip Rev Dev Biol, 2016. **5**(3): p. 311-62.

103.   Hafner, M., et al., *Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP.* Cell, 2010. **141**(1): p. 129-41.

104.   Chi, S.W., et al., *Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.* Nature, 2009. **460**(7254): p. 479-86.

105.   Helwak, A., et al., *Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding.* Cell, 2013. **153**(3): p. 654-65.

106.   Zisoulis, D.G., et al., *Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans.* Nat Struct Mol Biol, 2010. **17**(2): p. 173-9.

107.   Jin, Y., et al., *Evaluating the microRNA targeting sites by luciferase reporter gene assay.* Methods Mol Biol, 2013. **936**: p. 117-27.

108.   Wolter, J.M., et al., *Detection of miRNA Targets in High-throughput Using the 3'LIFE Assay.* J Vis Exp, 2015(99): p. e52647.

109.   Lee, Y.S. and A. Dutta, *MicroRNAs in cancer.* Annu Rev Pathol, 2009. **4**: p. 199-227.

110.   Calin, G.A., et al., *Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.* Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15524-9.

111.   Cimmino, A., et al., *miR-15 and miR-16 induce apoptosis by targeting BCL2.* Proc Natl Acad Sci U S A, 2005. **102**(39): p. 13944-9.

112.   Bonci, D., et al., *The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities.* Nat Med, 2008. **14**(11): p. 1271-7.

113.   Aqeilan, R.I., G.A. Calin, and C.M. Croce, *miR-15a and miR-16-1 in cancer: discovery, function and future perspectives.* Cell Death Differ, 2010. **17**(2): p. 215-20.

114.   Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

115.   Calin, G.A., et al., *Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2999-3004.

116.   Johnson, S.M., et al., *RAS is regulated by the let-7 microRNA family.* Cell, 2005. **120**(5): p. 635-47.

117.   Paranjape, T., et al., *A 3'-untranslated region KRAS variant and triple-negative breast cancer: a case-control and genetic analysis.* Lancet Oncol, 2011. **12**(4): p. 377-86.

118.   Chin, L.J., et al., *A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk.* Cancer Res, 2008. **68**(20): p. 8535-40.

119.   Esquela-Kerscher, A. and F.J. Slack, *Oncomirs - microRNAs with a role in cancer.* Nat Rev Cancer, 2006. **6**(4): p. 259-69.

120.   Sheedy, P. and Z. Medarova, *The fundamental role of miR-10b in metastatic cancer.* Am J Cancer Res, 2018. **8**(9): p. 1674-1688.

121.   Ma, L., J. Teruya-Feldstein, and R.A. Weinberg, *Tumour invasion and metastasis initiated by microRNA-10b in breast cancer.* Nature, 2007. **449**(7163): p. 682-8.
122.   Garofalo, M., et al., *miR221/222 in cancer: their role in tumor progression and response to therapy.* Curr Mol Med, 2012. **12**(1): p. 27-33.

123.   Song, J., et al., *Potential Value of miR-221/222 as Diagnostic, Prognostic, and Therapeutic Biomarkers for Diseases.* Front Immunol, 2017. **8**: p. 56.
124.   Iorio, M.V. and C.M. Croce, *MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review.* EMBO Mol Med, 2012. **4**(3): p. 143-59.

125.   Calin, G.A. and C.M. Croce, *MicroRNA signatures in human cancers.* Nat Rev Cancer, 2006. **6**(11): p. 857-66.

126.   Lan, H., et al., *MicroRNAs as potential biomarkers in cancer: opportunities and challenges.* Biomed Res Int, 2015. **2015**: p. 125094.

127.   Larrea, E., et al., *New Concepts in Cancer Biomarkers: Circulating miRNAs in Liquid Biopsies.* Int J Mol Sci, 2016. **17**(5).

128.   Schwarzenbach, H., D.S. Hoon, and K. Pantel, *Cell-free nucleic acids as biomarkers in cancer patients.* Nat Rev Cancer, 2011. **11**(6): p. 426-37.

129.   Kotagama, K., Y. Chang, and M. Mangone, *miRNAs as Biomarkers in Chronic Myelogenous Leukemia.* Drug Dev Res, 2015. **76**(6): p. 278-85.

130.   Chen, P.S., J.L. Su, and M.C. Hung, *Dysregulation of microRNAs in cancer.* J Biomed Sci, 2012. **19**: p. 90.

131.   Janssen, H.L., et al., *Treatment of HCV infection by targeting microRNA.* N Engl J Med, 2013. **368**(18): p. 1685-94.

132.   Rupaimoole, R. and F.J. Slack, *MicroRNA therapeutics: towards a new era for the management of cancer and other diseases.* Nat Rev Drug Discov, 2017. **16**(3): p. 203-222.

133.   Kent, W.J., *BLAT--the BLAST-like alignment tool.* Genome Res, 2002. **12**(4): p. 656-64.

134.   Liu, X., et al., *Cell-specific effects of miR-221/222 in vessels: molecular mechanism and therapeutic application.* J Mol Cell Cardiol, 2012. **52**(1): p. 245-55.

135.   Marsischky, G. and J. LaBaer, *Many paths to many clones: a comparative look at high-throughput cloning methods.* Genome Res, 2004. **14**(10B): p. 2020-8.

136.   Yang, X., et al., *A public genome-scale lentiviral expression library of human ORFs.* Nat Methods, 2011. **8**(8): p. 659-61.

137.   Schaefer, U., S. Schmeier, and V.B. Bajic, *TcoF-DB: dragon database for human transcription co-factors and transcription factor interacting proteins.* Nucleic Acids Res, 2011. **39**(Database issue): p. D106-10.

138.   Zhang, H.M., et al., *AnimalTFDB: a comprehensive animal transcription factor database.* Nucleic Acids Res, 2012. **40**(Database issue): p. D144-9.

139.   Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.

140.   Cook, K.B., et al., *RBPDB: a database of RNA-binding specificities.* Nucleic Acids Res, 2011. **39**(Database issue): p. D301-8.

141. Cormier, C.Y., et al., *Protein Structure Initiative Material Repository: an open shared public resource of structural genomics plasmids for the biological community.* Nucleic Acids Res, 2010. **38**(Database issue): p. D743-9.

142. Stinson, S., et al., *TRPS1 targeting by miR-221/222 promotes the epithelial-to-mesenchymal transition in breast cancer.* Sci Signal, 2011. **4**(177): p. ra41.

143. Fornari, F., et al., *MiR-221 controls CDKN1C/p57 and CDKN1B/p27 expression in human hepatocellular carcinoma.* Oncogene, 2008. **27**(43): p. 5651-61.

144. Ciafre, S.A., et al., *Extensive modulation of a set of microRNAs in primary glioblastoma.* Biochem Biophys Res Commun, 2005. **334**(4): p. 1351-8.

145. Galardi, S., et al., *miR-221 and miR-222 expression affects the proliferation potential of human prostate carcinoma cell lines by targeting p27Kip1.* J Biol Chem, 2007. **282**(32): p. 23716-24.

146. Garofalo, M., et al., *miR-221&222 regulate TRAIL resistance and enhance tumorigenicity through PTEN and TIMP3 downregulation.* Cancer Cell, 2009. **16**(6): p. 498-509.

147. le Sage, C., et al., *Regulation of the p27(Kip1) tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation.* EMBO J, 2007. **26**(15): p. 3699-708.

148. Chistiakov, D.A., et al., *Human miR-221/222 in Physiological and Atherosclerotic Vascular Remodeling.* Biomed Res Int, 2015. **2015**: p. 354517.

149. Easow, G., A.A. Teleman, and S.M. Cohen, *Isolation of microRNA targets by miRNP immunopurification.* RNA, 2007. **13**(8): p. 1198-204.

150. Chi, S.W., G.J. Hannon, and R.B. Darnell, *An alternative mode of microRNA target recognition.* Nat Struct Mol Biol, 2012. **19**(3): p. 321-7.

151. Baek, D., et al., *The impact of microRNAs on protein output.* Nature, 2008. **455**(7209): p. 64-71.

152. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

153. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.

154.   Johnson, C.D., et al., *The let-7 microRNA represses cell proliferation pathways in human cells.* Cancer Res, 2007. **67**(16): p. 7713-22.

155.   Lupini, L., et al., *miR-221 affects multiple cancer pathways by modulating the level of hundreds messenger RNAs.* Front Genet, 2013. **4**: p. 64.

156.   Wu, Y.H., et al., *The manipulation of miRNA-gene regulatory networks by KSHV induces endothelial cell motility.* Blood, 2011. **118**(10): p. 2896-905.

157.   Zongaro, S., et al., *The 3' UTR of FMR1 mRNA is a target of miR-101, miR-129-5p and miR-221: implications for the molecular pathology of FXTAS at the synapse.* Hum Mol Genet, 2013. **22**(10): p. 1971-82.

158.   Godshalk, S.E., et al., *A Variant in a MicroRNA complementary site in the 3' UTR of the KIT oncogene increases risk of acral melanoma.* Oncogene, 2011. **30**(13): p. 1542-50.

159.   Togliatto, G., et al., *MIR221/MIR222-driven post-transcriptional regulation of P27KIP1 and P57KIP2 is crucial for high-glucose- and AGE-mediated vascular cell damage.* Diabetologia, 2011. **54**(7): p. 1930.

160.   Ichimura, A., et al., *MicroRNA-34a inhibits cell proliferation by repressing mitogen-activated protein kinase kinase 1 during megakaryocytic differentiation of K562 cells.* Mol Pharmacol, 2010. **77**(6): p. 1016-24.

161.   Kneitz, B., et al., *Survival in patients with high-risk prostate cancer is predicted by miR-221, which regulates proliferation, apoptosis, and invasion of prostate cancer cells by inhibiting IRF2 and SOCS3.* Cancer Res, 2014. **74**(9): p. 2591-603.

162.   Hu, G., et al., *miR-221 suppresses ICAM-1 translation and regulates interferon-gamma-induced ICAM-1 expression in human cholangiocytes.* Am J Physiol Gastrointest Liver Physiol, 2010. **298**(4): p. G542-50.

163.   Zhang, X., et al., *Increased expression of microRNA-221 inhibits PAK1 in endothelial progenitor cells and impairs its function via c-Raf/MEK/ERK pathway.* Biochem Biophys Res Commun, 2013. **431**(3): p. 404-8.

164.   Lal, A., et al., *miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements.* Mol Cell, 2009. **35**(5): p. 610-25.

165.   Cevec, M., C. Thibaudeau, and J. Plavec, *NMR structure of the let-7 miRNA interacting with the site LCS1 of lin-41 mRNA from Caenorhabditis elegans.* Nucleic Acids Res, 2010. **38**(21): p. 7814-21.

166.   Seiler, C.Y., et al., *DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1253-60.

167.   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.

168.   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.

169.   Mayr, C., *Regulation by 3'-Untranslated Regions.* Annu Rev Genet, 2017. **51**: p. 171-194.

170.   Conne, B., A. Stutz, and J.D. Vassalli, *The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology?* Nat Med, 2000. **6**(6): p. 637-41.

171.   Delay, C., et al., *Alzheimer-specific variants in the 3'UTR of Amyloid precursor protein affect microRNA function.* Mol Neurodegener, 2011. **6**: p. 70.

172.   Rehfeld, A., et al., *Alterations in polyadenylation and its implications for endocrine disease.* Front Endocrinol (Lausanne), 2013. **4**: p. 53.

173.   Bartel, D.P., *MicroRNAs: target recognition and regulatory functions.* Cell, 2009. **136**(2): p. 215-33.

174.   Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**(6772): p. 901-6.

175.   Shin, C., et al., *Expanding the microRNA targeting code: functional sites with centered pairing.* Mol Cell, 2010. **38**(6): p. 789-802.

176.   Chen, K. and N. Rajewsky, *The evolution of gene regulation by transcription factors and microRNAs.* Nat Rev Genet, 2007. **8**(2): p. 93-103.

177.   Brown, K.C., et al., *ALG-5 is a miRNA-associated Argonaute required for proper developmental timing in the Caenorhabditis elegans germline.* Nucleic Acids Res, 2017. **45**(15): p. 9093-9107.

178. Landgraf, P., et al., *A mammalian microRNA expression atlas based on small RNA library sequencing.* Cell, 2007. **129**(7): p. 1401-14.

179. Eisenberg, I., et al., *Distinctive patterns of microRNA expression in primary muscular disorders.* Proc Natl Acad Sci U S A, 2007. **104**(43): p. 17016-21.

180. Jima, D.D., et al., *Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs.* Blood, 2010. **116**(23): p. e118-27.

181. Mangone, M., et al., *UTRome.org: a platform for 3'UTR biology in C. elegans.* Nucleic Acids Res, 2008. **36**(Database issue): p. D57-62.

182. Lee, M.-H. and T. Schedl, *RNA-binding proteins*, in *WormBook*, T.C.e.R. Community, Editor. April 18, 2006, WormBook.

183. Kudlow, B.A., L. Zhang, and M. Han, *Systematic analysis of tissue-restricted miRISCs reveals a broad role for microRNAs in suppressing basal activity of the C. elegans pathogen response.* Mol Cell, 2012. **46**(4): p. 530-41.

184. Jannot, G., et al., *GW182-Free microRNA Silencing Complex Controls Post-transcriptional Gene Expression during Caenorhabditis elegans Embryogenesis.* PLoS Genet, 2016. **12**(12): p. e1006484.

185. Alberti, C., et al., *Cell-type specific sequencing of microRNAs from complex animal tissues.* Nat Methods, 2018. **15**(4): p. 283-289.

186. McGhee, J.D., *The C. elegans intestine*, in *WormBook*, T.C.e.R. Community, Editor. March 27, 2007, WormBook.

187. Gieseler, K., H. Qadota, and G.M. Benian, *Development, structure, and maintenance of C. elegans body wall muscle*, in *WormBook*, T.C.e.R. Community, Editor. April 13, 2017, WormBook.

188. Frokjaer-Jensen, C., et al., *Improved Mos1-mediated transgenesis in C. elegans.* Nat Methods, 2012. **9**(2): p. 117-8.

189. Frokjaer-Jensen, C., et al., *Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon.* Nat Methods, 2014. **11**(5): p. 529-34.

190. Bukhari, S.I., et al., *The microRNA pathway controls germ cell proliferation and differentiation in C. elegans.* Cell Res, 2012. **22**(6): p. 1034-45.

191. Blazie, S.M., et al., *Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles.* BMC Biol, 2015. **13**: p. 4.

192. Qian, X., et al., *Enolase 1 stimulates glycolysis to promote chemoresistance in gastric cancer.* Oncotarget, 2017. **8**(29): p. 47691-47708.

193. DePina, A.S., et al., *Regulation of Caenorhabditis elegans vitellogenesis by DAF-2/IIS through separable transcriptional and posttranscriptional mechanisms.* BMC Physiol, 2011. **11**: p. 11.

194. Blackwell, T.K., et al., *SKN-1/Nrf, stress responses, and aging in Caenorhabditis elegans.* Free Radic Biol Med, 2015. **88**(Pt B): p. 290-301.

195. Smith-Vikos, T., et al., *MicroRNAs mediate dietary-restriction-induced longevity through PHA-4/FOXA and SKN-1/Nrf transcription factors.* Curr Biol, 2014. **24**(19): p. 2238-46.

196. Grosshans, H., et al., *The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in C. elegans.* Dev Cell, 2005. **8**(3): p. 321-30.

197. Heid, P.J., et al., *The zinc finger protein DIE-1 is required for late events during epithelial cell rearrangement in C. elegans.* Dev Biol, 2001. **236**(1): p. 165-80.

198. Ding, X.C., F.J. Slack, and H. Grosshans, *The let-7 microRNA interfaces extensively with the translation machinery to regulate cell differentiation.* Cell Cycle, 2008. **7**(19): p. 3083-90.

199. Sun, L., et al., *microRNAs Involved in the Control of Innate Immunity in Candida Infected Caenorhabditis elegans.* Sci Rep, 2016. **6**: p. 36036.

200. Moerman, D.G. and B.D. Williams, *Sarcomere assembly in C. elegans muscle*, in *WormBook*, T.C.e.R. Community, Editor. January 16, 2006, WormBook.

201. Enright, A.J., et al., *MicroRNA targets in Drosophila.* Genome Biol, 2003. **5**(1): p. R1.

202. Liu, M., et al., *Rho GTPase regulation by miRNAs and covalent modifications.* Trends Cell Biol, 2012. **22**(7): p. 365-73.

203. Chen, X., et al., *RhoC is a major target of microRNA-93-5P in epithelial ovarian carcinoma tumorigenesis and progression.* Mol Cancer, 2015. **14**: p. 31.

204. Farley, B.M., J.M. Pagano, and S.P. Ryder, *RNA target specificity of the embryonic cell fate determinant POS-1.* RNA, 2008. **14**(12): p. 2685-97.

205. Wu, Y., et al., *MicroRNA-214 regulates smooth muscle cell differentiation from stem cells by targeting RNA-binding protein QKI.* Oncotarget, 2017. **8**(12): p. 19866-19878.

206. Pillman, K.A., et al., *miR-200/375 control epithelial plasticity-associated alternative splicing by repressing the RNA-binding protein Quaking.* EMBO J, 2018. **37**(13).

207. Zhi, L., et al., *mir-355 Functions as An Important Link between p38 MAPK Signaling and Insulin Signaling in the Regulation of Innate Immunity.* Sci Rep, 2017. **7**(1): p. 14560.

208. Martinez, N.J., et al., *Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity.* Genome Res, 2008. **18**(12): p. 2005-15.

209. Vasquez-Rifo, A., et al., *Developmental characterization of the microRNA-specific C. elegans Argonautes alg-1 and alg-2.* PLoS One, 2012. **7**(3): p. e33750.

210. Porta-de-la-Riva, M., et al., *Basic Caenorhabditis elegans methods: synchronization and observation.* J Vis Exp, 2012(64): p. e4019.

211. Frokjaer-Jensen, C., et al., *Single-copy insertion of transgenes in Caenorhabditis elegans.* Nat Genet, 2008. **40**(11): p. 1375-83.

212. Moore, M.J., et al., *Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis.* Nat Protoc, 2014. **9**(2): p. 263-93.

213. Kurn, N., et al., *Novel isothermal, linear nucleic acid amplification systems for highly multiplexed applications.* Clin Chem, 2005. **51**(10): p. 1973-81.

214. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.

215.   Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol, 2010. **28**(5): p. 511-5.

216.   Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

217.   Evans (ed.), T.C., *Transformation and microinjection*, in *WormBook*, T.C.e.R. Community, Editor. April 6, 2006, WormBook.

218.   Kozomara, A. and S. Griffiths-Jones, *miRBase: annotating high confidence microRNAs using deep sequencing data.* Nucleic Acids Res, 2014. **42**(Database issue): p. D68-73.

219.   Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics.* Nucleic Acids Res, 2008. **36**(Database issue): p. D154-8.

220.   Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature.* Nucleic Acids Res, 2006. **34**(Database issue): p. D140-4.

221.   Griffiths-Jones, S., *The microRNA Registry.* Nucleic Acids Res, 2004. **32**(Database issue): p. D109-11.

222.   Keren, H., G. Lev-Maor, and G. Ast, *Alternative splicing and evolution: diversification, exon definition and function.* Nat Rev Genet, 2010. **11**(5): p. 345-55.

223.   Wahl, M.C., C.L. Will, and R. Luhrmann, *The spliceosome: design principles of a dynamic RNP machine.* Cell, 2009. **136**(4): p. 701-18.

224.   Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nat Genet, 2008. **40**(12): p. 1413-5.

225.   Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-6.

226.   Kim, M.S., et al., *A draft map of the human proteome.* Nature, 2014. **509**(7502): p. 575-81.

227.   Baralle, F.E. and J. Giudice, *Alternative splicing as a regulator of development and tissue identity.* Nat Rev Mol Cell Biol, 2017. **18**(7): p. 437-451.

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

229. Montes, M., et al., *RNA Splicing and Disease: Animal Models to Therapies.* Trends Genet, 2019. **35**(1): p. 68-87.

230. Venables, J.P., J. Tazi, and F. Juge, *Regulated functional alternative splicing in Drosophila.* Nucleic Acids Res, 2012. **40**(1): p. 1-10.

231. Wen, J., A. Chiba, and X. Cai, *Computational identification of tissue-specific alternative splicing elements in mouse genes from RNA-Seq.* Nucleic Acids Res, 2010. **38**(22): p. 7895-907.

232. Kuroyanagi, H., et al., *The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo.* Mol Cell Biol, 2007. **27**(24): p. 8612-21.

233. Xu, Q., B. Modrek, and C. Lee, *Genome-wide detection of tissue-specific alternative splicing in the human transcriptome.* Nucleic Acids Res, 2002. **30**(17): p. 3754-66.

234. Blencowe, B.J., *Alternative splicing: new insights from global analyses.* Cell, 2006. **126**(1): p. 37-47.

235. Tan, J.H. and A.G. Fraser, *The combinatorial control of alternative splicing in C. elegans.* PLoS Genet, 2017. **13**(11): p. e1007033.

236. Wu, J.Y. and T. Maniatis, *Specific interactions between proteins implicated in splice site selection and regulated alternative splicing.* Cell, 1993. **75**(6): p. 1061-70.

237. Chen, M. and J.L. Manley, *Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches.* Nat Rev Mol Cell Biol, 2009. **10**(11): p. 741-54.

238. Matlin, A.J., F. Clark, and C.W. Smith, *Understanding alternative splicing: towards a cellular code.* Nat Rev Mol Cell Biol, 2005. **6**(5): p. 386-98.

239. Caceres, J.F., et al., *Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors.* Science, 1994. **265**(5179): p. 1706-9.

240.   Zhu, J., A. Mayeda, and A.R. Krainer, *Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins.* Mol Cell, 2001. **8**(6): p. 1351-61.

241.   Boutz, P.L., et al., *MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development.* Genes Dev, 2007. **21**(1): p. 71-84.

242.   Tamburino, A.M., S.P. Ryder, and A.J. Walhout, *A compendium of Caenorhabditis elegans RNA binding proteins predicts extensive regulation at multiple levels.* G3 (Bethesda), 2013. **3**(2): p. 297-304.

243.   Egan, C.R., et al., *A gut-to-pharynx/tail switch in embryonic expression of the Caenorhabditis elegans ges-1 gene centers on two GATA sequences.* Dev Biol, 1995. **170**(2): p. 397-419.

244.   Marshall, S.D. and J.D. McGhee, *Coordination of ges-1 expression between the Caenorhabditis pharynx and intestine.* Dev Biol, 2001. **239**(2): p. 350-63.

245.   Ohno, G., et al., *Muscle-specific splicing factors ASD-2 and SUP-12 cooperatively switch alternative pre-mRNA processing patterns of the ADF/cofilin gene in Caenorhabditis elegans.* PLoS Genet, 2012. **8**(10): p. e1002991.

246.   Kabat, J.L., S. Barberan-Soler, and A.M. Zahler, *HRP-2, the Caenorhabditis elegans homolog of mammalian heterogeneous nuclear ribonucleoproteins Q and R, is an alternative splicing factor that binds to UCUAUC splicing regulatory elements.* J Biol Chem, 2009. **284**(42): p. 28490-7.

247.   Heintz, C., et al., *Splicing factor 1 modulates dietary restriction and TORC1 pathway longevity in C. elegans.* Nature, 2017. **541**(7635): p. 102-106.

248.   Masuda, K., et al., *Tissue- and age-dependent expression of RNA-binding proteins that influence mRNA turnover and translation.* Aging (Albany NY), 2009. **1**(8): p. 681-98.

249.   Aryal, B., et al., *MicroRNAs and lipid metabolism.* Curr Opin Lipidol, 2017. **28**(3): p. 273-280.

250.   Spartz, A.K., R.K. Herman, and J.E. Shaw, *SMU-2 and SMU-1, Caenorhabditis elegans homologs of mammalian spliceosome-associated proteins RED and fSAP57, work together to affect splice site choice.* Mol Cell Biol, 2004. **24**(15): p. 6811-23.

251.   Shin, C. and J.L. Manley, *Cell signalling and the control of pre-mRNA splicing.* Nat Rev Mol Cell Biol, 2004. **5**(9): p. 727-38.

252.   Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.

253.   Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nat Protoc, 2012. **7**(3): p. 562-78.

254.   Kamath, R.S. and J. Ahringer, *Genome-wide RNAi screening in Caenorhabditis elegans.* Methods, 2003. **30**(4): p. 313-21.

255.   Ahringer (ed.), J., *Reverse genetics*, in *WormBook*, T.C.e.R. Community, Editor. April 6, 2006, WormBook.

256.   Chuang, J.C. and P.A. Jones, *Epigenetics and microRNAs.* Pediatr Res, 2007. **61**(5 Pt 2): p. 24R-29R.

257.   Vasilatou, D., et al., *Epigenetic alterations and microRNAs: new players in the pathogenesis of myelodysplastic syndromes.* Epigenetics, 2013. **8**(6): p. 561-70.

APPENDIX A

GENES FROM THE H3'UTROME SCREENED USING THE 3'LIFE ASSAY

RefSeq ID	Alias	Predicted?				Validated target of miR-221/let-7c	Validation References (PMIDs)	RAS pathway	Top hit in the 3`LIFE screen	
		TargetScan [44]	PicTar [45]	DIANA microT [46]	miRanda				miR-221	let-7c
NM_000076	CDKN1C	miR-221	miR-221	miR-221	miR-221	miR-221	21461636	-	-	-
NM_001195259	TGFBR3L	-	-	-	-	-	-	-	-	-
NM_005252	FOS	miR-221	miR-221	miR-221	let-7c/miR-221	miR-221	20299489	Y	-	-
NM_001950	E2F4	-	-	-	-	-	-	-	-	-
NM_024832	RIN3	-	-	-	-	-	-	Y	-	-
NM_001271007	STRA13	-	-	-	-	-	-	-	-	-
NM_002467	MYC	-	-	-	let-7c	-	-	Y	-	-
NM_198256	E2F6	let-7c	-	let-7c	let-7c	-	-	-	-	-
NM_001789	CDC25A	let-7c	let-7c	let-7c	let-7c	let-7c	25909324	-	-	Y
NM_007241	SNF8	-	-	-	-	-	-	-	-	-

NM_004040	RHOB	-	-	-	-	-	-	Y	-	Y
NM_001665	RHOG	-	-	-	-	-	-	Y	-	-
NM_014245	RNF7	let-7c	let-7c	-	let-7c	-	-	-	-	Y
NM_002199	IRF2	miR-221	miR-221	miR-221	miR-221	miR-221	24607843	-	-	-
NM_014212	HOXC11	let-7c	let-7c	-	let-7c	-	-	-	-	-
NM_133639	RHOV	-	-	-	-	-	-	Y	-	-
NM_004310	RHOH	-	-	-	-	-	-	Y	-	-
NM_001039111	TRIM71	let-7c	-	let-7c	let-7c	let-7c	17890240	-	-	Y
NM_005343	HRAS	-	-	-	-	-	-	Y	-	-
NM_001191	BCL2L1	let-7c	let-7c	let-7c	let-7c	let-7c	23562878	Y	-	Y
NM_005080	XBP1	-	-	-	-	-	-	-	Y	-
NM_004958	FRAP1	-	-	-	let-7c/miR-221	-	-	-	Y	-

NM_007299	BRCA1	-	-	-	-	-	-	-	Y	-
NM_002228	JUN	-	-	-	-	-	-	-	-	-
NM_005225	E2F1	-	-	-	-	-	-	-	-	-
NM_001126115	TP53	let-7c	-	let-7c	let-7c	-	-	-	-	-
NM_198443	NRN1L	-	-	-	-	-	-	-	-	-
NM_005474	HDAC5	-	-	-	-	-	-	-	-	-
NM_001014835	PAK4	-	-	-	-	-	-	Y	-	Y
NM_001654	ARAF	-	-	-	let-7c	-	-	Y	-	-
NM_000201	ICAM1	-	-	-	-	miR-221	20110463	-	-	-
NM_004579	MAP4K2	-	-	let-7c	-	-	-	-	-	-
NM_001243926	MAPKAPK3	-	-	-	miR-221	-	-	Y	-	-
NM_032390	MKI67IP	-	-	-	-	-	-	-	-	-
NM_004535	MYT1	-	-	-	-	-	-	-	-	Y
NM_005904	SMAD7	-	-	-	-	-	-	-	-	-

NM_001143976	WEE1	miR-221	-	-	miR-221	miR-221	23630541	-	Y	-
NM_152688	KHDRBS2	miR-221	miR-221	miR-221	miR-221	-	-	-	-	-
NM_002755	MAP2K1	-	-	-	miR-221	-	-	Y	-	-
NM_005914	MCM4	-	-	-	-	-	-	-	-	Y
NM_153048	FYN	-	-	-	-	-	-	-	-	-
NM_003804	RIPK1	-	-	-	-	-	-	-	-	-
NM_015235	CSTF2T	-	-	-	-	-	-	-	-	Y
NM_003253	TIAM1	-	-	-	miR-221	-	-	Y	-	-
NM_032638	GATA2	-	-	-	miR-221	-	-	-	-	-
NM_015401	HDAC7A	-	-	-	-	-	-	-	-	-
NM_001145138	RELA	-	-	-	-	-	-	Y	-	-
NM_005923	MAP3K5	-	-	-	miR-221	-	-	-	-	-
NM_019884	GSK3A	-	-	-	-	-	-	-	Y	-
NM_024680	E2F8	-	-	-	-	-	-	-	-	-

NM_001128167	PAK3	-	-	-	let-7c	-	-	Y	-	Y
NM_015895	GMNN	-	-	-	-	-	-	-	Y	Y
NM_001256295	ETS2	miR-221	miR-221	-	miR-221	miR-221	21715310	Y	Y	-
NM_004556	NFKBIE	-	-	-	-	-	-	-	-	-
NM_001211	BUB1B	-	-	-	-	-	-	-	Y	Y
NM_001185081	FMR1	miR-221	-	miR-221	miR-221	miR-221	23390134	-	Y	-
NM_000222	KIT	miR-221	-	miR-221	miR-221	miR-221	21119596	Y	Y	Y
NM_005228	EGFR	-	-	-	-	-	-	Y	-	-
NM_001025243	IRAK1	-	-	-	-	-	-	-	-	-

NM_002915	RFC3	-	-	-	-	-	-	-	-	-
NM_021795	ELK4	-	-	let-7c	-	-	-	-	-	-
NM_006343	MERTK	-	-	-	-	-	-	-	Y	Y
NM_003668	MAPKAPK5	-	-	-	-	-	-	-	-	-
NM_005391	PDK3	-	-	-	-	-	-	-	-	Y
NM_001260	CDK8	-	-	-	miR-221	-	-	-	-	-
NM_001786	CDC2	-	-	-	-	-	-	-	-	-
NM_004526	MCM2	-	-	-	-	-	-	-	-	Y
NM_001128620	PAK1	let-7c/miR-221	miR-221/let-7c	-	let-7c/miR-221	miR-221	23333386	Y	-	-
NM_004333	BRAF	-	-	-	-	-	-	Y	-	Y
NM_020529	NFKBIA	-	-	-	-	-	-	Y	Y	Y
NM_002229	JUNB	-	-	-	let-7c	-	-	-	-	-
NM_002913	RFC1	-	-	-	let-7c	-	-	-	-	-

NM_002908	REL	-	-	-	-	-	-	Y	-	-
NM_002880	RAF1	-	-	-	-	-	-	Y	-	-
NM_012115	CASP8AP2	-	-	-	miR-221	-	-	-	-	-
NM_013254	TBK1	-	-	-	-	-	-	Y	-	-
NM_003954	MAP3K14	-	-	-	-	-	-	-	-	-
NM_004052	BNIP3	-	-	-	-	-	-	-	-	-
NM_003106	SOX2	-	-	-	-	-	-	-	-	-
NM_000346	SOX9	-	-	-	-	-	-	-	Y	-
NM_006914	RORB	-	-	miR-221	miR-221	-	-	-	-	Y
NM_005634	SOX3	-	-	-	-	-	-	-	-	-
NM_004189	SOX14	-	-	-	-	-	-	-	-	-
NM_018419	SOX18	-	-	-	-	-	-	-	-	-
NM_031439	SOX7	-	-	-	-	-	-	-	-	-
NM_001363	DKC1	-	-	-	-	-	-	-	-	-
NM_022454	SOX17	-	-	-	-	-	-	-	-	-