Characterization of Transcriptional and Post-transcriptional Regulation Of *Lin-42*/*Period* During Post-embryonic Development of *C. elegans*

Tracy James

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Biological Sciences

Diya Banerjee, chair
Daniela Cimini
Jeffrey R. Kuhn
Pablo Sobrado

September 11th 2012
Blacksburg, Virginia

Keywords: *C. elegans*, post-embryonic, transcriptional, *lin-42*, *ztf-16*, EMSA, circadian *period*, RNAi

Characterization of Transcriptional and Post-transcriptional Regulation of *Lin-42*/*Period* During Post-embryonic Development of *C. elegans*

Tracy James

Abstract

Period, which is broadly conserved in metazoans, regulates circadian timing of neurophysiology as well as cell fate specification. Studies in mouse and humans indicate that *period* functions as a tumor suppressor and controls adult stem cell differentiation. However, regulation of *period* function in developmental pathways has not been characterized and appears to be different from its regulation and function in circadian pathways. *lin-42* is the *Caenorhabditis elegans* ortholog of *period* and has both circadian and developmental timing functions. During post-embryonic larval development, cyclic expression and function of *lin-42* controls stage-specific and reiterative cell fate choices of a subset of epidermal stem cells called seam cells. We are studying *lin-42* regulation of seam cell fate during *C. elegans* larval development as a model for understanding the mechanisms of *period* regulation of adult stem cell fate in mammals.

This dissertation describes the research undertaken to characterize the *cis*-regulatory elements and the t*rans*-regulatory factors that control *lin-42* expression. We used direct molecular interaction assays (Electrophoretic Mobility Shift Assay, EMSA) (Chapter 2) followed by an RNA interference (RNAi)-based genetic screen (Chapter 3) to identify *lin-42* transcriptional regulators. Using the EMSA, we identified three 50 to 100 base pair regions

(binding regions, BR1-3) in the *lin-42* 5'noncoding sequences that were bound with specificity by *C. elegans* nuclear proteins. These binding regions represent putative *cis*-regulatory elements that may serve as transcription factor binding sites (TFBSs). We attempted to identify by mass spectrometry the proteins that bind to the BR sequences. We also used Phylogenetic Footprinting and bioinformatics screens to identify candidate *C. elegans* transcription factors (TFs) that may bind to putative TFBSs within the BR sequences. Using an RNAi-based screen, we tested the candidate TF genes for potential genetic interactions with *lin-42*. We identified ZTF-16, a member of the Hunchback/Ikaros zinc-finger transcription factor family, as a potential *lin-42* activator and, using quantitative real-time PCR, confirmed that *ztf-16* mutation results in down-regulation and loss of cycling expression of *lin-42*. We further determined that loss of *ztf-16* results in seam cell development defects that phenocopy *lin-42* loss-of-function, thus validating ZTF-16 as a transcriptional activator of *lin-42*.

Dedications

Dedicated to my baby girl.

Acknowledgements

At the outset I would like to express my sincere gratitude to my advisor and mentor, Diya Banerjee for her continued support and guidance throughout the course of completion of this dissertation. Her passion for the field and the depth of her knowledge are inspirational. Being the first graduate student of the lab, I had the good fortune of learning from her first-hand. I thank her for placing her trust in me and pushing me constantly to the next level.

I would like to specially thank the members of my committee, Drs. Daniela Cimini, Jeffrey Kuhn and Pablo Sobrado for their timely inputs, constructive criticism and understanding over the past five years.

I thank all the members of the Banerjee lab, especially Nathan Lee Roberts, who worked with me intensively over the last two years. Nate's unrelenting effort and commitment helped fuel the project forward at a critical time. I thank him with all my heart. Autumn Timpano- for teaching me the ropes and for her initial contribution to the project. Nimisha Khanduja- my friend and confidante, thank you for being there and making the last five years a time that I will remember fondly.

Words cannot convey my gratitude towards my parents- Mary James and James Verghese. Thank you both for breaking the mold of conventional parents and allowing your daughter to cross the seven seas in pursuit of her dreams. I know this day would not have been possible without your prayers and blessings. And finally, my husband, Santosh Abraham- for being my punching bag and shoulder to lean on- thank you!

Table of contents

Abstract ii

Dedications iv

Acknowledgements v

Table of contents vi

List of figures viii

List of tables x

CHAPTER 1 : Literature Review 1

SUMMARY 1
TEMPORAL CONTROL OF DEVELOPMENT IN *CAENORHABDITIS ELEGANS* BY HETEROCHRONIC GENES 2
DEVELOPMENTAL FUNCTIONS OF *lin-42* 10
CIRCADIAN FUNCTIONS OF *lin-42* 18
SYNOPSIS AND RATIONALE FOR PROJECT 25

CHAPTER 2 : Molecular Interaction Screen for *Cis*-regulatory Elements and Transcription Factors that Control *lin-42* Expression 36

SUMMARY 36
BACKGROUND AND RATIONALE 37
MATERIALS AND METHODS 44
RESULTS 47
DISCUSSION 55

CHAPTER 3 : Identification and Characterization of the *lin-42* Transcriptional Regulator *ztf-16* by RNAi Screen 73

SUMMARY 73
BACKGROUND AND RATIONALE 74
MATERIALS AND METHODS 81
RESULTS 83
DISCUSSION 94

CHAPTER 4 : Conclusions & Future Directions 111

REFERENCES 114

APPENDICES 126

APPENDIX A 126
Manuscript- *lin-42*/*Period* is Post-transcriptionally Regulated by *let-7* MicroRNAs During *C. elegans* Post-embryonic Stem Cell Development 126
APPENDIX B 174
Protocol for *C. elegans* Nuclear Extracts 174
APPENDIX C 176
Primers for EMSA Probes 176
APPENDIX D 178
C. elegans EMSA Protocol 178
APPENDIX E 180
C. elegans Homologs of *Drosophila* and Vertebrate Transcription Factors Predicted to Bind 180
DNA Motifs in *lin-42B*/*C* Promoter Regions BR1, BR2 and BR3 180
APPENDIX F 185
PCR Primers Used in Making RNAi Constructs 185
APPENDIX G 186
Genotyping Primers 186

List of figures

Figure 1.1: Life cycle of the nematode worm *C. elegans*. 28
Figure 1.2: Post-embryonic lineages of *C. elegans* hypodermal seam cells 29
Figure 1.3: Heterochronic mutations of seam cell lineages. 30
Figure 1.4: The *C. elegans* post-embryonic heterochronic pathway 31
Figure 1.5: *lin-42* isoforms 32
Figure 1.6: Expression profiles of *lin-42A*, *lin-42B* and *lin-42C* during larval stages L2 to L4 33
Figure 1.7: The *period* transcriptional/translational auto-regulatory feedback loop of the *Drosophila melanogaster* circadian timing pathway 34
Figure 2.1: Overview of eukaryotic regulatory DNA 59
Figure 2.2: Schematic of Electrophoretic Mobility Shift Assay (EMSA) 60
Figure 2.3: Western analysis of α-Tubulin and Histone H2A in cytoplasmic and nuclear fractions of *C. elegans* cellular extracts. 61
Figure 2.4: Optimization of cold competition EMSA conditions and effect of varying monovalent/divalent salt concentrations using control EMSA probe *myo-2*. 62
Figure 2.5: Cyclic expression of *lin-42B* over the course of larval development in *C. elegans* 63
Figure 2.6: UCSC Genome Browser view of *C. elegans lin-42* genomic region. 64
Figure 2.7: Identification of Binding Region 1 (BR1) by EMSA screen. 65
Figure 2.8: Identification of Binding Region 2 (BR2) by EMSA screen. 66
Figure 2.9: EMSA screen with probes 7 to 10 shows no specific binding of nuclear protein 67
Figure 2.10: Identification of Binding Region 3 (BR3) by EMSA screen. 68
Figure 2.11: Re-screening of BR1 with 50 bp EMSA probes. 69
Figure 2.12: Stage-specific EMSAs of BR2 and BR3. 70
Figure 3.1: Expression patterns of *let-7* family microRNAs during larval development and logic of the *let-7(lf)* lethality suppression/enhancement RNAi screen. 97
Figure 3.2: Flowchart of *let-7(lf)* lethality suppression/enhancement RNAi screen and interpretation of possible results 98
Figure 3.3: Phylogenetic Footprinting and results of *in silico* screens of *lin-42B/C* 5' non-coding sequences 99

Figure 3.4: Results of the *let-7(lf)* lethality suppression/enhancement RNA interference screen in *let-7(n2853)* and VT1066 mutant genetic backgrounds. 100
Figure 3.5: Expression of *lin-42* isoforms in wildtype and *ztf-16(lf)* genetic backgrounds. 101
Figure 3.6: *lin-42* isoform-specific expression over larval stages L2, L3 and L4 in N2/wildtype worms on RNAi of EV, *nhr-23* and *nhr-25* 102
Figure 3.7: Microarray expression data of mRNA levels of *ztf-16, nhr-23* and *nhr-25* compared to that of *lin-42* in N2 wildtype worms over time-course of larval development 103
Figure 3.8: Seam cell division patterns during larval development in wildtype and *ztf-16(ok3028)* mutants 104
Figure 3.9: Extent of gonad migration and the corresponding developmental age of the larval worm. 105
Figure 3.10: Seam cell fusion in wildtype and *ztf-16(ok3028)* mutants 106
Figure 3.11: Alae analysis across early-L4 to the adult stages in *ztf-16(ok3028)* versus wildtype *(wIs78*) worms. 107
Figure 3.12: *lin-42A, B and C* isoform levels in *sma-2(ok3019)* mutant and *sma-2* expression pattern over larval development. 108

List of tables

Table 1.1: *lin-42* loss-of-function mutant strains and *lin-42(RNAi)* 35

Table 2.1: Common components and variant components of EMSA binding buffer. 71

Table 2.2: *C. elegans* proteins identified by LC/MS-MS analysis of protein-DNA complex formed by EMSA of 200 bp probe 5 encompassing BR2 72

Table 3.1: List of 43 candidate *C. elegans* transcription factor genes that were tested in *let-7(lf)* lethality suppression/enhancement RNAi screen 109

Table 3.2: Summary of *let-7(lf)* lethality suppression/enhancement RNAi screen results. 110

CHAPTER 1 : Literature Review

lin-42, the *C. elegans* ortholog of *period*, and its role in the regulation of post-embryonic development

SUMMARY

Both spatial and temporal specificity is important for the normal development of multicellular organisms. In *C. elegans*, there exists a set of regulators known as the heterochronic genes that function to ensure that post-embryonic larval development occurs with temporal specificity. These temporal regulator genes are expressed and function at specific times during larval development to specify particular cell fates. One such heterochronic gene, *lin-42*, the *C. elegans* ortholog of the circadian timing gene *period*, is expressed cyclically throughout larval development and regulates cell division and differentiation. Like *period* in most animal species, *C. elegans lin-42* also regulates circadian rhythms in physiology and behavior in the adult worm. Recent studies have uncovered that mammalian *period* has developmental functions and controls adult stem cell differentiation and acts as a tumor suppressor. However, the molecular mechanisms of *period* regulation and function in this developmental capacity are not known. We are studying the regulation and function of *C. elegans lin-42* during post-embryonic larval development as a model for understanding the developmental functions of *period*. Specifically, we are studying the transcriptional regulation of *lin-42* in the larval epidermal seam cells, which are tissue or adult stem cells, in order to understand *period* regulation of stem cell development in vertebrates.

TEMPORAL CONTROL OF DEVELOPMENT IN *CAENORHABDITIS ELEGANS* BY HETEROCHRONIC GENES

Developmental Heterochrony

The development of multi-cellular organisms requires precision in regulation, given the need for coordination of multiple processes at the level of cells, tissues and the organism, and the dimension of time is a critical axis for development. Normal development depends on the accurate timing of cell division, cell-fate commitment and differentiation to occur within precise spatial boundaries. 'Heterochrony' is the term that describes the changes in the relative timing of developmental events, and includes relative variations in timing of a developmental event between a mutant and wild-type organism, or between distinct species [1]. Heterochronic genes can be thought of as the temporal equivalents of homeotic genes, in that while mutations in homeotic genes result in transformation as to *where* particular cell fates are expressed, mutations in heterochronic genes result in transformations as to *when* particular cell fates are expressed. In the roundworm *Caenorhabditis elegans (C. elegans)*, the heterochronic genes encode components of a molecular developmental timing mechanism. Like other developmental mutations, heterochronic defects alter cell fate. However, rather than altering spatial identity, sexual identity, or cell type, they disrupt the timing of expression of a specific cell fate. Two general phenotypes are seen in heterochronic mutants— 'precocious'- in which developmental events are skipped or occur earlier, and 'retarded'- in which they are repeated or occur later than they are supposed to [2]. Heterochronic mutations may affect specific tissue types (epidermal, neural, vulval, intestinal, etc.) and may also affect specific events during development (cell-division pattern, cell-cycle length, molting, terminal differentiation, etc.). These heterochronic genes may be stage-specifically expressed and repressed to ensure that developmental events occur with temporal specificity during larval development in *C. elegans* [3].

C. elegans as a model system for studying developmental timing

A number of features make *C. elegans* a suitable model organism for the study of development, specifically developmental heterochrony. *C. elegans* is a small (adults are approximately 1 mm) free-living, non-pathogenic soil nematode that has a short, temperature-dependent life cycle: *C. elegans* goes through a reproductive life cycle in 5.5 days at 15°C, in 3.5 days at 20°C, and in 2.5 days at 25°C. *C. elegans* has two sexes - hermaphrodites (XX) and males (X0). The hermaphrodites are self-fertilizing and produce about 300 progeny during their 2-week life spans. These self-fertilizing hermaphrodites allow for the rapid creation and maintenance of isogenic populations without the necessity of setting up crosses for each generation. However, the presence of rare males, which form by random non-disjunction of X chromosomes to form X0 individuals, allow for genetic out-crossing. In addition, the *C. elegans* genome is compact (~20,000 genes in 100 Mbps), has been completely sequenced and a wide variety of genetic molecular tools exist, such as RNA interference, (RNAi) libraries and gene-knockout libraries. Transgenes can be introduced and chromosomally integrated in *C. elegans* by microinjections or ballistic micro-bombardment [4-7]. The worms are also amenable to RNAi induced gene-knockdown by feeding them on dsRNA-expressing bacteria, soaking them in dsRNA and dsRNA injections. Most significantly, for the study of developmental timing, the complete somatic cell lineage, that is the developmental fate of every somatic cell (959 in the adult hermaphrodite; 1031 in the adult male), has been traced from the single cell embryo to the adult worm. Another attractive feature is that the *C. elegans* epidermis is un-pigmented, and thus cellular development can be monitored *in vivo* and in real-time by microscopy. A final advantage to working with *C. elegans*, and a feature that is unique among the currently used set of

multicellular eukaryotic model organisms, is that these nematodes can be kept in cryogenic suspension and revived after months to years of storage [8].

Post-embryonic larval development in *C. elegans*

The *C. elegans* life cycle comprises of embryonic, larval and adult developmental stages (**Figure 1.1**). Post-embryonic larval development consists of four stages (L1 to L4), each separated by a molt, which is the shedding of an old cuticle that is pushed off and replaced by a newly synthesized cuticle. A number of precursor or blast cells (approximately 50 to 55 cells in the newly hatched larva) continue development after the completion of embryonic development, and divide according to a schedule associated with progression through the larval stages. The post-embryonic cell lineages derived from these precursor blast cells exhibit stage-specific patterns of cell division and cell-fate determination, i.e. they divide in invariant temporal and spatial patterns during the four larval stages and give rise to a fixed number of cells with determined fates [9]. Somatic issues that develop during the larval stages include the epidermis, the vulva, neural cells, muscles and the somatic gonad. Transition from the larval stage to the adult stage is marked by terminal differentiation of all somatic blast cell lineages to form the reproductively competent adult animal. In the *C. elegans* adult, only the germ-line tissues that produce oocytes and sperm continue to divide and differentiate while all somatic cells types are terminally differentiated. The post-embryonic larval development of epidermal Seam Cells and vulval cells serve as model tissues for the study of molecular mechanisms that control temporal cell fate identity.

Epidermal Seam Cells as a model tissue for studying the temporal regulation of cell fate

Regulatory cues during development must be tightly coordinated with cell division programs to give rise to an appropriate number of cells before cell cycle exit and terminal differentiation. This is exemplified by the tight balance between cell proliferation and differentiation in stem cell maintenance, achieved by asymmetric cell division in which one daughter cell differentiates and the other remains proliferative [10]. The molecular mechanisms by which this division pattern is regulated or how the cell makes the choice between self-renewal and differentiation is still poorly understood. The hallmark features of stem cells, namely self-renewal and asymmetric division, are exhibited by the *C. elegans* epidermal Seam Cells, and the temporal progression of cell fates in these epidermal cell lineages represents a simple model for stem cell lineages in general.

The *C. elegans* epidermis, which is often referred to as the hypodermis, is the outermost cell layer and is responsible for synthesizing and secreting a collagen cuticle that contains and protects the other tissues of the worm and form ridges, called alae, that aid in locomotion. The seam cells are a type of adult or tissue stem cells that arranged in two rows along the left and right lateral midlines of the animal. After completion of embryonic development, 10 seam cells (H0, H1-2, V1-6 and T) enter into larval development and undergo stage-specific patterns of cell division and differentiation to form 16 terminally differentiated seam cells in the adult worm that secrete adult-specific alae (**Figure 1.2**). During each larval stage, the majority of the seam cells divide in classic stem cell pattern showing asymmetric division with self-renewal. During seam cell asymmetric division, the posterior daughter cell maintains proliferative capacity and continues the seam cell lineage, i.e. it self-renews, while the anterior daughter cell migrates from the seam and differentiates to form part of the Hyp7 hypodermal syncytium that covers most of

the animal. At the end of larval development, during the L4 to adult molt, the seam cells switch to their adult developmental program as they terminally differentiate, exit permanently from the cell cycle, fuse along their lateral axes and synthesize the collagenous ridges of adult alae, which distinguish the adult worm cuticle morphologically from larval stage cuticles [8]. While most seam cells show asymmetric division with self-renewal at every larval stage, there are stage-specific variations to this general pattern. For example, at the L1 stage V1-V4 and V6 seam cells divide once, whereas they divide twice in the L2 stage (L2-specific proliferative division) and then resume the single division pattern through L3 and L4 [8]. (**Figure 1.2**) Genetic screens in *C. elegans* have identified heterochronic genes that when mutated cause alterations to the invariant patterns of stage-specific seam cell division and differentiation during larval development. Studying how these heterochronic genes function to control the correct timing of larval seam cell development in *C. elegans* will allow us to understand temporal regulation of animal stem cell development in general.

Regulation of Seam Cell division by heterochronic genes during larval development

In *C. elegans*, a dedicated set of genes, components of the heterochronic pathway, control the timing of various stage-specific events during larval development. Mutations in heterochronic genes cause these stage-specific events (cell division pattern, terminal differentiation, alae formation, etc.) to occur prematurely (precocious mutant) or to be delayed (retarded mutant) relative to developmental progression in unaffected tissues [11, 12] (**Figure 1.3**) For example, loss-of-function mutations of heterochronic genes like *lin-41* or *lin-42* cause precocious phenotypes in which seam cell terminal differentiation is executed prematurely in the L4 stage instead of at the adult stage. In *lin-28(lf)* mutants, the L2-specific proliferative division

is skipped altogether and seam cells precociously execute the L3 and L4-specific divisions and terminally differentiate at the L4 stage. Conversely, mutations of the *let-7* microRNAs (*mir-48, mir-241 and mir-84)* retard this differentiation event [13-17].

Tissues affected by heterochronic mutations include the hypodermis, including seam cells, the vulva, the somatic gonad, muscle cells and neural cells [18-22]. Mutations in heterochronic genes therefore affect the development of multiple tissues, indicating that the heterochronic pathway is not tissue specific and that it may function to coordinate cell fate programs in various tissues during post-embryonic development. The products of the heterochronic genes include transcriptional and translational regulators, nuclear hormone receptors and small regulatory RNAs (microRNAs) [3]. Although the complete network of genetic and molecular interactions among these genes has not been mapped, epistasis and suppressor mutation analyses have helped to order these genes into the *C. elegans* post-embryonic heterochronic pathway. (**Figure 1.4**) The majority of known heterochronic genes are either expressed or repressed stage-specifically to determine specific cell fates, while some, such as *lin-46* and *lin-42*, are cyclically expressed throughout the larval stages [2, 23]. In the following sections, key heterochronic genes that affect seam cell development will be described.

lin-4, *lin-14* and *lin-28* regulation of early (L1/L2) larval development

lin-4 (abnormal cell lineage), the first gene found to encode a microRNA, is also the most upstream gene identified in the heterochronic pathway. MicroRNAs are small 22 to 28 nucleotide long non-coding regulatory RNAs that are found in most animal phyla. MicroRNAs downregulate translational expression of a target gene by binding to complementary sequences in the 3'UTR of the target gene transcript [24, 25]. The mechanism of down-regulation by miRs is

not entirely clear but can involve both translational repression and mRNA cleavage and degradation [26]. *lin-4* was the first molecule to be identified as a miR and the description of *lin-4* interactions with its target genes established a context for RNA-mediated gene silencing [27].

lin-4 loss-of-function (lf) mutants show retarded larval development, i.e. they reiterate the L1 seam cell division pattern, execute extra larval molts and do not form adult specific alae [28]. *lin-4* functions to post-transcriptionally down regulate the heterochronic genes *lin-14* and *lin-28* [29, 30]. (**Figure 1.4**) LIN-14 levels are high during the early L1 stage and are required for L1 stage division pattern to occur. After the L1 divisions, LIN-14 levels drop concomitant with increasing levels of *lin-4* to allow progression to the L2 stage, and thus allow L2 proliferative divisions to take place [30, 31]. If LIN-14 levels are not repressed after L1, the L1 division pattern is reiterated in place of the normal L2 to L4 division patterns, resulting in a reduced number of seam cells compared to wildtype, and other developmental defects including impaired locomotion and sterility [32, 33]. *lin-28,* the second target of *lin-4* microRNA, encodes a RNA binding protein whose activity is required during the L2 stage for normal development [34].

<u>*hbl-1* and regulation of the transition from early (L1/L2) to late (L3/L4) larval development</u>

Hunchback-like-1 (*hbl-1*), the *C. elegans* homolog of the HUNCHBACK (HBL) transcription factor, regulates seam cell fates at the transition from early to late larval development by maintaining L2 cell fates and repressing L3 cell fates. Therefore, *hbl-1* levels need to be repressed in order for progression from L2 to L3. At this developmental transition point, two parallel independent pathways converge on *hbl-1* to repress its levels, one involving *lin-4*/*lin-28*/*lin-46* and the other involving *daf-12*/*let-7* family microRNAs *mir-48, mir-84* and *mir-241* (referred to as *let-7-fam)* [14, 35]. (**Figure 1.4**) At the end of the L2 stage, the

expression of *lin-28* declines by the repressive activity of miR *lin-4* [15, 29]. This leads to de-repression of the gene immediately downstream, *lin-46*, which acts to specify L3 cell fates by inhibiting *hbl-1* expression [35]. The second pathway, independent of the first, involves the *let-7*-fam miRs (*mir-48, mir-84* and *mir-241*) that start to be expressed during L1 stage and are expressed maximally at the L2 stage, and function together redundantly to post-transcriptionally inhibit *hbl-1* [14]. The expression of the *let-7*-fam miRs is induced by nuclear receptor DAF-12 and its steroidal ligand, Dafachronic Acid, by direct activation of the promoters of the *let-7-fam* miRs [36]. Therefore, *mir-48, mir-84,* and *mir-241* function to control appropriate timing of seam cell division and L2-to-L3 transition through 3'UTR-mediated down-regulation of *hbl-1* [37]. Loss-of-function mutations of *lin-46*, the *let-7* family miRs and *daf-12*, all lead to reiteration of the L2-specific proliferative division of the seam cells at the L3 stage due to persistence of HBL-1 expression [14, 35, 38]. Thus, at the L3 stage, repressed levels of HBL-1 lead to the transition from the early to the late larval developmental stages. The complete inhibition of *hbl-1* at this point is also necessary to de-repress the transcription of the miR *let-7*, whose expression is requisite at the late larval stages to regulate transition into the adult stage [39].

let-7, *lin-41* and *lin-29* regulation of the transition from late larval (L3/L4) to adult development

By the end of the L3 stage, *hbl-1* expression is completely turned off and this leads to de-repression of *let-7* miR [39]. Thus, *let-7* expression beginning at L3 and peaking during L4 [16] allows for post-transcriptional down-regulation of a host of transcription factors that specify the late larval stages, thus controlling the transition from larval to adult stage [39]. These transcription factors include *lin-41, lin-42* and *daf-12* among others [16, 17, 40]. (**Figure 1.4**)

This up-regulation of *let-7* miR in the later stages is critical for the resultant down-regulation of LIN-41 which in turn relieves the inhibition of *lin-29*, a zinc-finger transcription factor that is responsible for the larval to adult switch by specifying adult-collagen gene expression, cessation of molting cycle and seam cell-fusion [12]. LIN-41 is believed to prevent translation of *lin-29* by directly binding to *lin-29* mRNA and preventing its translation in earlier larval stages. *lin-29* is thus the most downstream gene in the heterochronic pathway and is considered the most direct regulator of the larval-adult switch processes [17, 41-43]. *let-7(lf)* and *lin-29 (lf)* mutations delay or prevent the larval to adult transition and result in reiteration of larval specific seam cell division pattern at the adult stage.

Temporal expression of key transcription factors, such as HBL-1 and LIN-29, regulated by temporal expression of the miRs *lin-4* (L1 stage), *let-7* family miRs (L2/L3 stage) and *let-7* (L3/L4 stage) together constitute the heterochronic pathway that regulates seam cell development through the *C. elegans* larval stages.

DEVELOPMENTAL FUNCTIONS OF *lin-42*

The heterochronic genes described in the section above have temporally distinct functions and specify stage-specific cell fates during larval development. An exception is the heterochronic gene *lin-42*, which is cyclically expressed with each larval stage and has both stage-specific and reiterative developmental functions [13, 44]. Mutational analysis of *lin-42* has revealed that it regulates at least three different stage-specific developmental events in *C. elegans* larvae. i) *lin-42* activity during the L2 stage is required for L2-proliferative divisions of the seam cells, ii) *lin-42* activity during L4 stage is required to maintain L4 seam cell fates and inhibit adult fates, and iii) repression of *lin-42* during early L2 stage is required for commitment to

dauer, a developmentally quiescent state that is an alternative to continuing with reproductive development [13, 44-46]. *lin-42* also has reiterative functions and regulates the periods of lethargus and molting in between each of the larval stages [47]. In the following sections, we describe what is known of each of these functions of *lin-42* during *C. elegans* larval development.

lin-42 isoforms, expression and mutations

The *C. elegans lin-42* genomic region on chromosome II consists of 12 exons that are alternatively spliced or expressed from alternative promoters to give rise to three isoforms: *lin-42A, lin-42B* and *lin-42C*. (**Figure 1.5**) These *lin-42* isoforms not only have different 5' and 3' regulatory sequences but also encode proteins that differ in domain structure. For example, LIN-42B and LIN-42C contain PAS domains, which function to mediate protein-protein interactions, while LIN-42A lacks this domain. These structural differences suggest that LIN-42 protein isoforms may have different functions.

lin-42 expression, as detected by antibody staining and *lin-42* transcriptional reporter analyses, can be detected in many tissues of the developing worm, including the hypodermis, vulva, somatic gonad, muscles and intestines [44]. *lin-42* is also expressed during each of the four larval stages [13, 44]. This broad spatial and temporal expression pattern is consistent with a role for *lin-42* in coordinating developmental programs among various tissues and consistent with reiterative functions for *lin-42* during larval development. However, *lin-42* isoforms have different expression patterns. (Chapters 3 and 4) (**Figure 1.6**) *lin-42A* and *lin-42B* transcripts are cyclically expressed during larval development, with peaks of expression during inter-molts, i.e. during larval stages, and repression of expression during lethargus, when larvae molt. However,

lin-42A and *lin-42B* expression is repressed upon terminal differentiation at the end of larval development, with no expression during the adult stage. (Banerjee Lab, unpublished) In contrast, *lin-42C* is expressed at a basal level during larval development but is upregulated during the adult stage. (Banerjee Lab, unpublished) These differences in expression pattern, combined with differences in isoform protein structure, suggest that there may be isoform-specific differences in function. This possibility is supported by the fact that *lin-42* mutants that are impaired for different combinations of *lin-42* isoforms have different phenotypes. (**Table 1.1**) For example, *lin-42(n1089)* and *lin-42(mg152)* mutant strains produce functional LIN-42A, and show precocious seam cell differentiation and retain their eggs due to vulval developmental defects (egl phenotype, egg laying defective), but are otherwise wildtype [44]. In contrast, *lin-42(ok2385)* does not produce any functional LIN-42A or LIN-42B and has severe developmental abnormalities, including larval lethality [47]. (**Table 1.1**) Indeed, as we describe in the following sections, *lin-42A* and *lin-42B* together regulate larval seam cell division and differentiation [13, 44], *lin-42A* regulates lethargus and molting [47], and *lin-42C* likely regulates circadian timing of physiology and behavior in adult *C. elegans* (Banerjee Lab, unpublished).

<u>*lin-42* regulation of seam cell division and differentiation</u>

Mutational analysis indicates that *lin-42* controls seam cell fate both early and late during larval development. *lin-42* loss-of-function mutants (**Table 1.1**) display a strong precocious seam cell differentiation phenotype [44]. (**Figure 1.3**) Seam cells normally terminally differentiate at the end of the L4 stage and express adult specific alae as young adults. In *lin-42 (lf)* mutants, seam cells precociously exit the cell cycle after the L3 molt, fuse along their lateral axes and start forming alae during the L4 stage instead of the adult stage. This precocious phenotype indicates

that *lin-42* activity is required during the L4 stage to allow expression of L4 cell fates and to prevent transition to the adult stage. In mutant strains that do not produce *lin-42B* transcripts but can produce *lin-42A* (strains *n1089, mg152*, **Table 1.1**), L1, L2 and L3 seam cell division patterns are wildtype, suggesting that *lin-42* function is confined to the L4 stage. However, L2-stage specific defects are revealed if *lin-42(lf)* mutation is combined with a temperature sensitive *lin-14(lf)* mutation *lin-14(n179)* that shows wildtype seam cell development at the permissive temperature. These *lin-42(lf);lin-14(lf)* double mutants skip the L2-specific proliferative divisions, even at the permissive temperature, much like *lin-28* null mutants [12, 18, 45]. This phenotype indicates that *lin-42B* activity is required for the L2-specific proliferative division event, and that *lin-42A* activity is not sufficient for the L2-specific seam cell divisions. Taken together the L2 and L4 seam cell phenotypes in *lin-42 (lf)* mutants indicate that *lin-42* regulates stage-specific cell fates, specifically L2-specific symmetric seam cell division patterns and L4-specific seam cell terminal differentiation.

lin-42 regulation of the choice between reproductive development and dauer

Another stage-specific function of *lin-42* is as a critical component of the switch that regulates the larval decision between normal reproductive development and developmental arrest or dauer [46]. Its function in this capacity suggests a role for *lin-42* as a link between environmental cues and developmental responses.

Dauer versus reproductive development in C. elegans: In response to unfavorable environmental conditions, *C. elegans* larvae can undergo reversible developmental arrest at the second molt. These arrested developmental variants of the worm, referred to as dauer larvae, exhibit distinct morphological and behavioral characteristics compared to normal developing

larvae. Dauers are radially constricted and possess specialized cuticle with alae. Their oral orifices are closed by an internal plug and they have constricted pharynxes that do not pump. These dauer-specific morphological features make dauer larvae easily distinguishable from normal larvae [48, 49]. The dauer developmental arrest decision takes its cue from unfavorable environmental conditions, namely, high population density, food shortage and high temperature. Either one or a combination of these environmental cues at the L1 stage causes worms to suspend normal reproductive development after the L2 stage and to form dauers that are equipped for long-term survival [46]. From a genetic standpoint, the decision to go into developmental arrest means activation of L2-dauer genetic pathways (turning on dauer- specific genes) and the suspension of L2-L3 transition developmental timing pathways (turning off L3-specific genes). Although a number of proteins have been identified that regulate one program or the other, the nuclear hormone receptor transcription factor DAF-12 has been shown to be a critical contributing factor to both [50].

Overview of Dauer pathways: At the L1 stage, if *C. elegans* worms encounter unfavorable environmental conditions, the worms enter developmental arrest. This is in response to inputs from a number of signaling pathways - the Insulin (IG), TGF-β and Dafachronic Acid (DA) signaling pathways [46]. DAF-12 integrates the signals from these pathways and acts as a molecular switch that either promotes dauer formation and blocks reproductive development or inhibits dauer and promotes normal L2 developmental fates and the progression of reproductive development. Under optimal growth conditions, the Insulin, TGF-β and DA pathways convert dietary cholesterol into the steroid hormone, Dafachronic Acid (DA), the ligand for DAF-12 steroid hormone receptor. DA binds to DAF-12 to form a transcriptional complex that inhibits entry into dauer and allows for L2 stage developmental programs. Alternatively, under poor

dietary or growth conditions, DA synthesis is inhibited, and DAF-12 associates with the SHARP co-repressor homolog DIN-1 to form a transcriptional repressor complex. The DAF-12::DIN-1 repressor complex inhibits L2 stage developmental programs, and allows activation of the dauer developmental program [51, 52].

lin-42 is repressed during entry to dauer: Several regulators of the L2 versus L3 cell fate decision in the heterochronic pathway have been shown to play a role in influencing dauer decision, including *daf-12, hbl-1* and the *let-7* family miRs [52, 53]. *lin-42* – also a known L2-L3 regulator has been shown to play a major role as a stress-responsive negative regulator of dauer formation by acting antagonistically to DA ligand-free DAF-12. Together, LIN-42 and DAF-12 ensure appropriate developmental adaptation in stressful environments [46]. The balance between DA ligand-free and DA ligand-bound DAF-12 controls entry into dauer. Favorable growth conditions lead to optimal synthesis of DA resulting in ligand-bound DAF-12, which promotes reproductive development and dauer inhibition. Under these conditions, the presence of LIN-42 protein is necessary for L2 and L3 cell fates [13, 46]. At the time of stress-induced reduction of DA synthesis, LIN-42 is hypothesized to be the crucial factor in the decision to enter dauer or not. This model is based on experiments that show that *lin-42(lf)* is sufficient to induce dauer formation in the absence of environmental stress [46]. Conversely, in worm strains with mutations in the dauer induction pathway, forced expression of *lin-42* results in dauer formation even in the absence of environmental stress [46]. The model proposed from these studies is as follows: In the event of transient or mild stress that does not necessitate dauer entry, LIN-42 is hypothesized to prevent formation of the DAF-12::DIN-1 complex that initiates dauer. However, in the event of prolonged stress that necessitates dauer induction, both LIN-42 as well as DA levels drop in order for the shift in equilibrium to the ligand-free DAF-12::DIN-1

complex, which only in the absence of LIN-42 promotes entry into dauer and inhibits L3 transition. Thus, the dialogue between LIN-42 and DAF-12 further fine tunes the response to environmental stress by preventing unnecessary entry into dauer in the event of mild stress, thereby elevating the robustness of the system in a constantly changing environment. These studies also indicate a conserved role for *Per* family proteins like LIN-42 in responding to environmental stress [46].

lin-42 genetic interactions with daf-12: The role of *lin-42* in dauer seems to be as a promoter of continuous reproductive development under conditions of mild stress that does not necessitate dauer entry. In contrast, its levels need to be significantly repressed in order for the worm to enter into dauer when conditions of high stress mandate dauer development in order for survival [46]. The role of *lin-42* in this capacity is through dialogue with the ligand-bound form of DAF-12. Genetic Yeast Two Hybrid assays indicate that the B and C isoforms of LIN-42 interact with DAF-12 *in vitro* and the strength of this interaction is dependent on the presence of the DAF-12 ligand-binding domain (LBD). This interaction is suggestive of a possible complex formation that might transcriptionally activate dauer [13, 44]. Besides dauer decision, *lin-42* and *daf-12* functions also intersect during larval development of the worm. They both have heterochronic phenotypes in various tissues, including seam cells, vulva and somatic gonad [13, 38] *daf-12* alleles play a general role in advancing L2 stage programs and *daf-12* mutants show retarded heterochronic phenotypes - L2-specific cellular programs of division are repeated at the L3 stage and somatic gonad morphogenesis and migration is delayed. *lin-42* mutants on the other hand show precocious phenotypes [46]. When *lin-42* was depleted in various *daf-12* mutant backgrounds, the experiments revealed that a straightforward epistatic relationship does not exist

between these two genes. Rather, these two genes, *lin-42* and *daf-12*, are thought to act in parallel pathways that integrate environmental and cellular signals to control cell fate.

lin-42 regulation of lethargus and ecdysis

The roles of *lin-42* during larval development described thus far involve *lin-42A* and lin-*42B* regulation of larval stage-specific events. However, both *lin-42A* and *lin-42B* are cyclically expressed throughout larval development (**Figure 1.6**) and *lin-42A* in particular has a reiterative role in regulating lethargus and ecdysis [47]. The end of each larval stage in *C. elegans* is marked by a molt, at which point a stage-specific fresh cuticle is synthesized and the old one is shed. The process of molting entails the separation of the cuticle from the hypodermis, followed by formation of the new cuticle, and finally the shedding of the old cuticle or ecdysis. Before the larva progresses into the molt phase, it enters into a period of brief quiescence called lethargus. During this period, the worm stops feeding, pharyngeal pumping ceases, locomotion stops and the cuticle loosens. Thereafter, at ecdysis, the worm makes a hole in the cold cuticle at the head region through which it emerges, shedding the old cuticle and starts feeding immediately [8]. Thus, the cuticle is renewed every 8 to 10 hours at 25°C, concomitant with the completion of each larval stage, and molting occurs a total of four times before the worm enters adulthood. Although many heterochronic genes when mutated affect the number of times the larvae molt, no mutations have been reported to affect the timing of the molts itself [16, 41, 54] *lin-42* expression cycles in phase with the larval molts and particularly the cycling expression of the *lin-42A* isoform seems to be essential for rapid, rhythmic and productive molting cycles [34]. Recently published work has shown that *lin-42A* null mutant worms (*lin-42(ok2385)*, **Table 1.1**) that exhibit prolonged and variable duration of the larval stages were restored to wildtype phenotype by an extra-chromosomal array carrying the *lin-42A* coding sequence. The same study

also showed that forced over expression of the *lin-42A* isoform resulted in anachronistic larval molts and lethargy [47]. Thus, the *lin-42A* isoform appears to play a major and specific role in the timing and execution of the rhythmic molts during development of the larva. It is interesting to note that lethargus has been molecularly defined as being similar to sleep in vertebrate animals [55]. Thus, a direct parallel can be drawn between *lin-42* developmental function and *period* circadian function in regulating sleep/wake patterns. In *C. elegans*, the cyclic expression of *lin-42* appears to correlate with and is necessary for the pattern of active cell division and differentiation during each larval stage, followed by a period of developmental quiescence during lethargus. This is similar to cyclic expression of *period* that correlates with and is necessary for active physiology and movement during daylight hours in many animals, followed by inactivity and sleep during the night. In the next section, we explore this connection between the developmental timing regulator *lin-42* and the circadian timing regulator *period* in more detail.

CIRCADIAN FUNCTIONS OF *lin-42*

Interestingly, *lin-42* is the homolog of *period*—a core component of the molecular circadian timing pathway that controls the 24-hour periodicity of many physiological processes and behaviors [13]. Like *period*, *lin-42* encodes a protein with the PAS domain, a protein-protein interaction domain that is found in a number of core circadian pathway molecules. The PAS domain is thus named because of its presence in PER, ARNT (aryl hydrocarbon receptor nuclear translocator) and SIM (*Drosphila single-minded*) proteins. Several PAS proteins contain bHLH (Basic helix-loop-helix) motifs N-terminal to the PAS domain, however, *lin-42* and *period* both lack bHLH motifs in their N-termini. The *period* PAS domain consists of direct repeats of 50

amino acids called the PASA and PASB repeats. The PASB repeats of *lin-42* and *period* are the most similar [13]. Both *period* and *lin-42* are cyclically expressed. However, while *C. elegans lin-42* expression oscillates with an approximate 8 to 10 hours periodicity that correlates with sequential developmental stages, *period* expression in insects and vertebrate animals oscillates with a 24-hour periodicity that correlates with the day/night light cycle [44]. In addition to the similarities in structure and expression profile, *lin-42* and *period* also share similarity in function. Period has been extensively studied as a regulator of neuro-endocrine physiology and behavior, while *C. elegans lin-42* has been defined as a regulator of cell fate during post-embryonic development. However, *lin-42* has recently been found to regulate circadian rhythms of behavior in adult *C. elegans*, while evidence is mounting that *period* has developmental functions, specifically in regulating cell proliferation and adult stem cell differentiation in mammals [56-59]. In the following sections, we describe what is known of the circadian functions of *lin-42* and the developmental functions of *period*.

The circadian timing pathway and the role of *period*

One class of biological temporal regulators exemplified by the *C. elegans* heterochronic genes times the onset of developmental events. A second class of biological temporal regulators comprises the 'clock' components that regulate 24-hour periodicity, or circadian periodicity, of biological processes, such as sleep-wake cycles in humans. These 24-hour 'clock' regulators comprise the circadian pathway, which controls most aspects of animal physiology and behavior, and which is widely conserved across metazoan phyla. *period* was one of the first genes of the circadian pathway identified by forward genetic screens in *Drosophila,* and *period* loss of function mutants have aberrant circadian phenotypes of either long or short circadian rhythms

and are completely arrhythmic [60, 61]. Subsequent genetic and molecular analyses over the last two decades have identified other circadian regulators in both flies and vertebrates and a molecular mechanism of how the circadian clock works has emerged. Described below is a simplified model of the circadian clock centered on regulation of *period*. The animal circadian clock consists of transcriptional activators, CLOCK (CLK) and CYCLE (CYC) in *Drosophila* or BMAL in mammals, which turn on expression of *period (per)* and *timeless (tim)* in *Drosophila* or *cryptochrome (cry)* in mammals. (**Figure 1.7**) This transcriptional activation and subsequent translation into protein constitutes the positive arm of the circadian cycle. In the cytoplasm, the stability and localization of PER is dependent on its phosphorylation status, which is regulated by kinases, such as casein kinase I (CKIe), and phosphatases, such as Protein Phosphatase 2A (PP2A) [62]. Other post-translational modifications that regulate circadian clock molecules include acetylation, phosphorylation, ubiquitination and sumoylation [63]. PER phosphorylation by CKIe protects PER from degradation and allows its association with TIM or CRY. The PER-CKIe-TIM/CRY complex is translocated into the nucleus where it represses the CLK-CYC/BMAL complex and thus shuts off transcription of the *period* gene. This auto-feedback of PER constitutes the negative arm of the circadian cycle [63]. Hence, both PER and TIM/CRY serve as repressors whose activity is controlled by temporal changes in their states of phosphorylation leading to a transcription-translational feedback loop. The pace of the molecular circadian clock is therefore set by the temporal lag from the activation of the auto-repressors (PER and TIM), their translation, and nuclear accumulation to subsequent degradation, which requires about 24 hours [60].

lin-42 and *C. elegans* circadian rhythms

It has been widely believed that *C. elegans* and other short-lived nematodes do not possess

circadian rhythms. However, a number of studies have now established the existence of circadian rhythms in *C. elegans*. The earliest of such studies showed that starved L1 larvae have circadian rhythmicity in resistance to hyperosmotic stress [64]. Studies published in parallel showed that adult *C. elegans* have an overt temperature-compensated circadian rhythm in locomotor activity following entrainment to light and dark cycles [65]. Adult *C. elegans* also exhibit circadian rhythms in response to abiotic stimuli like osmotic and oxidative stress, thus demonstrating that stress responses vary daily and providing further evidence of rhythmic gene expression patterns that control these behaviors in *C. elegans* [66]. Genome-wide expression profiling experiments have identified light- and temperature-entrained oscillating transcripts in *C. elegans* that exhibited rhythmic expression with temperature-compensated 24-hour periods, suggesting that these transcripts are under circadian regulation [67]. This growing body of evidence suggests that there exists a molecular machinery sustaining circadian rhythms in nematodes. *lin-42* has been shown to regulate circadian patterns of behavior in adult *C. elegans* [68]. *lin-42(lf)* mutants exhibit a significant increase in the circadian period of locomotion in adult *C. elegans* as compared to wildtype worms [68]. Furthermore, genetic rescue of *lin-42(lf)* mutation restores normal circadian period, indicating that the larval developmental timer *lin-42* also has a circadian/physiological function in the worm [68].

In addition to *lin-42/period,* the *C. elegans* genome contains sequence homologs of many of the core circadian clock genes of *Drosophila* and mammals [13, 69]. However, none of these seems to play a role with respect to circadian function in the worm, as determined by lack of circadian timing defects on mutation of these genes. What is known of the *C. elegans* circadian molecular machinery suggests that it is somewhat divergent from the conserved circadian pathway that has been extensively studied in *Drosophila* and vertebrate species [64, 65]. Cycling

gene expression pattern is a key characteristic of proteins involved in circadian function. Both core circadian regulators like *period*, *timeless*, *clock* and *cycle*, as well as target genes controlled by the core circadian regulators, show cyclic patterns of expression in vertebrates and invertebrates [69]. Genome-wide expression profiling has revealed that while many genes show a cyclic expression pattern with 24-hour periodicity during adult stages, *C. elegans* orthologs of *clock* and *cycle*, and *lin-42/period* itself, are not expressed with circadian periodicity during either larva or adult stages [67]. Moreover, neither *tim-1*, the homolog of *timeless,* nor *kin-20,* the homolog of *CKI,* has a cyclic expression pattern during larval development, although both interact genetically with *lin-42* to regulate larval developmental timing [69]. These differences between *C. elegans* and other animals suggest that the molecular mechanisms by which *lin-42* controls circadian patterns of locomotion and larval developmental timing do not involve a transcription-translation loop with negative feed-back regulation.

Developmental functions of *period*

The role of *period* in circadian control of physiology has been well established. Mammalian *Per1, Per2* and *Per3* are core clock protein genes that are key regulators of circadian rhythms both in the central clock in the hypothalamus and in peripheral tissues [60, 70]. More recent findings have implicated *period* as a regulator of stem cell differentiation and as a cancer tumor suppressor [56, 57, 70-74]. Although the mechanism of *period* function in circadian rhythms is well studied and understood, its role in regulating cell division and development is poorly understood from a mechanistic stand-point. However, there is growing *in vivo* and *in vitro* evidence implicating the *period* genes in regulating adult stem cell fates and the growth of tumors [56, 57, 75, 76].

Circadian rhythms are daily patterns of various biological processes directed by endogenous clocks. The circadian system consists of peripheral oscillators in most tissues of the body and a central pacemaker located in the suprachiasmatic nucleus (SCN) of the hypothalamus [77, 78]. Genes that function in the circadian system and their protein products constitute the molecular components that are regulated through positive/negative feedback loops to generate circadian rhythms. Beyond the primary clock genes, circadian regulation extends to include various clock-controlled genes including cell-cycle genes and tumor suppressors and oncogenes [79, 80]. Thus, aberrant expressions of clock genes can directly or indirectly affect genomic instability, cell proliferation, and cell cycle control, ultimately promoting tumorigenesis [78]. Both the circadian and cell cycle clocks, intrinsic to most cells, share a number of mechanistic and conceptual similarities– they are both based on auto-regulatory loops and rely on transcription, translation, protein-modification and degradation [81]. Although the circadian clock controls expression of cell-cycle genes, it can function independently of the cell cycle [79, 81].

period as a tumor suppressor: The core circadian gene *period* plays a key role in tumor suppression by regulating DNA damage-responsive pathways [59, 71, 82, 83]. Mice deficient in *mPer2* showed increased susceptibility to developing tumors and reduced apoptosis in thymocytes. The temporal expression patterns of various cell cycle genes (involved in cell proliferation and tumor suppression) like *cyclin D1, cyclin A, mdm-2* and *gadd45α* were shown to be deregulated in *mPer*2 mutants [82]. Additionally, transcription of the gene *c-myc* (involved in cell division and apoptosis) was found to be upregulated and that of the gene *p53* (plays a critical role in G1-S checkpoint) was found to be downregulated in *Per2* mutant mice [82]. Thus, over-expression of *c-myc* allows cells to progress through checkpoints despite DNA damage and

decreased levels of *p53* leads to impaired apoptosis and ultimately tumor development owing to accumulation of damaged cells [82, 84]. Consistent with this observation, over expression of *period* in human cancer cell lines increased their sensitivity to DNA damage and apoptosis; in contrast, down regulation of *period* was associated with protection against ionizing radiation induced apoptosis [71]. Deregulated *period* expression has been implicated in various forms of cancer like hepatic cancer and pancreatic cancers [85], human gliomas [58], colorectal cancers [86] and breast cancers [75].

period regulation of adult stem cell development: Growing evidence over the recent years has established the role of *period* in the development of different types of stem cells. Recent work has unveiled a functional link between the non-oscillating, constitutive expression of the circadian gene *mPer2* and the intrinsic control of neural progenitor cells (NPCs) proliferation, cell death and neurogenesis in the adult dentate gyrus of adult mice. It was shown in this study that functional impairment of *mPer2* increased both the number of proliferating NPCs and the production of immature newborn neurons in the DG of adult mice [56]. *Per2* gene expression in neural stem cells was also shown to be associated with a suppression of their proliferative capacity [87]. Another publication reported that glucocorticoids-mediated repression of *Per3* is critical for advancing mesenchymal stem cells to the adipocyte fate, thereby establishing the role of *period* as a potent mediator of stem cell fates [76]. *Per1* expression in osteoblasts was shown to be activated by parathyroid hormone (PTH) signaling, involving *PerI* in regulating transcriptional events of bone metabolism [88]. Furthermore, osteoblasts lacking *Per* genes, tend to have a higher bone mass [89]. Development of hematopoietic stem cells and progenitor cells was also shown to be affected by circadian rhythms, particularly in regeneration of the stem cell niche [90]. Additionally, human as well as mouse hematopoietic stem cells show clear circadian

rhythms in *period* gene expression in the bone marrow, demonstrating that *period* may be developmentally regulated in these stem cells [91, 92].

SYNOPSIS AND RATIONALE FOR PROJECT

C. elegans lin-42 is cyclically expressed during post-embryonic larval development and has stage-specific and reiterative functions. Stage-specific functions include regulation of symmetric and asymmetric seam cell division patterns during L2 stage, seam cell terminal differentiation during L4 stage and the choice between reproductive development and dauer at early L2 stage [13, 44]. Reiterative functions involve regulating the timing of lethargus and ecdysis to take place at the completion of each of the four larval stages [47]. In each of these roles, the isoform-specific expression patterns of *lin-42* are of paramount importance. *lin-42B* expression is required to allow L2 and L4 stage-specific developmental programs to occur but needs to be downregulated on the completion of the larval stage in order to allow progression to the next stage. Similarly, *lin-42A* needs to be expressed during each larval stage but has to be repressed to allow lethargus and ecdysis to take place between stages. Finally, expression of both *lin-42A* and *lin-42B* needs to be completely inhibited to enable either entry into dauer at L2 stage or terminal differentiation at late L4 stage. Therefore, the cyclic *lin-42* expression pattern is not merely a consequence of the larval developmental stages but instead drives the period of these stages and enables stage-specific developmental programs to take place.

How is the temporally dynamic *lin-42* expression pattern generated and regulated? *lin-42* is the *C. elegans* ortholog of *period* and, like *period*, regulates both physiological and developmental pathways, is cyclically expressed and at the molecular level is regulated by molecules such as CKIe that function in the circadian pathway [13, 44, 68, 93]. Thus, it is

reasonable to propose that *lin-42* cyclic expression is controlled by mechanisms similar to those that control *period* expression, i.e. a cell autonomous molecular oscillation based on a transcription/translation negative feedback loop that is driven by cell extrinsic signals. However, there are two pieces of evidence that suggest that *lin-42* mechanisms of expression in the heterochronic pathway is likely to be different from control of *period* expression in the circadian pathway. First, there appears to be no negative auto-regulation of *lin-42* expression – *lin-42A* and *lin-42B* transcripts continue to cycle during larval development in *lin-42(mg152)* mutants that do not produce any functional LIN-42B protein [13]. (**Table 1.1**) Second, the *C. elegans* homologs of *Clock* and *Cycle/Bmal*, the circadian pathway transcription factors that control *period* expression, do not regulate *lin-42* transcription during larval development. (Chapter 3) Additionally, the signaling pathways that control the circadian timing pathway are likely different from the signaling pathways that control the heterochronic pathway. The mammalian central circadian clock in the SCN that regulates physiology and behavior is regulated by extrinsic, environmental stimuli such as light and temperature that are transduced by cellular signaling pathways. The *C. elegans* heterochronic pathway is regulated by developmental signals that are intrinsic to the animal, although the developmental timing pathway may be modified by environmental signals, as exemplified by the ability to switch to the dauer stage. Together these facts argue that *lin-42,* expression is controlled by transcriptional regulators other than *Clock* and *Cycle/Bmal* that function to integrate and transduce intrinsic and extrinsic signals. We hypothesize that *period* is also regulated by a different set of signaling pathways and transcriptional regulators in its capacity as a developmental regulator compared to *period* regulation in the circadian pathway. Our goal is to identify and characterize the transcriptional regulators of *lin-42* in *the C. elegans* larval heterochronic pathway that controls seam stem cell

development, and subsequently to use this knowledge to understand *period* regulation and function in mammalian stem cell development.

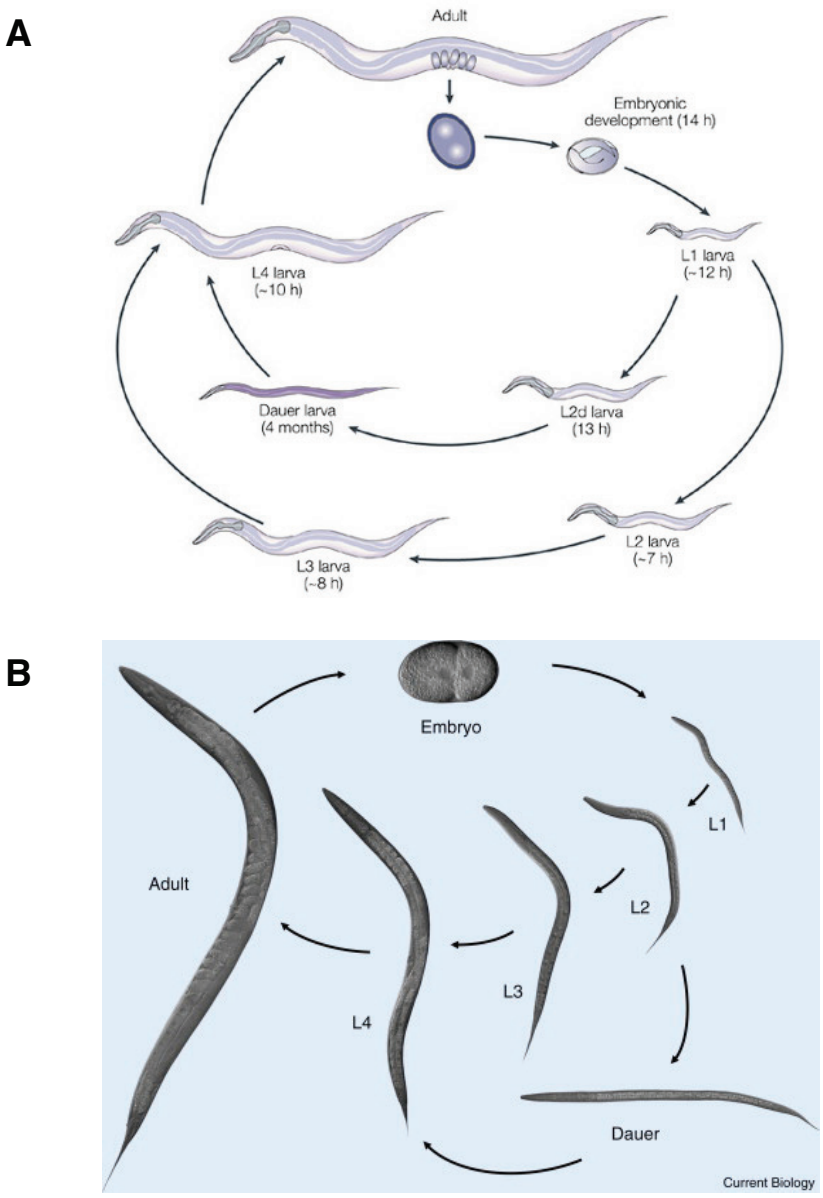


Figure 1.1: Life cycle of the nematode worm *C. elegans*.

After embryogenesis and hatching, *C. elegans* develop through four larval stages (L1 L2, L3 and L4) that culminate in the reproductively capable adult stage. Larval worms grow in overall body size during each of the larval stages resulting in an adult worm that is approximately 1 mm long. Under adverse environmental conditions, L1 larvae enter into an alternate state consisting of L2d and dauer. The dauer stage is reversible and worms reenter active development at the L4 stage when environmental conditions improve. The complete C. elegans life cycle takes about 2.5 weeks but dauer larvae can survive for up to 4 months. **(A)** The numbers in parentheses indicate the approximate length of each stage at 20-22°C. Diagram taken from [94]. **(B)** Diagram taken from [95].

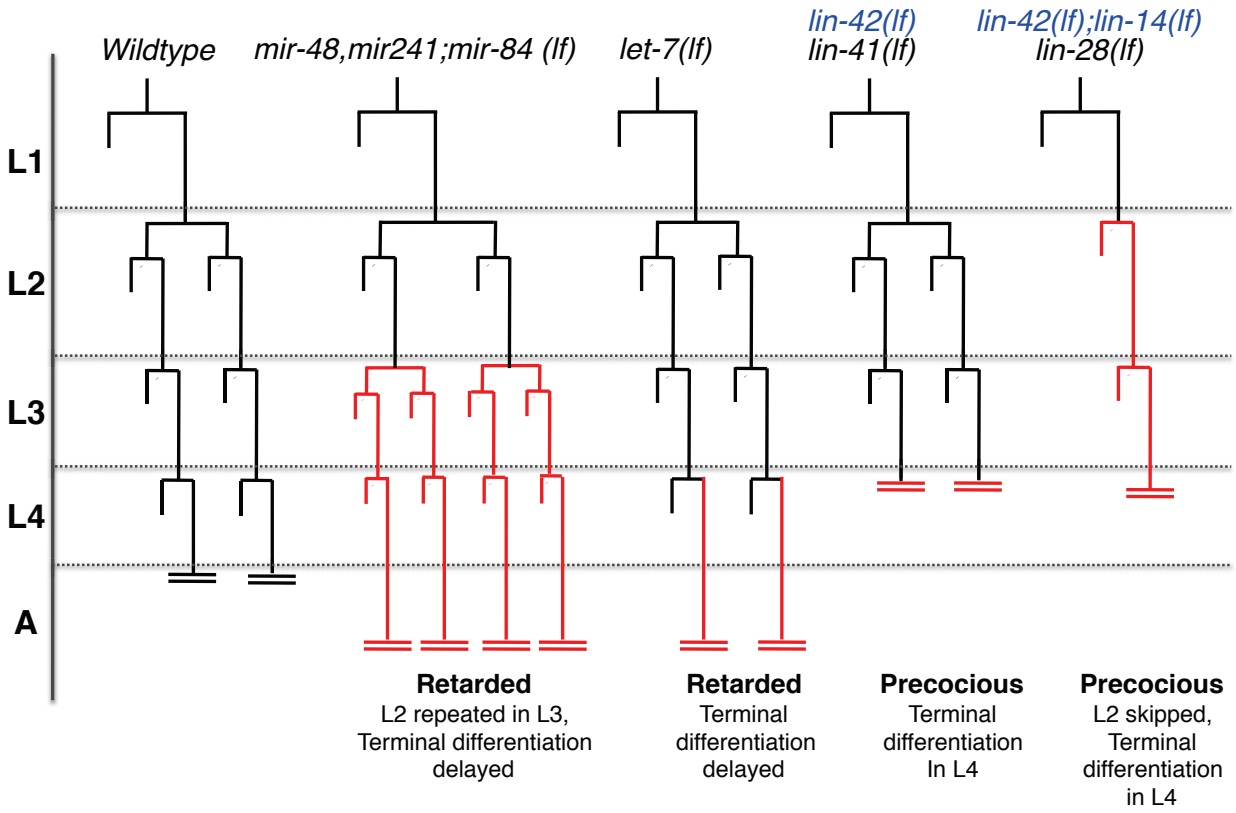


Figure 1.2: Post-embryonic lineages of *C. elegans* hypodermal seam cells.

Retarded and precocious developmental alterations to lineage of V1-4, V6 seam cells caused by loss-of-function heterochronic gene mutations. Aberrant seam cell division and differentiation patterns are highlighted in red. L1-L4: larval stages 1 to 4, A: adult stage, double horizontal bars: alae secretion, terminal differentiation. Figure compiled from [14-17].

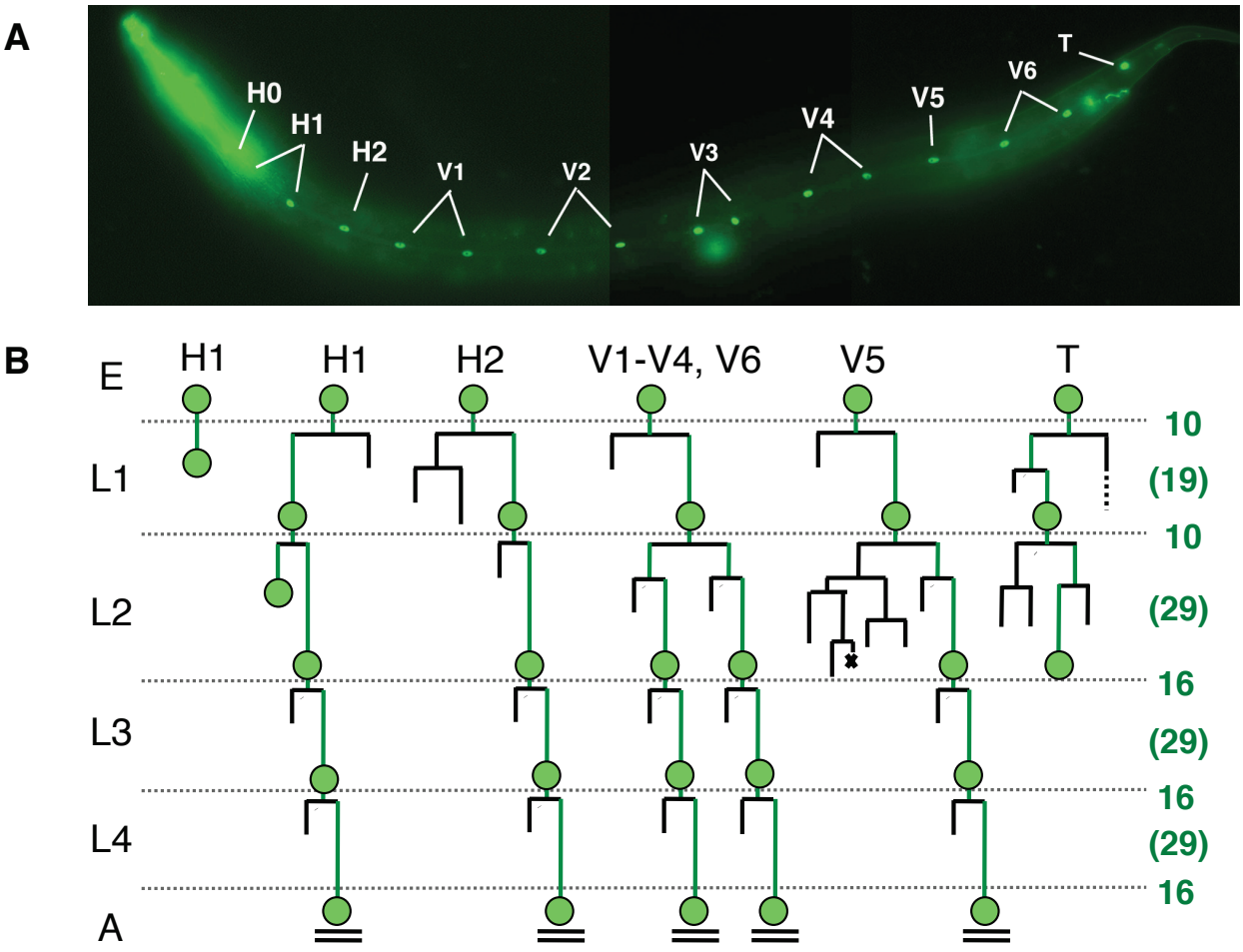


Figure 1.3: Heterochronic mutations of seam cell lineages.

(A) Adult wildtype *C. elegans* worm in which seam cell nuclei have been have visualized with a GFP reporter (strain *wIs78*). The positions of the 16 adult seam cells are marked. Figure from [93].
(B) Larval lineages of hypodermal seam cells. Ten seam cells (green dots, H0-H2, V1-V6 and T), formed during embryogenesis (E), undergo stage-specific symmetric and asymmetric divisions during the four larval stages (L1-L4) to give rise to 16 terminally differentiated seam cells in the adult worm (A) that secrete collagenous ridges called alae (double horizontal bars). Horizontal dotted lies represent transition points between stages during which larvae enter lethargus and undergo ecdysis or molting. During each larval stage, the majority of seam cells undergo asymmetric division with self-renewal, giving rise to a daughter cell that migrates away from the seam and differentiates (black vertical lines) or undergoes apoptosis (marked wit X), and another daughter cell that does not differentiate and retains seam cell identity, i.e. self-renews (green vertical lines). Numbers in green indicate the number of seam cells at the beginning and end of each stage, and green number in parentheses indicate the approximate number of GFP nuclei during seam cell division at every larval stage. Figure adapted from [9].

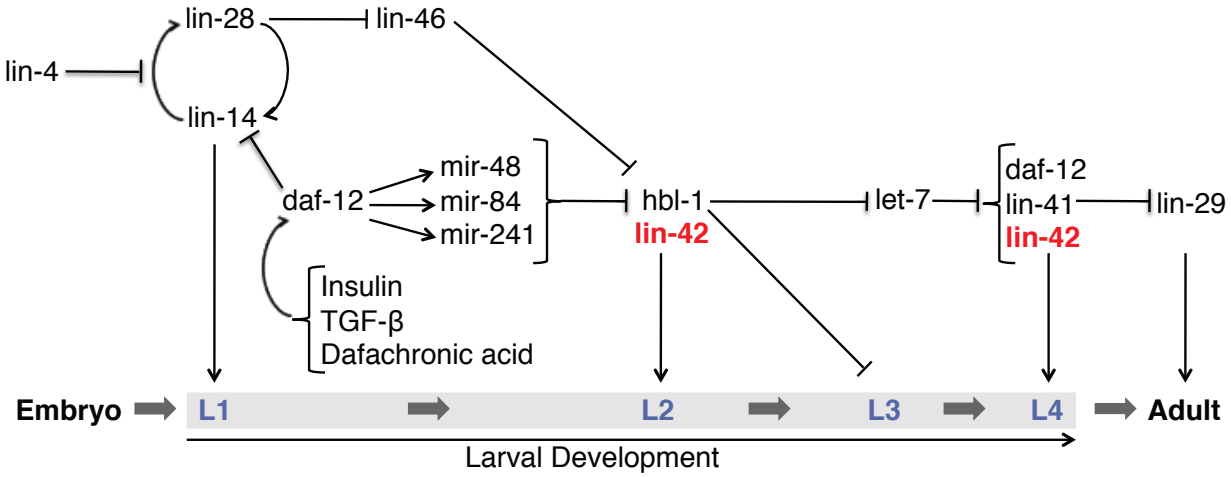


Figure 1.4: The *C. elegans* post-embryonic heterochronic pathway

Heterochronic gene products consist of transcriptional and translational regulators (*lin-14*, *lin-28*, *hbl-1*, *lin-41*, *lin-29*), nuclear hormone receptor transcription factors (*daf-12*), microRNAs (*lin-4*, *mir-48*, *mir-84*, *mir-241*, *let-7)* and other molecules with undefined molecular function (*lin-46*, *lin-42*). L1-L4: larval stages 1 to 4. Genes are positioned according to their approximate times of function relative to the larval stages. Lines with arrowheads (→) indicate activation and lines with bars (⊣) indicate repression. Refer to text for description of the relationships among the heterochronic genes.

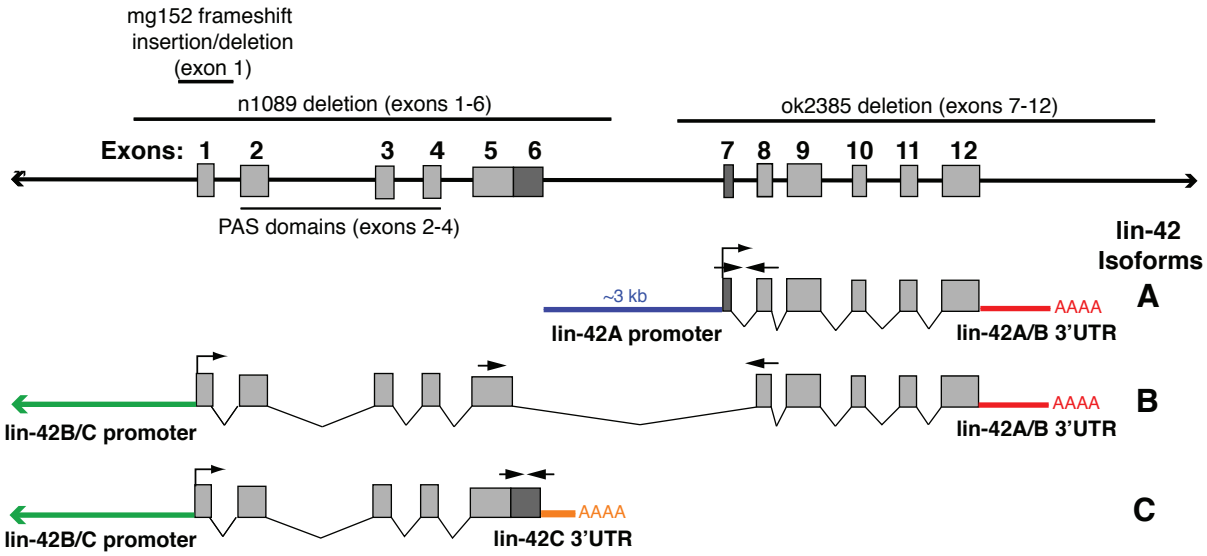


Figure 1.5: *lin-42* isoforms

The *lin-42* genomic region on *C. elegans* chromosome II consists of 12 exons that are alternatively spliced or expressed from different promoters to give rise to three isoforms: *lin-42A*, *lin-42B* and *lin-42C*. Exons 2, 3 and 4 encode the PAS domain (absent in the *A* isoform). The *A* and *B* isoforms share a common 3' UTR (red), while the *B* and *C* isoforms share a common promoter region (green). *lin-42A* has a unique promoter (blue) and *lin-42C* has a unique 3'UTR (orange). Deletions and mutations in the three *lin-42* loss-of-function mutant strains, *mg152, n1089* and *ok2385,* are shown. Self-facing pairs of arrows indicate the positions of primers used for isoform-specific quantitative real-time PCR of *lin-42* transcripts (Chapters 3 and Appendix A).

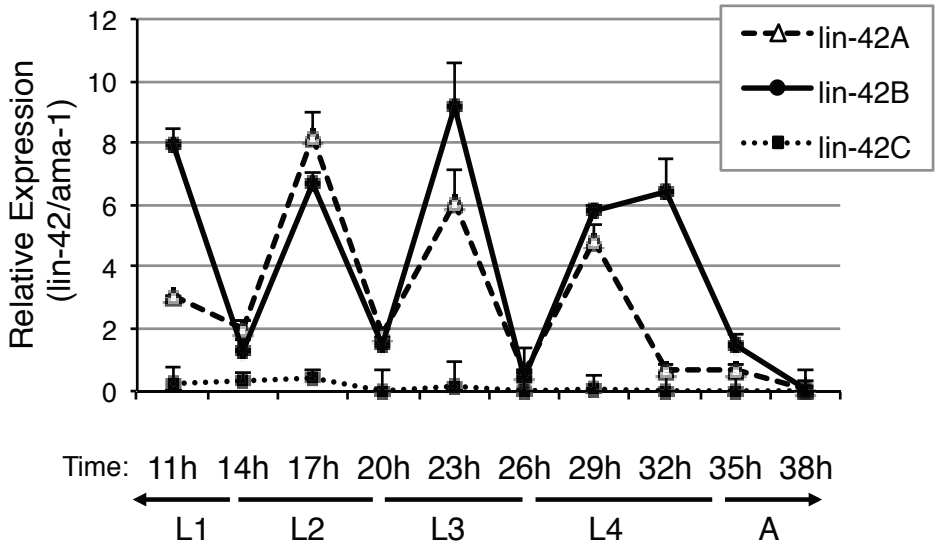


Figure 1.6: Expression profiles of *lin-42A*, *lin-42B* and *lin-42C* during larval stages L2 to L4

Synchronized populations of N2/wildtype *C. elegans* were harvested at 3-hour intervals starting at 11 hrs post addition of starved L1 worms to bacterial food, and resumption of feeding and development, and continuing until 38 hrs. This time frame encompasses late L1 to L4 larval stages and early adult stage, as indicated on the X-axis. The levels of *lin-42* transcripts relative to *ama-1* transcripts (internal reference gene) were measured by quantitative real time PCR (qRT-PCR) using isoform-specific primers and probes (experimental details in Chapters 3 and Appendix A). Error bars represent standard deviation.

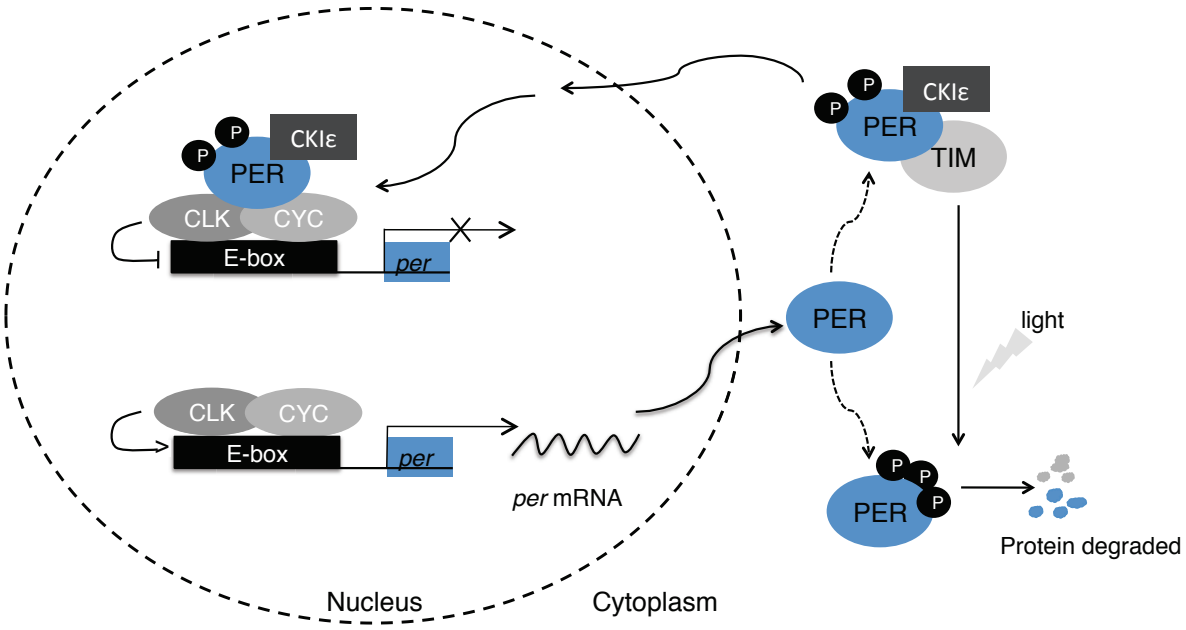


Figure 1.7: The *period* transcriptional/translational auto-regulatory feedback loop of the *Drosophila melanogaster* circadian timing pathway

The binding of CLOCK (CLK) and CYCLE (CYC) transcription factors to E-box elements in *period* promoter drives transcriptional activation of *period* (*per*). *per* mRNA is translocated into the cytoplasm where it is translated into PERIOD protein (PER), which is a phosphorylation target for Casein Kinase I epsilon (CKIε) and other kinases. Depending on the sites and extent of phosphorylation, PER may be either targeted for degradation or stabilized by interaction with Timeless (TIM) and CKIε, and subsequently translocated back into the nucleus. In the nucleus, the PER/ CKIε complex inhibits the CLK/CYC heterodimer and thus prevents transcription of *period*. Timeless is transcribed by the CLK/CYC heterodimer, and thus *timeless* transcription is also inhibited by PER/ CKIε nuclear complex. TIM protein stability in the cytoplasm is regulated via a light-dependent pathway, thus connecting the PER regulatory loop to cell extrinsic environmental signals. (Adapted from [61, 96])

Mutation	Strain	mRNA/Protein levels			Phenotype
		LIN-42A	LIN-42B	LIN-42C	
N/A	N2/wildtype	+/+	+/+	+/+	Wildtype
Frame-shift insertion/ deletion (Exon 1)	*lin-42(mg152)*	+/+	+/0	+/0	pvl, egl, precocious seam cell differentiation, fertile
Deletion (Exons 1-6)	*lin-42(n1089)*	+/+	0/0	0/0	pvl, egl, precocious seam cell differentiation, fertile
Deletion (Exons 7-12)	*lin-42(ok2385)*	0/0	0/0	+/+	pvl, egl, (high penetrance), precocious seam cell differentiation, unc larval lethality
N/A	*lin-42A(RNAi)*	↓/↓	↓/↓	+/+	egl, pvl
N/A	*lin-42B(RNAi)*	+/+	↓/↓	↓/↓	egl, pvl

Table 1.1: *lin-42* loss-of-function mutant strains and *lin-42(RNAi)*

List of known *lin-42* loss-of-function mutant strains and *lin-42(RNAi)* describing *lin-42* isoforms that are not expressed in each worm line and the associated phenotypes. (+) indicates wildtype levels of expression, (0) indicates null expression, (↓) indicates reduced levels of expression. pvl: protruding vulva phenotype, egl: egg-laying defective phenotype, unc: uncoordinated movement phenotype.

.

CHAPTER 2 : Molecular Interaction Screen for *Cis*-regulatory Elements and Transcription Factors that Control *lin-42* Expression

SUMMARY

C. elegans lin-42 controls the timing of cell fate decisions during larval development and is cyclically expressed during each of the four larval stages. Alterations to this expression pattern, leading to loss of expression, over-expression or temporal misexpression, result in abnormal cellular differentiation and lethality. In order to characterize and identify transcriptional regulators of *lin-42*, we used Electrophoretic Mobility Shift Assays (EMSA) to detect direct molecular interactions between putative transcriptional factor proteins and *lin-42 cis*-regulatory DNA sequences. Using EMSA we screened 1.7 kb of the *lin-42B/C* 5' regulatory sequences and identified three DNA regions (Binding Regions, BRs 1-3), varying in size from 65 to 100 base pairs, which were bound with specificity by protein/s present in nuclear extracts from *C. elegans* larvae. We found that the protein/DNA complexes for BR2 and BR3 were not developmental stage specific, i.e. they formed at each of the four larval stages or were present throughout larval development. We attempted to identify the nuclear proteins that bind BR2 using mass spectrometry. The proteins identified were abundant and ubiquitous heat shock factors and cytoskeletal proteins, and none of the identified proteins have a role in transcriptional regulation. These results indicate that future attempts at protein identification by mass spectrometry must include enrichment steps in order to concentrate low abundance transcriptional regulators and thus increase the chances for detection. Despite the lack of success with mass spectrometry, the results of the EMSA screen form the basis for future bioinformatics analysis of the binding regions to identify putative *cis*-regulatory elements and candidate transcription factors that bind

to these elements. These candidate *cis*- and *trans*-regulatory factors can then be tested for transcriptional relevance using a variety of genetic and molecular approaches (as described in Chapter 3).

BACKGROUND AND RATIONALE

The *C. elegans period* homolog *lin-42*, which controls the timing of cell fate specification during larval development, has an oscillatory pattern of expression [13, 44]. *lin-42* expression oscillates with an approximate 8-hour period, with peaks in expression during the intermolts or larval stages, down-regulation of expression during molts, or the transitions between larval stages, and repression of *lin-42* expression on completion of larval development. Reduction or loss of expression and temporal misexpression of *lin-42* results in cell division and differentiation defects in the epidermal seam cells, somatic gonad, sex myoblasts, and vulva tissues, and also alters the timing and period of molting and lethargus [13, 44, 46, 47]. These phenotypes indicate that the cyclic expression pattern of *lin-42* that is coordinated with larval stages is important for its developmental functions. However, the molecular mechanism by which the spatio-temporal expression patterns of *lin-42* are generated and regulated remains unknown. While *lin-42* is the *C. elegans* ortholog of the circadian timing gene *period*, the *C. elegans* homologs of the transcription factors that control *period* expression in the circadian pathway do not control *lin-42* expression during *C. elegans* larval development (**Chapter 3**). To characterize transcriptional regulation of *lin-42*, we sought to identify the *cis*-regulatory elements within *lin-42* promoter and enhancer regions and the *trans*-regulatory factors that bind them to effect the dynamic *lin-42* expression pattern.

Overview of eukaryotic transcriptional regulation

In eukaryotes, transcription begins with the binding of the pre-initiation complex (PIC) to sequences within the core promoter that are positioned approximately 35 to 40 bp around the transcription start site (TSS) [97]. The PIC consists of RNA Polymerase II (RNA Pol II), linked via the Mediator co-activator complex to the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH [98-100]. Recruitment and binding of the PIC to a core promoter is primarily determined by the general transcription complex TFIID, which recognizes and binds a consensus TATA sequence, known as a TATA box, or AT-rich sequences within core promoters [101]. Although recruitment of the PIC to the core promoter is essential for transcription, sustained transcription above a basal level is driven by gene-specific regulatory transcription factors (RTFs) that bind to additional *cis*-regulatory elements outside the core promoter and directly or indirectly interact with the PIC to modulate its activity. (**Figure 2.1**) RTF binding sites are usually clustered within proximal promoters, which are located 250 to 300 bp upstream of the core promoter, and enhancers, whose positions can be highly variable. Eukaryotic enhancer elements may be found in the inter-genic region upstream of the core promoter, or may be distally located dozens or hundreds of kilobases 5' or 3' of the gene, or may be found in the introns of the gene [102]. Together, the proximal promoter and enhancer elements bound by regulatory transcription factors function to elevate transcriptional activity above the basal level and specify spatial and temporal expression patterns. In order to define how the dynamic *lin-42* spatio-temporal expression patterns are generated, we sought (1) to identify the *cis*-regulatory DNA elements, that is the proximal promoters and enhancers, that drive *lin-42* expression, and (2) to identify the *trans*-factors, that is the regulatory transcription factors, that bind to these *cis*-regulatory elements to control *lin-42* expression patterns.

C. elegans transcription regulatory sequences

Owing to the compact size of the *C. elegans* genome, *cis*-regulatory sequences that are bound by regulatory transcription factors are generally found relatively proximal to *C. elegans* gene coding regions [103]. The vast majority of *C. elegans* gene promoters appear to lie within 2 kb upstream of the transcription start site [104]. Most *C. elegans* housekeeping genes fall into this category, such as the *hsp-16* family of genes and vitellogenin (*vit*) genes [105-107]. There are however a few notable exceptions to this general rule. In certain genes like *egl-1*, the *cis*-regulatory elements are located farther than 2 kb downstream of the coding region and within sequences beyond an adjacent, unrelated gene [108]. In another example, reporter gene expression of *lin-39* requires approximately 30 kb of sequence spanning the region both upstream and downstream of the *lin-39* protein-coding region, indicating that some *C. elegans* genes have both complex and distal *cis*-acting control elements [109, 110]. *C. elegans cis*-regulatory elements have also been found within the intronic regions of a gene, particularly when the introns at the beginning of the coding region span several hundred base pairs [111]. For example, the nuclear hormone receptor gene *nhr-23* is regulated by an enhancer present in the first large intron [112]. Despite these exceptions, 2 kb upstream of the ATG site is deemed a reasonable starting point for analysis of most *C. elegans* promoters and *cis*-regulatory elements [102]. Therefore, we focused our analysis on approximately 2 kb of non-coding sequence immediately upstream of the ATG start sites of *lin-42* coding sequences.

Experimental Approach

Our experimental approach was based on the following hypothesis: *regulatory transcription factors that control lin-42 transcription bind with specificity to cognate sequences within a 2 kb non-coding region 5' of the lin-42 ATG start site*. To identify *lin-42 cis*-regulatory sequences and *trans*-factors, we used Electrophoretic Mobility Shift Assays (EMSA), a biochemical approach based on direct protein-DNA interaction between transcription factors and promoter sequences. EMSA is based on the simple principle that a protein-DNA complex migrates through a non-denaturing gel more slowly than unbound DNA. Under optimized binding conditions, transcription factors will bind with high specificity to promoter sequences that contain cognate sites, but will not bind or will bind with low specificity to sequences that do not contain cognate binding sites. When binding reactions between transcription factors in cellular extract and labeled fragments of promoter DNA are resolved by electrophoresis through a non-denaturing, polyacrylamide gel, the unbound labeled DNA separates from labeled DNA bound by protein, and the difference in mobility, the mobility shift, can be detected by chemiluminescence or autoradiography of the DNA label (**Figure 2.2**). The specificity of the binding between protein, the putative transcription factors, and DNA can be assayed by carrying out 'cold competition' with specific and non-specific competitors. For competition with a specific competitor, un-labeled or 'cold' DNA of the same sequence as the labeled probe is added in excess to the binding reaction. The excess of un-labeled DNA out-competes the labeled DNA in binding protein in a concentration dependent manner, and this results in the fading and ultimate disappearance of the upward shifted band representing the complex of protein and labeled DNA. If instead the un-labeled DNA is a non-specific competitor, that is it does not contain cognate binding sites for the proteins in the cellular extract, then addition of excess

unlabeled non-specific competitor will have little or no effect on the formation of the protein-DNA complex, and no change will be observed in the position or intensity of the upward shifted band representing the complex of protein and labeled DNA. EMSA can thus be used to define promoter regions that contain putative *cis*-regulatory sites, and to isolate proteins that bind to these regulatory sites. EMSA-positive DNA fragments that are bound by protein with specificity can be tested for regulatory function by deficiency and sufficiency transcriptional reporter assays, and the specific sequence of the regulatory element can be determined using DNA Foot-printing [113, 114]. The identity of the proteins that bind with specificity to the EMSA-positive DNA fragment can be determined by Mass Spectrometry [115]. The regulatory relevance of these identified proteins can be assayed in a number of ways, such as by assaying the phenotypic effects and effect on *lin-42* transcript levels when the protein function is impaired by mutation or by expression knockdown.

Alternative methods: a comparison of the One Hybrid assay with the EMSA approach

The One-hybrid assay is an alternative method, typically performed in Yeast that is routinely used to identify genes encoding proteins that bind a DNA element of interest. The One-hybrid assay is a two-component system that involves a known DNA sequence, the 'bait', which is bound by an unknown protein, the 'prey'. Interaction between prey and bait results in transcriptional activation of a reporter gene. The DNA bait entails the construction of a reporter plasmid that contains multiple copies of the DNA binding site or promoter element upstream of a TATA box and a reporter gene such as *lacZ* or HIS3. This DNA bait-reporter construct is then stably integrated into the yeast genome. The protein prey comes from a cDNA library that is constructed such that the encoded proteins are expressed as a fusion with a strong transcriptional

activation domain (AD). These protein-AD expression constructs are then co-transformed into the yeast cells carrying the DNA bait-reporter construct. If a prey transcription factor interacts with a bait cis-regulatory element, then the AD of the prey fusion protein activates expression of the reporter gene [116].

In comparison to the EMSA approach, the Yeast One-hybrid assay possesses several advantages. The primary advantages are that binding between transcription factor and *cis*-regulatory element occurs *in vivo*, as compared to the requirement for *in vitro* optimization of binding conditions for EMSA, and that the identity of a protein that activates reporter transcription is immediately known from the cDNA sequence. However, the yeast One-hybrid assay also has a number of disadvantages compared to EMSA that led us to adopt the latter approach. The primary disadvantages of using the One-hybrid method are as follows: (1) A *C. elegans* cDNA library is not commercially available and would have had to be constructed, which would be time-consuming and expensive. A major drawback of yeast One-hybrid assay is the high incidence of false negatives and false positives. (2) False negatives may occur because (i) DNA binding is impaired in the transcription factor-AD fusion protein, or (ii) because accessory factors required for binding are absent, or (iii) because a low abundance transcription factor or a factor that has temporally or spatially restricted expression may not be represented in the cDNA library, or (iv) because a transcription factor requires post-translational modifications that do not occur in yeast cells. (3) False positives may occur because (i) the protein-AD fusion allows for inappropriate and non-specific DNA binding and subsequent reporter activation, or (ii) the protein-AD fusion activates reporter transcription without DNA binding. 4) Activation of the reporter gene in One-hybrid assay is dependent on direct interaction between one transcription factor and the bait DNA element. This excludes the One-hybrid assay from

identifying proteins that may be co-factors that interact with other proteins as part of a multi-protein complex to bind the DNA element of interest, or from identifying transcription factors that require interaction with cofactors for DNA binding.

Despite the obvious drawback of being an *in vitro* assay, for which binding conditions needed optimization, we chose to use EMSA for our study for the following reasons: (1) Nuclear extracts were used for the assay, which facilitated the formation of native protein complexes, since protein modifications would be intact. (2) EMSA allows for further identification of not just proteins that bind directly to the DNA but also of those proteins that may be critical components of a multi-protein complex that drives transcription (e.g. kinases and phosphatases that may alter transcription factor activity by changing its state of phosphorylation). (3) With a biochemical approach like the EMSA, it was possible to modify or fine-tune the binding conditions and nuclear extraction protocols based on the *lin-42* expression profile to answer specific questions or improve the outcome of the results, e.g., nuclear extracts were prepared from synchronized populations of worms that were harvested at time points corresponding to larval stages when *lin-42* expression peaked in order to enrich for relevant transcription factors of *lin-42*. (4) With an EMSA, the likelihood of false positives is higher than false negatives since a positive result is based only on binding of protein to DNA, which could later be tested in functional assays. (5) EMSA is effective and sensitive even when the protein is at a low concentration [117, 118].

Using EMSA we screened approximately 1.7 kb of *lin-42B/C* 5' regulatory region, containing putative promoter and enhancer elements, for sequences that are bound by protein. We identified three regions (binding regions, BR1-3), varying in length from 65 bp to 100 bp,

that are bound with specificity by proteins present in *C. elegans* nuclear extracts, which contain the putative transcriptional regulators responsible for controlling *lin-42* expression.

MATERIALS AND METHODS

C. elegans strains and staging: N2 (wildtype) worms were obtained from the CGC (*C. elegans* Genetics Center), grown at 25°C on NGM (nematode growth medium) plates seeded with *E.coli* strain OP50. Larval populations were synchronized by hatching eggs from an egg prep and starving the L1 larvae in M9 medium in the absence of bacterial food. Larval worms were harvested by washing from the plates with M9 medium, and a further three times with M9 to remove bacteria. The larvae were then pelleted by centrifugation, the supernatant removed and the pellet was flash-frozen in liquid nitrogen. For L2 extracts, synchronized populations of larvae were harvested at 9 and 10 hours post resumption of L1 feeding, for L3 extracts synchronized populations were harvested at 18 and 19 hours post resumption of L1 feeding, and for L4 extracts synchronized populations were harvested at 27 and 28 hours post resumption of L1 feeding.

PAGE and Western Blotting: 20μg of nuclear and cytoplasmic extracts were subjected to SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis. 12% denaturing gels were used to run out the prepared extracts and were analyzed for expression of nuclear (Histone H2A) and cytoplasmic proteins (α-Tubulin) by western blotting. The cellular and nuclear fractions were both probed with primary monoclonal antibody mouse Tubulin α-Ab 2 (1:100) (Neomarkers, Catalog number MS-581-P1) and Histone-H2A polyclonal antibody anti goat C-19 (1:1000) (Santa Cruz Biotechnology, Catalog number sc-8648). Secondary antibodies used were

peroxidase-conjugated goat anti-mouse IgG (1:10000) for α-Tubulin 2 (Jackson Immunoresearch, catalog number 115-035-003) and peroxidase-conjugated donkey anti-goat IgG (1:5000) for Histone H2A (Santa Cruz Biotechnology, catalog number sc-2020).

C. elegans nuclear extracts: Populations of whole worms were flash frozen in liquid Nitrogen and the frozen worm pellets were used to make nuclear extracts using NE-PER Nuclear and Cytoplasmic Extraction Kit (Pierce Thermo Scientific, catalog number 78833). The worm pellets were lysed using a sonicator and kit instructions were followed for extracting the nuclear and cytoplasmic fractions. For pooled extracts, pellets from all three stages were used. For stage-specific extracts, only pellets from the corresponding stages were used. Throughout the extraction, the lysates were kept on ice and 100X Protease and Phosphatase inhibitor cocktail (Pierce Thermo Scientific, catalog number 78441) was added to prevent protein degradation and maintain phosphorylation states. The protein concentrations of the extracts were determined by Bradford assay. If the extracts were not used immediately they were flash frozen in liquid nitrogen and stored at -80°C for up to 6 months. The detailed protocol for preparation of *C. elegans* nuclear extracts is in **Appendix B**.

Electrophoretic mobility shift assays (EMSAs): The 50 to 200 bp DNA fragments used for EMSA probes were PCR amplified from *C. elegans* genomic DNA purified from N2/wildtype worms. The complete list of primers used PCR amplification of DNA probes (for *lin-42*, *myo-2* and *egl-18* genes) is in **Appendix C**. The PCR DNA fragments were non-radioactively labeled with biotin using DNA 3' End Biotinylation Kit (Pierce Thermo Scientific, catalog number 89818). The biotinylated probes were diluted to appropriate concentrations and incubated along

with nuclear extracts in 1X binding buffer (10X: HEPES-262.25mM; Glycerol-30%; KCl-600mM, $MgCl_2$- 40mM, Spermidine- 40mM, DTT-5mM) with the following additives: 100X inhibitor cocktail, 80mM Zn; 80mM Mg, 1μg/μL Poly dI:dC, 1% NP-40 and 1μg/μL BSA (1:100). The binding reaction was carried out in a total volume of 20μL for 1 hour on ice and the reactions were electrophoresed through a non-denaturing 5% TBE gel at 145-150V for 50mins using 0.5X TBE until the tracking dye was three quarters run. Gel-transfer onto a nylon membrane was carried out at 380mA for 30 mins in 0.5X TBE. Once transferred, the labeled DNA was cross-linked onto the membrane using an auto cross-linker at 120mJ/cm^2. The signal was developed using the Chemiluminescent Nucleic Acid Detection Module (Pierce Thermo Scientific, catalog number 89880).

EMSA with specific and non-specific competition: For competition with specific competitor, unlabeled (i.e., non-biotinylated) probe was added to an EMSA binding reaction in ~50-fold excess. For competition with a non-specific competitor, unlabeled 200 bp *myo-2* promoter fragment was added to an EMSA in ~50-fold excess. All other conditions for EMSA were identical to the non-competition assays. For each probe tested, EMSAs with specific/non-specific competition were repeated a minimum of two times if no specific binding was detected and a minimum of three times if specific binding was detected. Fresh nuclear extract was prepared for each set of duplicated EMSAs. The detailed protocol for EMSA is in **Appendix D**.

Mass spectrometry

Sample preparation: To prepare samples for mass spectroscopic analyses, duplicate EMSA reactions were run on separate non-denaturing gels, and only one gel was transferred onto a nylon membrane and developed for signal. The developed film was used as a guide to cut out the

shifted band from the other undeveloped gel. The gel samples were then sent to the Virginia Tech Mass Spectrometry Facility for trypsinization and treatment before analysis by mass spectrometry. All procedures at the Mass Spectrometry Facility were carried out by Dr. Keith Wray.

RESULTS

Preparation of *C. elegans* nuclear extracts

In eukaryotic cells, transcription of nuclear genes occurs exclusively in the nucleus. Thus, as cell-free nuclear extracts serve as an enriched source of regulatory transcription factors, we chose to carry out EMSA with nuclear extracts rather than whole cell extracts. Since there are no published protocols for preparation of *C. elegans* nuclear extracts, we adapted protocols used for human and other vertebrate cells types (**Appendix B**) and also tested a commercially available kit for use with *C. elegans* (Pierce, NE-PER Nuclear and Cytoplasmic Extraction Kit).

For both the lab-developed protocol and the commercial kit, *C. elegans* larval populations were mechanically lysed by either sonication or using a Dounce homogenizer. More efficient release of cells from the worm body and cell lysis were achieved using a sonicator compared to the Dounce homogenizer, as determined by observation under the dissecting microscope and DAPI staining of nuclei to determine the integrity of the nucleus. With the Dounce homogenizer, we observed that many worms remained partially intact, with very modest release of nuclei from cells that had been released from tissues. In contrast, sonication effectively dissociated worms into cells and we observed that the majority of cells had lysed but that nuclei remained intact. The presence of proteases and nucleases can severely affect the activity of the extracts and to prevent protein degradation, protease and phosphatase inhibitors were added to the lysates at

every step. In addition, the temperature of the cellular extracts was maintained at 4°C or below throughout the protocol to prevent protein degradation. The relative efficiencies of the lab-developed protocol and the commercial kit in separating nuclear and cytoplasmic factions of cellular extracts were compared by Western blot analysis. We used antibody against Histone 2A, a highly conserved nuclear specific protein that is a component of nucleosomes, and antibody against alpha-Tubulin, a highly conserved component of the microtubule cytoskeleton that is localized in the cytoplasm, to assay the integrity of the nuclear factions (**Figure 2.3**). For both the lab-developed protocol and the commercial kit, we found enrichment of Histone 2A staining in the nuclear faction compared to the cytoplasmic faction. Although there was greater enrichment of Histone 2A staining with the lab-developed protocol, this enrichment was not consistent. Furthermore, there was consistently more Tubulin staining in the nuclear factions from the lab-developed protocol compared to the nuclear factions from the commercial kit. These results indicate that the lab-developed protocol does not separate the cytoplasmic and nuclear factions as well, and as consistently, as the commercial kit. We thus used the commercial kit for preparation of *C. elegans* nuclear extracts in all subsequent assays.

To further increase our chances of identifying relevant transcriptional factors by the EMSA approach, we used nuclear extracts from *C. elegans* larvae at specific developmental time points. *lin-42B* is dynamically expressed with peaks of expression during each larval stage (**Figure 2.5**). *lin-42B* transcriptional activators must be active during induction of the *lin-42B* message, but may not be active when *lin-42B* levels drop. Thus, instead of using nuclear extracts from worms from a mixed population at different developmental stages, we synchronized worm populations and harvested larvae at time points corresponding to *lin-42B* induction during L2, L3 and L4 stages (**Figure 2.5**). For initial EMSA screening, we pooled nuclear extracts from

synchronized L2, L3, and L4 populations. However, to assay whether binding of protein to a particular probe is developmental stage-specific, we used L2, L3 and L4 stage nuclear extracts separately (**Figure 2.12**).

Optimization of protocol for EMSA

There are only a few published protocols for EMSA with *C. elegans* reagents, and these protocols vary widely in the buffers and binding conditions used. We thus developed our own EMSA protocol, which was based on these published protocols, and tested it using validated transcription factor binding sequences as EMSA probes with *C. elegans* nuclear extracts. The transcription factors DAF-3 binds the sequence GTCTG at position 3049636 (that is, at -416 bp) in the *myo-2* promoter, and the transcription factor LIN-39 binds the sequence TGATTAT(G/T)(G/A) in the *egl-18* promoter [119, 120]. *myo-2* is expressed in the head and pharyngeal cells, and *egl-18* is expressed in the head and hypodermal seam cells, during multiple stages of larval development [119, 120]. Since *myo-2* and *egl-18* are expressed in tissues and at developmental time points that overlap with *lin-42* expression patterns, [13, 44] we reasoned that EMSA conditions that allow for specific and stable binding of nuclear protein to the *myo-2* and *egl-18* promoter probes would also allow us to detect transcription factor binding at *lin-42* promoter probes. We had difficulty amplifying an *egl-18* promoter fragment containing the LIN-39 binding site, and thus we used a 200 bp *myo-2* promoter fragment, containing the DAF-3 binding site, as an EMSA probe for the majority of our assays. In EMSA using nuclear extract from larval stage *C. elegans* worms, we detected two shifted bands (**Figure 2.4A**, lanes 1 and 6, white box) that were competed away by increasing amounts of unlabeled *myo-2* probe, that is by specific competitor DNA (**Figure 2.4A**, lanes 2-5, 7-10, white box). These two bands

presumably represent the specific DAF-3/*myo-2* complexes. However, a second set of shifted bands that formed at the top of the gel was unaffected by increasing amounts of unlabeled *myo-2* probe, indicating that these bands represent non-specific binding between protein and the probe DNA. A fifty-fold excess, the lowest amount of unlabeled specific competitor, was sufficient to compete away the specific protein/*myo-2* DNA complex formed, and thus all specific and non-specific competitor DNA probes in subsequent competition EMSAs were used at 50-fold excess. The results of these experiments validated our nuclear extract and EMSA protocols and methodology.

To optimize the EMSA, we created a list of common reagents used in published EMSA protocols and their roles in promoting optimal binding between transcription factors and promoter sequences [118, 121-123]. (**Table 2.1**) One of the critical parameters in EMSA is the concentrations of mono- and divalent salts, specifically potassium (K^+) and magnesium (Mg^{2+}), which have to be as close to the microenvironment of the nucleus as possible, although the exact ionic requirement may vary depending on the specific transcription factor [124, 125]. We tested the effect of varying potassium and magnesium concentrations on formation of the putative DAF-3/myo-2 bands in EMSA (**Figure 2.4B**). We observed that the DAF-3/*myo-2* shifted bands formed at all salt concentrations tested, but the sharpest bands formed at 100 mM or 150 mM potassium and 5 mM magnesium. The potassium concentration was further increased to 200 mM with magnesium concentration kept constant at 5 mM (**Figure 2.4A**), and these salt conditions were used for all binding assays thereafter.

EMSA screen of 1.7 kb region of *lin-42B/C* 5' regulatory sequences

For the purpose of identifying *cis*-regulatory elements in the *lin-42B/C* promoter, we assayed 1.7 kb of non-coding sequence upstream of the ATG site. Our rationale for focusing on this region was three-fold: First, the majority of C. *elegans* gene promoters lie within 2 kb of the ATG translation start site [102, 104]. Second, phylogenetic footprinting, i.e. comparative sequence analysis between *C. elegans* and closely related nematode species, shows that the highest levels of non-coding sequence conservation are within this 2 kb region (**Figure 2.6**). Third, a transcriptional reporter construct with 2 kb of *lin-42B/C* promoter recapitulates LIN-42B spatio-temporal expression patterns (Banerjee Lab, unpublished). Thus, we predicted that the main regulatory promoter/enhancer elements that control *lin-42* expression lie within the proximal 2 kb regulatory region of *lin-42B/C* open reading frame.

To assay *lin-42B/C* promoter by EMSA, we used sixteen 200 bp probes, with 100 bp overlaps, to cover 1.7 kb of the proximal regulatory DNA (**Figure 2.6**). Two hundred base pairs is the largest length of probe that has been successfully used in published EMSAs. The overlaps among probes allow us to narrow down binding sites to 100 bp regions. For example in Figure 2.5, specific binding by protein of probes 2 and 3 but not of probes 1 and 4 would indicate that the protein binding sites are in the 100 bp region common to the distal end of probe 2 and proximal end of probe 3. The probe overlaps also ensure that binding elements that fall between two adjoining probes are represented intact in the middle of the overlapping probe. For example in Figure 2.5, a protein binding site that is at the junction of probes 1 and 3 may not be intact in either probe but is present intact in the middle of probe 2 and can be bound by protein.

Using EMSA with nuclear extracts from *C. elegans* larvae, we identified three regions within 1.7 kb of the *lin-42B/C* 5' regulatory sequences that are bound by protein with specificity.

These regions are designated binding regions 1 to 3 (BR 1-3). BR1 falls within the first 300 bp of regulatory sequence upstream of the transcriptional ATG start site of *lin-42B/C* that was assayed with EMSA probes 1 to 4 (**Figure 2.7A**). Probes 2 and 3 gave EMSA shifted bands that were competed away with specific competitor but not with non-specific competitor, while probes 1 and 4 did not form any specific protein/DNA complexes (**Figure 2.7B**). These results define a 100 bp region common to probes 2 and 3 that we designated BR1.

BR2 is located at approximately 500 bp of regulatory sequence upstream of the transcriptional ATG start site of *lin-42B/C* that assayed with EMSA probes 4 to 6 (**Figure 2.8A**). Of these 3 probes, only probe 5 formed a specific protein/DNA complexes suggesting that the relevant binding site lies at the junction of probes 4 and 6 (**Figure 2.8B**). To test this possibility, we created a smaller 65 bp long probe 4-6, that spanned the sequence at the junction of probes 4 and 6 and used it in EMSA. We found that probe 4-6 formed a high molecular weight protein/DNA complex that was competed away by specific competitor and not by non-specific competitor (**Figure 2.8B**). We designated this 65 bp region as BR2.

The region from 500 bp to 1100 bp, assayed with probes 6 to 11, did not form specific protein/DNA complexes in EMSA (**Figure 2.9**). However, we identified another binding region, BR3, upstream of 1100 bp assayed with probes 12 to 16 (**Figure 2.10A**). EMSA of probes 12 and 13 gave shifted bands that were competed away with specific competitor but not with non-specific competitor, while probes 11, 14, 15 and 16 did not form any specific protein/DNA complexes (**Figure 2.10B**). A smaller 100 bp probe, probe 12-13, consisting of the overlap sequence between probes 12 and 13, also gave a specific EMSA shift. These results define a 100 bp region between 1200 and 1300 bp, common to probes 12 and 13, that we designated BR3.

The three binding regions contain putative *cis*-regulatory elements that are bound with specificity by *trans*-regulatory factors, such as transcription factors. Transcription factor binding sites are usually smaller than 20 bp long. We thus attempted to narrow down the 100 bp BR1 to smaller *cis*-regulatory elements. To this end, we designed primers to amplify twelve 50 bp EMSA probes (probes 1.1-1.12) spanning probes 1 to 3 and covering 350 bp of promoter sequence including BR1 (**Figure 2.11**). Three of the twelve probes (1.1, 1.6 and 1.7) gave us specific shifts, defining two 25 bp regions from 0 to -25 and -150 to -175 of the *lin-42B/C* proximal 5' promoter (**Figure 2.11**). However, EMSA shifts with these 50 bp probes were not consistently reproducible, and the results are not consistent with the EMSA shifts obtained with the 200 bp probes. In contrast, the shifted bands using the 200 bp probes for BR1, the 65 bp probe for BR2 and the 100 bp probe for BR3 formed consistently in multiple EMSAs and behaved in the same manner in multiple competition assays (minimum of three EMSAs for each probe with different preparations of nuclear extracts). We thus did not attempt to use similar 50 bp probes to further assay BR3.

Stage-specific formation of protein/DNA EMSA complexes

C. elegans goes through four larval developmental stages L1, L2, L3 and L4, and *lin-42* expression oscillates in phase with these four stages. We thus asked whether the protein or protein complexes that bind to the binding regions do so in a stage-specific manner. We used nuclear extracts that were obtained from developmentally synchronized populations at L2, L3 and L4 stages (**Figure 2.5**) for EMSAs with probe 4-6, which defines BR2, and probe 12-13, which defines BR3. For both probes, the same pattern of EMSA shifted bands was formed with each stage-specific extract as with the pooled stage extract (**Figure 2.12**). These results indicate

that the protein/DNA complexes formed at BR2 and BR3 do not form stage-specifically. These proteins may thus remain bound to the DNA throughout larval development or these complexes may form reiteratively during each larval stage.

Identification of proteins that bind BRs by Mass Spectrometry

We attempted to use mass spectrometry to identify the proteins that bind with specificity to the *lin-42B/C* promoter binding regions. For mass spectrometry, we cut the band representing the specifically bound protein/DNA complex on probe 5 from the non-denaturing gel and sent this sample to the Mass Spectrometry Facility. As a control, we also excised an equivalent gel slice from an area of the gel without any band formation. The proteins associated with the gel samples were analyzed by LC/MS-MS (liquid chromatography and tandem mass spectrometry), and the proteins identities determined from the *C. elegans* protein database. LC/MS-MS, also known as Tandem Mass Spectrometry is routinely used in proteomics for the identification of complex protein mixtures. A sample mixture, once separated by simple methods like SDS-PAGE, following protease digestion (usually trypsinization) can be analyzed for component peptide sequences by Tandem MS. Even in complex samples where peptide masses may overlap, the high-resolution mass spectrometer used in LC/MS-MS can identify individual peptides [126]. Sample preparation and all LC/MS-MS protocols were carried out by Dr. Keith Wray at the Virginia Tech Mass Spectrometry Facility. We made two attempts at protein identification, using the same EMSA probe and experimental conditions. We were provided with a list of the proteins identified by LC/MS-MS in the two attempts (**Table 2.2**). Unfortunately, the majority of the proteins identified represent ubiquitous and high abundance proteins such as heat shock factors, cytoskeletal components and other housekeeping proteins. Moreover, the analysis did not un-

cover any proteins that were enriched in the shifted band slice and not present in the control gel slice with no bands. Thus, none of the proteins identified represent meaningful candidates for further analysis as transcriptional regulators.

DISCUSSION

Several published methods are available for isolation of nuclear proteins from *C. elegans* embryos, oocytes and adults, but none for larval stage worms. We thus developed a protocol for isolation of nuclear extracts from *C. elegans* larvae using published protocols as well as a commercially available kit designed primarily for extraction of nuclear protein from HeLa cells. We also developed an EMSA protocol for *C. elegans* and optimized the binding conditions using a known promoter element in the *myo-2* promoter that is bound by transcription factor (DAF-3). Using these optimized conditions we were able to resolve 1700 bp of *lin-42B/C* 5' regulatory sequence into three separate regions of binding activity (BR1, BR2 and BR3) that contain putative *cis*-regulatory sites.

In EMSA, as with any *in vitro* assay, the question arises as to whether the results obtained represent functionally relevant DNA-protein interactions that reflect *in vivo* reality. While validation of candidate factors identified in the screen depends on additional assays, it is important to reiterate that our methodology selects for specific and biologically relevant interactions between protein and DNA. Cold competition assays were carried out with both specific and non-specific competitors to ensure that proteins bound to DNA probes with high specificity. We also used poly dI:dC carrier DNA in all EMSAs to further reduce non-specific binding. Furthermore, we used EMSA probe DNA concentrations in the femtomolar range (10^{-15}) to encourage specific and biologically relevant interactions. Affinity and kinetics studies of

transcription factor-target DNA binding have shown that EMSA probe DNA concentrations at no higher than the picomolar range (10^{-12}) are required for specific binding between probe DNA and transcription factors in typical nuclear extracts [127]. Theoretically, the lower the concentration required for binding or complex formation, the stronger the interaction and the lower the K_D (equilibrium binding constant). Using very low probe concentrations in the femtomolar range (approximately 3.5 femtomols/binding reaction), we obtained EMSA shifts that we were able to compete away with specific competitor and reform on addition of non-specific competitor. These EMSA protein-DNA complexes are thus likely to be biologically relevant. Finally, we have further confidence in our results because we only detected three regions of specific binding (BR1-3), which all fall within 5' non-coding regions that are the most highly conserved among the related nematode species (**Figures 2.7, 2.8, 2.10**). Encouragingly, we found no specific binding of DNA by nuclear protein in 5' non-coding regions that lacked any sequence conservation (**Figure 2.7** probe 4, **Figure 2.9** probe 9).

The three binding regions defined by the EMSA screen lie at approximately -200 to -300 (BR1), -500 (BR2), and -1200 to -1300 (BR3) in the *lin-42B/C* 5' regulatory region. BR1 may represent the proximal regulatory promoter, while the more distal binding regions, BR2 and BR3, may be part of enhancers to which tissue-specific or temporal regulatory elements may bind. EMSA with larval stage-specific nuclear extracts showed that the same protein complexes from L2-specific, L3-specific and L4-specific extracts bound to BR2 and BR3 probes. These results could indicate that transcription factors that bind to BR2 or BR3 are cyclically expressed and reiteratively activate *lin-42* transcription at each larval stage, thus generating the oscillatory *lin-42* expression patterns. Alternatively, proteins may remain bound to a binding region

throughout larval development, with transcriptional activity being modulated by periodic expression and/or activity of activators or repressors.

The proteins that bind to the BRs may not be transcription factors. For example, DNA binding proteins include those involved in chromatin remodeling and nucleosome positioning. These proteins also play a role in gene regulation by altering access to the gene and its regulatory sites, without actual physical interaction with the core transcription complex or polymerase II [128, 129].

The EMSA screen results are a springboard for future discovery and can be used in further assays to identify both the *cis*-regulatory elements and the *trans*-regulatory factors within the binding regions that control *lin-42* expression. Bioinformatic analysis of BR1, 2, and 3 can be used to identify putative *cis*-regulatory sites and candidate *C. elegans* transcription factors that bind these sites (**Chapter 3**). These candidate *cis*- and *trans*-factors can then be tested for transcriptional function and relevance using reverse genetic methods, such as RNA interference, phenotypic assays and measurement of transcription, such as quantitative PCR and reporter assays. Individual regulatory elements can be defined using techniques such as DNA footprinting. Although our initial mass spectrometric analysis did not yield any meaningful protein identities, the process of sample preparation for mass spectrometry can be optimized. For example, we did not take any steps to enrich for transcription factors, which are typically found in cells in low abundance. A number of methods have recently been developed to enrich for transcription factors prior to mass spectrometry [123, 130-132]. These methods include Oligonucleotide or Promoter Trapping, which entails attaching linkers or protein moieties to the promoter fragments (i.e. 'tailing') to enable linking the DNA onto magnetic beads or a sepharose column. The DNA-column is incubated with nuclear extracts to allow formation of the

transcription factor-DNA complexes. The columns are subsequently washed under low stringency conditions so as to remove non-specifically bound proteins and then under high stringency conditions to elute the transcription factor-enriched fraction for analysis by Mass spectrometry. With an enrichment step for low abundance transcription factors, protein identification by mass spectrometry retains the possibility of identifying multiple components of a transcriptional complex.

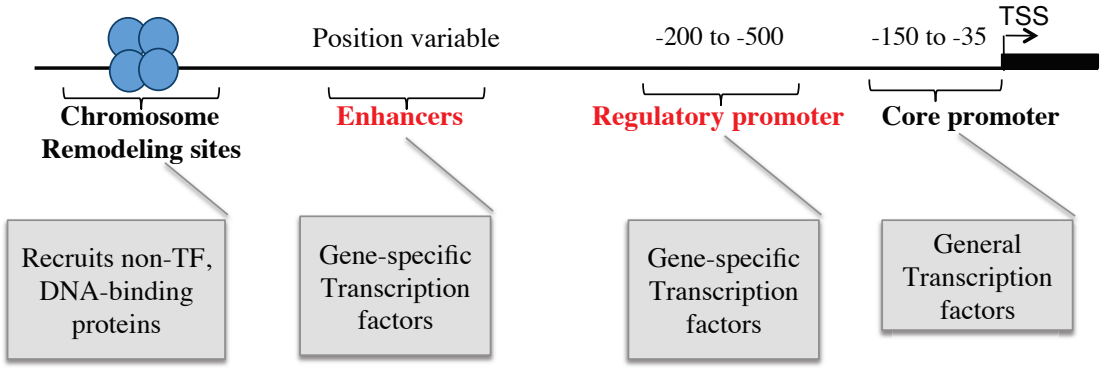


Figure 2.1: Overview of eukaryotic regulatory DNA

The four types of non-coding regulatory sequences that control transcription and their relative positions from the translational start site (TSS) are indicated. General transcription factors bind to the core promoter, gene-specific transcription factors bind to the regulatory promoter and proximal or distal enhancer sites and non-transcription factor DNA binding proteins bind to the chromosome remodeling sites. Blue circles represent nucleosomes formed by the Histone octamers. (Adapted from [129, 133])

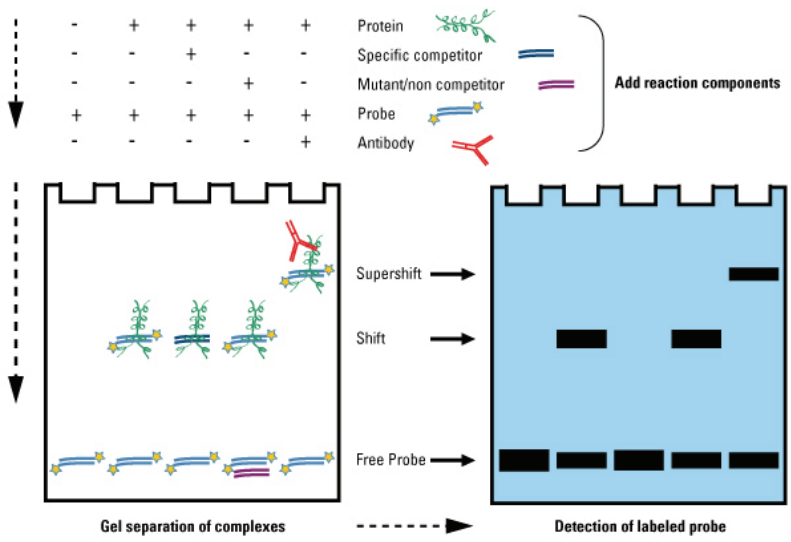


Figure 2.2: Schematic of Electrophoretic Mobility Shift Assay (EMSA).

Figure adapted from Fisher Thermo Scientific/Pierce Protein Methods Library (http://www.piercenet.com/EMSA)

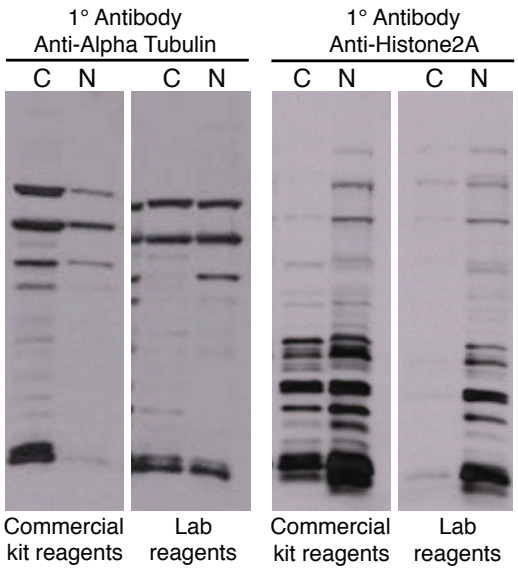


Figure 2.3: Western analysis of α-Tubulin and Histone H2A in cytoplasmic and nuclear fractions of *C. elegans* cellular extracts.

Cytoplasmic and nuclear extracts from larval stage *C. elegans* cells were prepared using a commercially available kit (Pierce NE-PER Kit, panels 1 and 3) and lab-made reagents (panels 2 and 4). Cytoplasmic (C) and nuclear (N) fractions were probed using polyclonal antibodies against α-Tubulin (panels 1 and 2) or against Histone-2A (panels 3 and 4).

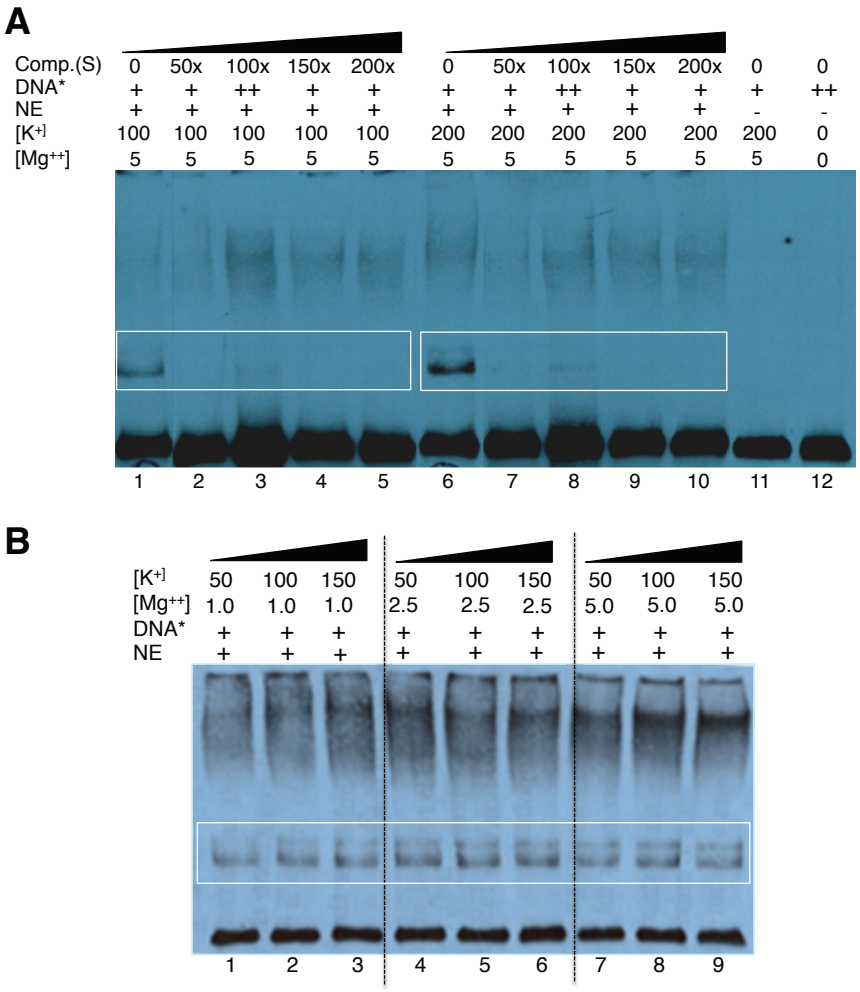


Figure 2.4: Optimization of cold competition EMSA conditions and effect of varying monovalent/divalent salt concentrations using control EMSA probe *myo-2*.

(A and B) Approximately 6 μg of total protein in nuclear extract (NE) was incubated with 3.5 fmols (+) or 7.0 fmols (++) of biotinylated *myo-2* probe DNA (DNA*) in varying concentrations (50-200 mM) of potassium salts ([K+]) and varying concentrations (1.0-5.0 mM) of magnesium salts ([Mg++]).

(A) Optimization of cold competition assay conditions using increasing excess of specific competitor DNA. Un-biotinylated *myo-2* probe was used as specific competitor (Comp.(S)) and 50-fold (50x) to 200-fold (200x) excess of 3.5 fmols was added as indicated. Two sets of shifted bands formed under the two sets of EMSA reaction conditions (lanes 1-5 and lanes 6-10), but only the lower band (in white frame) was competed away with increasing amounts of specific competitor DNA and reappeared on addition of 2-fold excess *myo-2* probe (lanes 3 and 8). Lanes 11 and 12: free probe control reactions.

(B) Effect of varying mono- and divalent salt concentrations on *myo-2* EMSAs. Three sets of varying potassium and magnesium salt concentrations were tested (lanes 1-3, lanes 4-6 and lanes 7-9). Specific shifted bands (lower bands in white frame) and non-specific bands (upper bands, unframed) formed under all conditions tested.

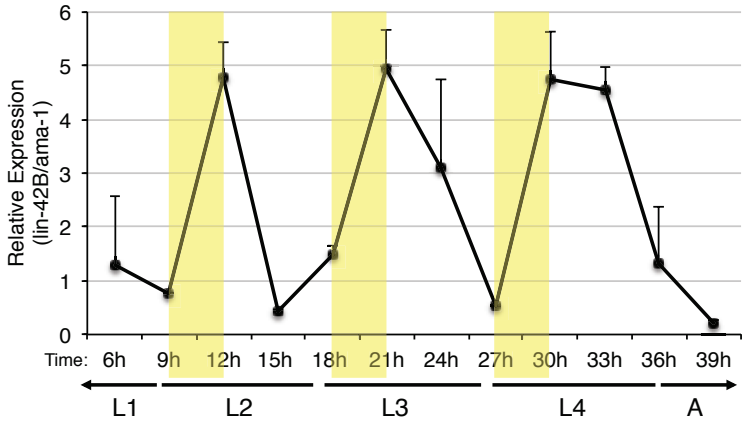


Figure 2.5: Cyclic expression of *lin-42B* over the course of larval development in *C. elegans*.

Synchronized populations of N2/wildtype *C. elegans*, fed on OP50 bacteria and grown at 25°C, were harvested at 3 hour intervals starting at 6 hrs post addition of starved L1 worms to bacterial food, and resumption of feeding and development, and continuing until 39 hrs. This time frame encompasses late L1 to L4 larval stages and early adult stage, as indicated on the X-axis. The levels of *lin-42B* transcript relative to *ama-1* transcripts (internal reference gene) were measured by quantitative real time PCR (experimental details in Chapter 3 and 4). Error bars represent standard deviation. The time frames highlighted in yellow represent times at which worms were harvested for preparation of stage-specific nuclear extracts, coinciding with the peaks in *lin-42B* mRNA expression. Data from Appendix A.

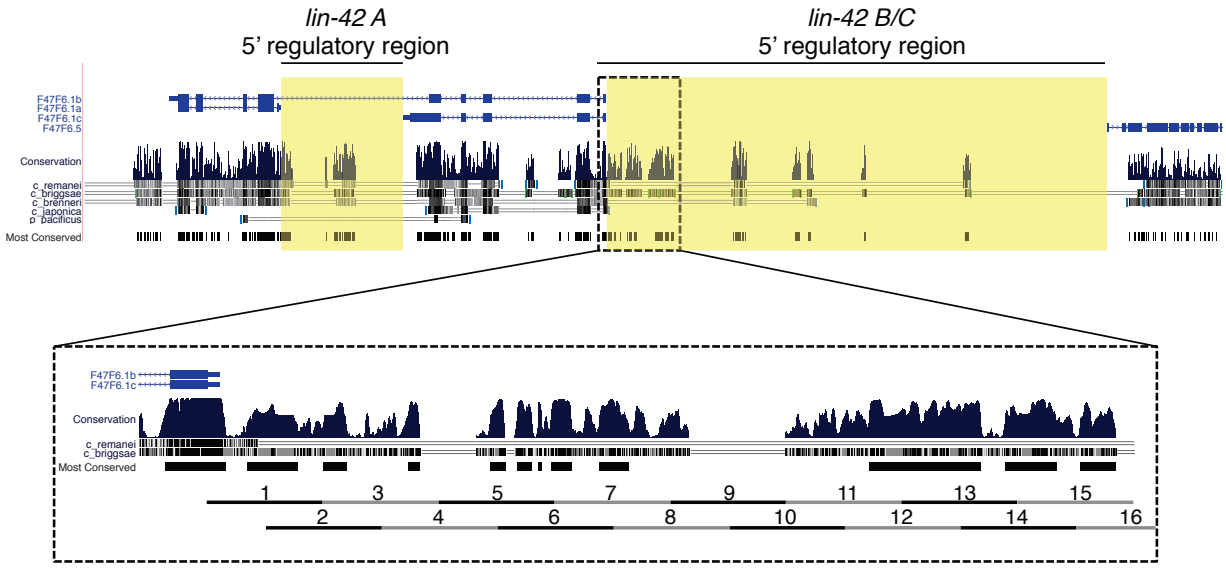


Figure 2.6: UCSC Genome Browser view of *C. elegans lin-42* genomic region.

Screen capture from UCSC Genome Browser (http://genome.ucsc.edu/) showing exons (blue boxes) and non-coding sequences (blue lines) for *C. elegans lin-42A, B* and *C* isoform (*F47F6.1a, b, c*) open reading frames on chromosome II. The 5' regulatory regions for *lin-42B/C* (inter-genic region between *lin-42B/C* and gene *F47F6.5*) and *lin-42A* (inter-genic region between *lin-42A* and *lin-42C* exons) are highlighted in yellow. The approximately 1.7 kb of *lin-42B/C* 5' non-coding DNA assayed by EMSA is framed in dotted lines and shown enlarged. Black and grey lines numbered 1 to 16 in the enlarged dotted-line frame show relative positions of 200 bp EMSA probes used to assay *lin-42B/C* regulatory sequences. The UCSC Genome Brower view of *lin-42* genomic region also shows the level of sequence conservation (blue peaks and black and grey bars) between *C. elegans* and other closely related (*C. briggsae*, *C. remanei*, *C. brenneri*) and distantly related (*C. japonica*, *P. pacificus*) nematode species. The regions of greatest conservation among the five species are indicated by the highest blue peaks and the black bars at the bottom of the screen capture (Most Conserved).

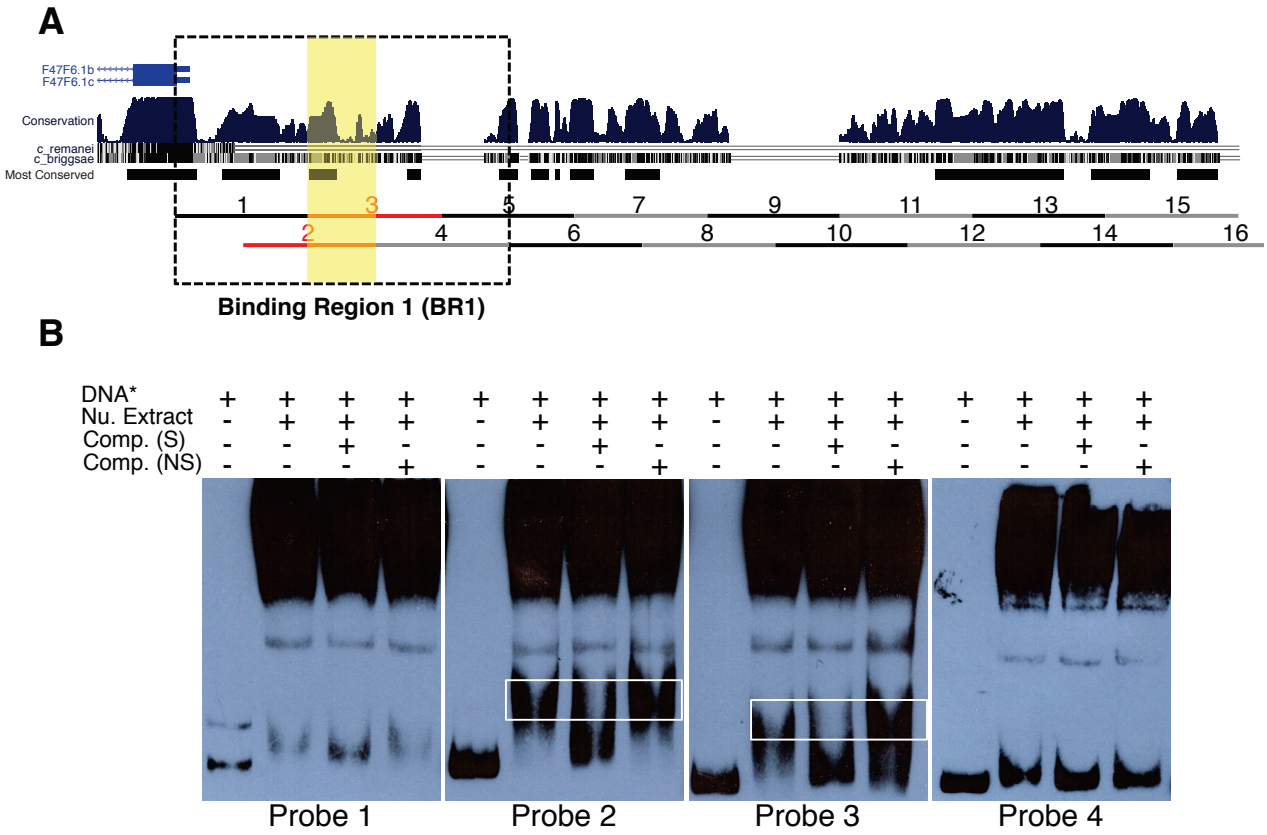


Figure 2.7: Identification of Binding Region 1 (BR1) by EMSA screen.

(A) UCSC Genome Browser view of *C. elegans lin-42B/C* 5' non-coding sequences described in Figure 2.5. The 500 bp of non-coding sequence screened by probes 1 to 4 is boxed with dotted lines. Probes highlighted in red produced specific shifts in EMSA. The 100 bp binding region 1 (BR1) defined by EMSA-positive probes 2 and 3 is highlighted in yellow.
(B) EMSA blots of 200 bp DNA probes 1 to 4. The white frames show shifted DNA probes that are bound by protein with specificity, as determined by competition assays (lanes 3 and 4 of each blot). DNA*: biotin labeled DNA probe; Nu. Extract: nuclear extract from *C. elegans* larval cells; Comp. (S): specific competitor DNA, unlabeled probe DNA added in 50-fold excess of the biotinylated probe; Comp. (NS): non-specific competitor DNA, unlabeled 200 bp *myo-2* probe added in 50-fold excess of the biotinylated probe. (+) indicates addition and (-) indicates omission.

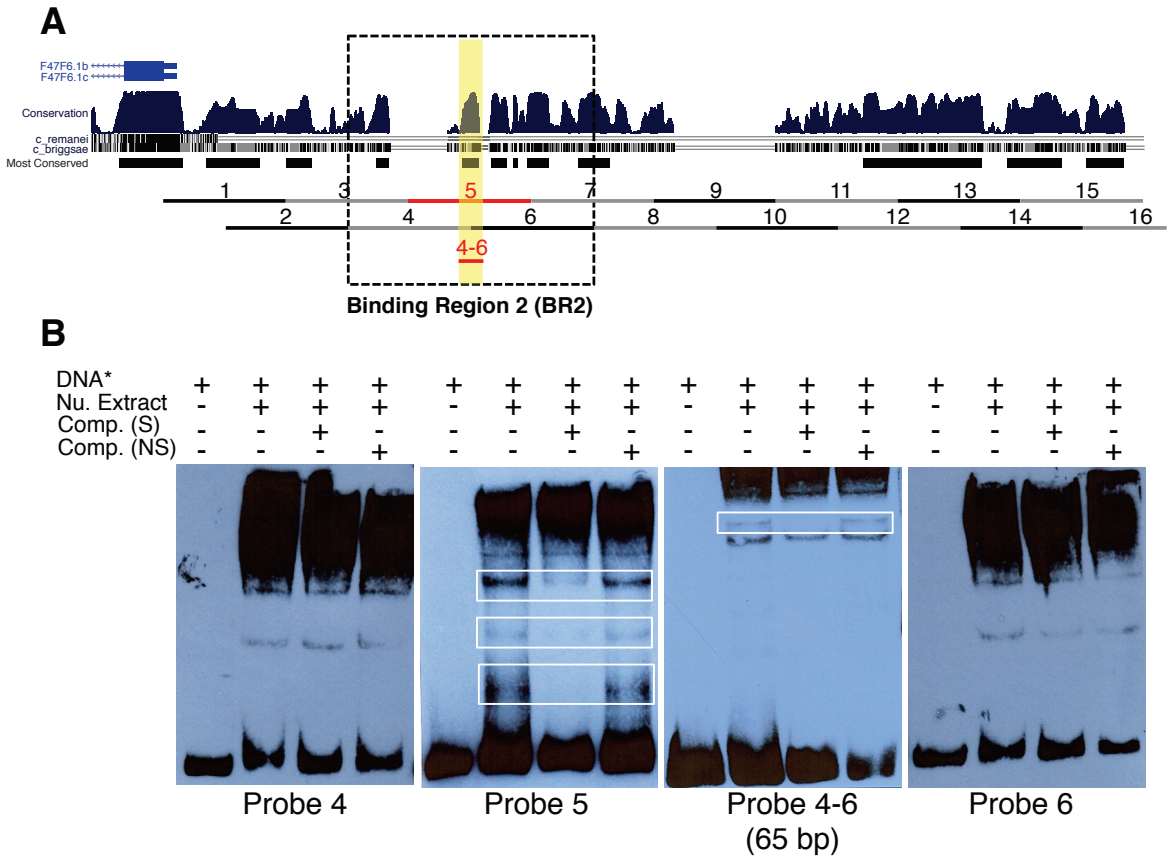


Figure 2.8: Identification of Binding Region 2 (BR2) by EMSA screen.

(A) UCSC Genome Browser view of *C. elegans lin-42B/C* 5' non-coding sequences described in Figure 2.5. The 400 bp of non-coding sequence screened by probes 4 to 6 is boxed with dotted lines. Probes highlighted in red produced specific shifts in EMSA. The 65 bp binding region 2 (BR2) defined by EMSA-positive probes 5 and 4-6 is highlighted in yellow.
(B) EMSA blots of 200 bp DNA probes 4, 5, 6 and 65 bp DNA probe 4-6. The white frames show shifted DNA probes that are bound by protein with specificity, as determined by competition assays (lanes 3 and 4 of each blot). DNA*: biotin labeled DNA probe; Nu. Extract: nuclear extract from *C. elegans* larval cells; Comp. (S): specific competitor DNA, unlabeled probe DNA added in 50-fold excess of the biotinylated probe; Comp. (NS): non-specific competitor DNA, unlabeled 200 bp *myo-2* probe added in 50-fold excess of the biotinylated probe. (+) indicates addition and (-) indicates omission.

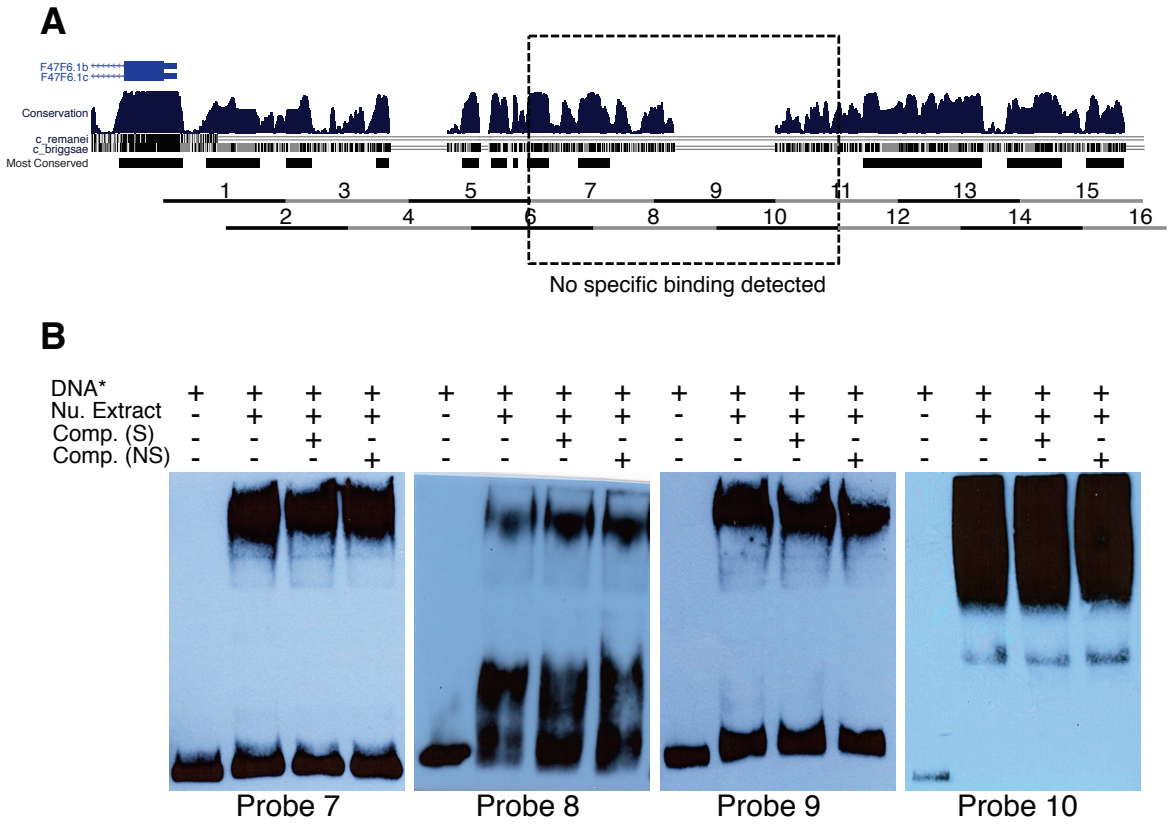


Figure 2.9: EMSA screen with probes 7 to 10 shows no specific binding of nuclear protein.

(A) UCSC Genome Browser view of *C. elegans lin-42B/C* 5' non-coding sequences described in Figure 2.5. The 500 bp of non-coding sequence screened by probes 7 to 10 is boxed with dotted lines. No specific binding of DNA probes by nuclear proteins was detected in this region.
(B) EMSA blots of 200 bp DNA probes 7, 8, 9 and 10. DNA*: biotin labeled DNA probe; Nu. Extract: nuclear extract from *C. elegans* larval cells; Comp. (S): specific competitor DNA, unlabeled probe DNA added in 50-fold excess of the biotinylated probe; Comp. (NS): non-specific competitor DNA, unlabeled 200 bp *myo-2* probe added in 50-fold excess of the biotinylated probe. (+) indicates addition and (-) indicates omission.

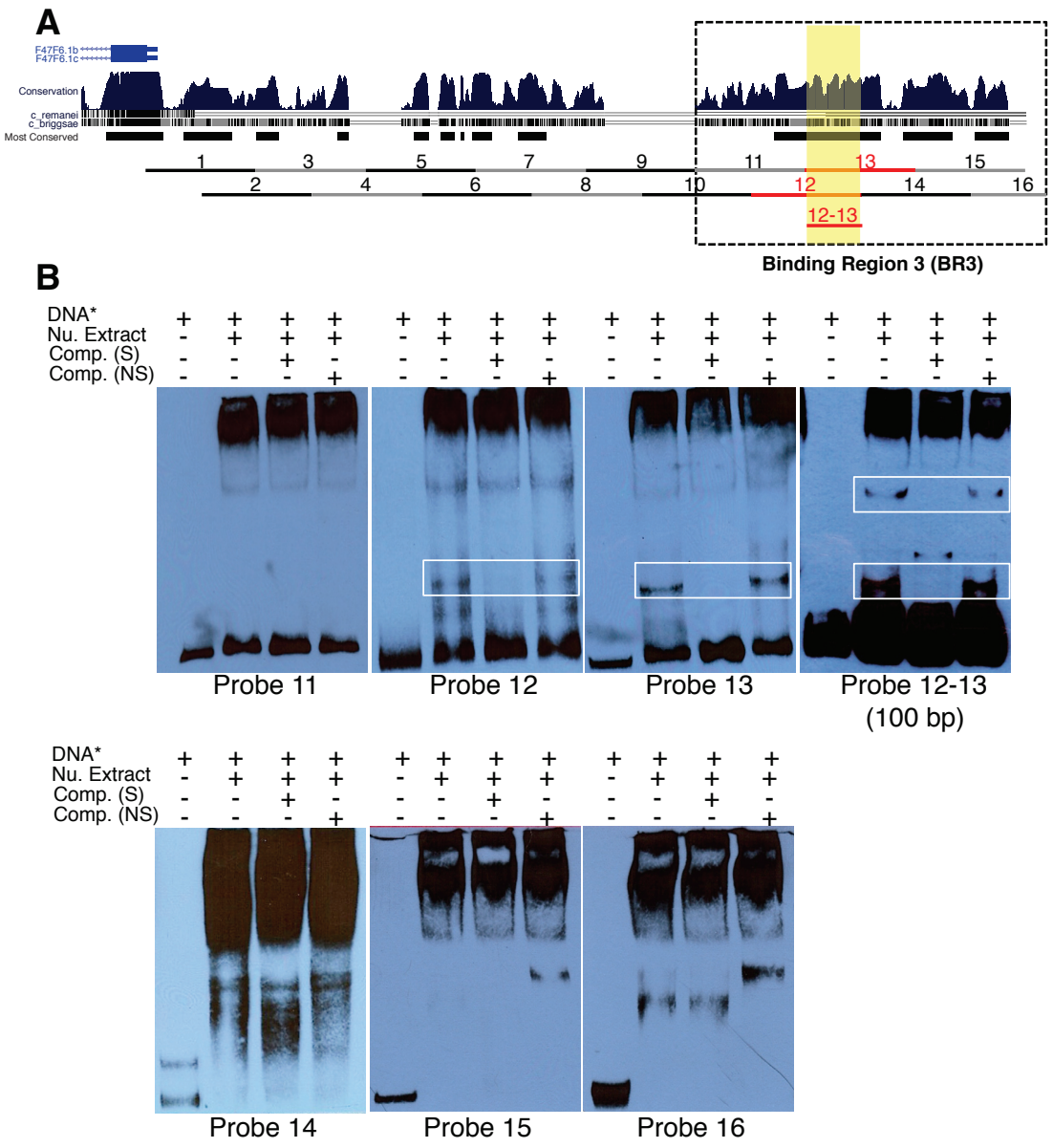


Figure 2.10: Identification of Binding Region 3 (BR3) by EMSA screen.

(A) UCSC Genome Browser view of *C. elegans lin-42B/C* 5' non-coding sequences described in Figure 2.5. The 600 bp of non-coding sequence screened by probes 11 to 16 is boxed with dotted lines. Probes highlighted in red produced specific shifts in EMSA. The 100 bp binding region 3 (BR3) defined by EMSA-positive probes 12, 13 and 12-13 is highlighted in yellow.
(B) EMSA blots of 200 bp DNA probes 11, 12, 13, 14, 15, 16 and 100 bp DNA probe 12-13. The white frames show shifted DNA probes that are bound by protein with specificity, as determined by competition assays (lanes 3 and 4 of each blot). DNA*: biotin labeled DNA probe; Nu. Extract: nuclear extract from *C. elegans* larval cells; Comp. (S): specific competitor DNA, unlabeled probe DNA added in 50-fold excess of the biotinylated probe; Comp. (NS): non-specific competitor DNA, unlabeled 200 bp *myo-2* probe added in 50-fold excess of the biotinylated probe. (+) indicates addition and (-) indicates omission.

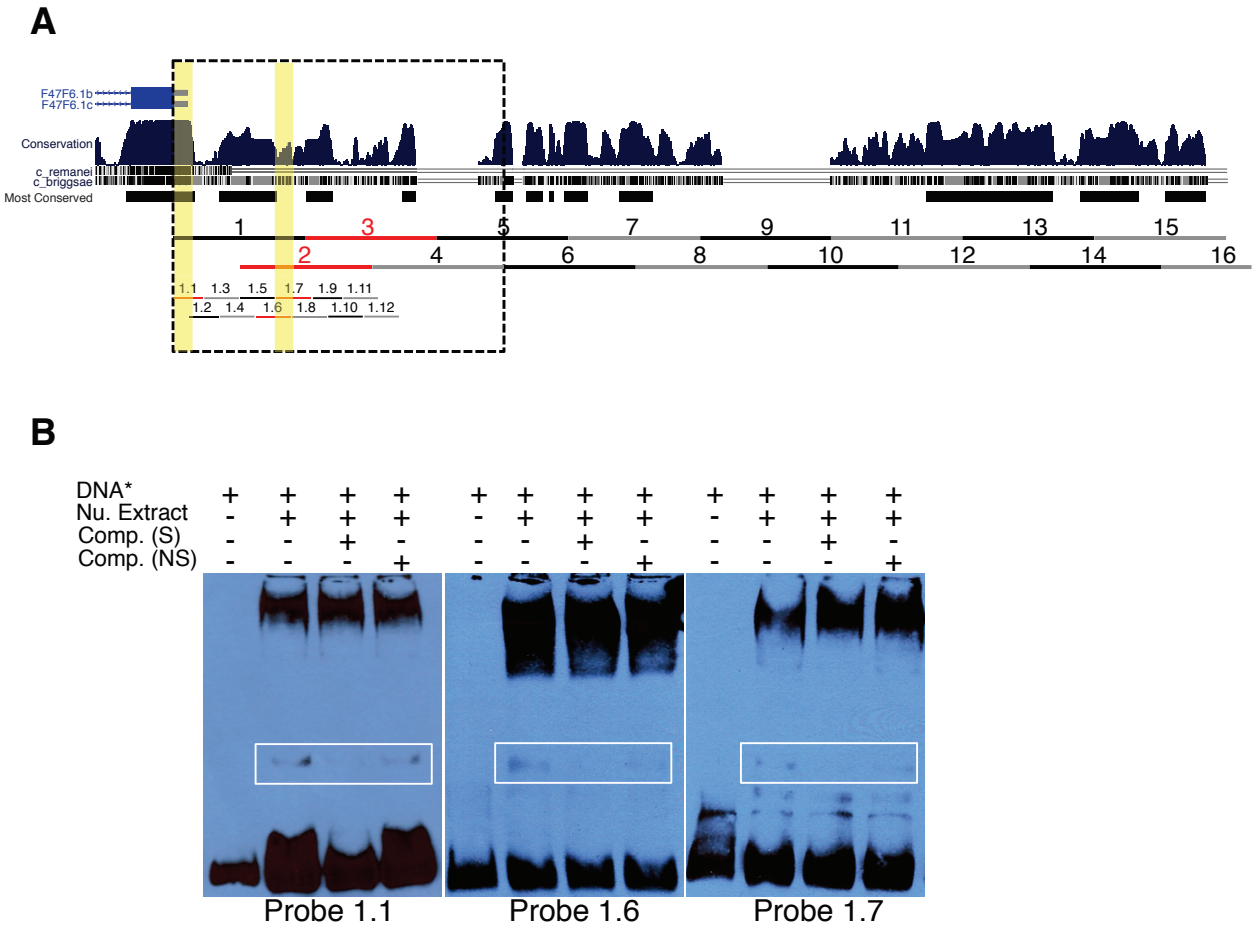


Figure 2.11: Re-screening of BR1 with 50 bp EMSA probes.

(A) UCSC Genome Browser view of *C. elegans lin-42B/C* 5' non-coding sequences described in Figure 2.5. The 500 bp of non-coding sequence screened by probes 1 to 4 is boxed with dotted lines. Part of this region was re-screened with 50 bp probes 1.1 to 1.12. Probes highlighted in red produced specific shifts in EMSA. The 25 bp binding regions defined by EMSA-positive probes 1.1, 1.6 and 1.7 are highlighted in yellow.
(B) EMSA blots of 50 bp DNA probes 1.1, 1.6 and 1.7. The white frames show shifted DNA probes that are bound by protein with specificity, as determined by competition assays (lanes 3 and 4 of each blot). DNA*: biotin labeled DNA probe; Nu. Extract: nuclear extract from *C. elegans* larval cells; Comp. (S): specific competitor DNA, unlabeled probe DNA added in 50-fold excess of the biotinylated probe; Comp. (NS): non-specific competitor DNA, unlabeled 200 bp *myo-2* probe added in 50-fold excess of the biotinylated probe. (+) indicates addition and (-) indicates omission.

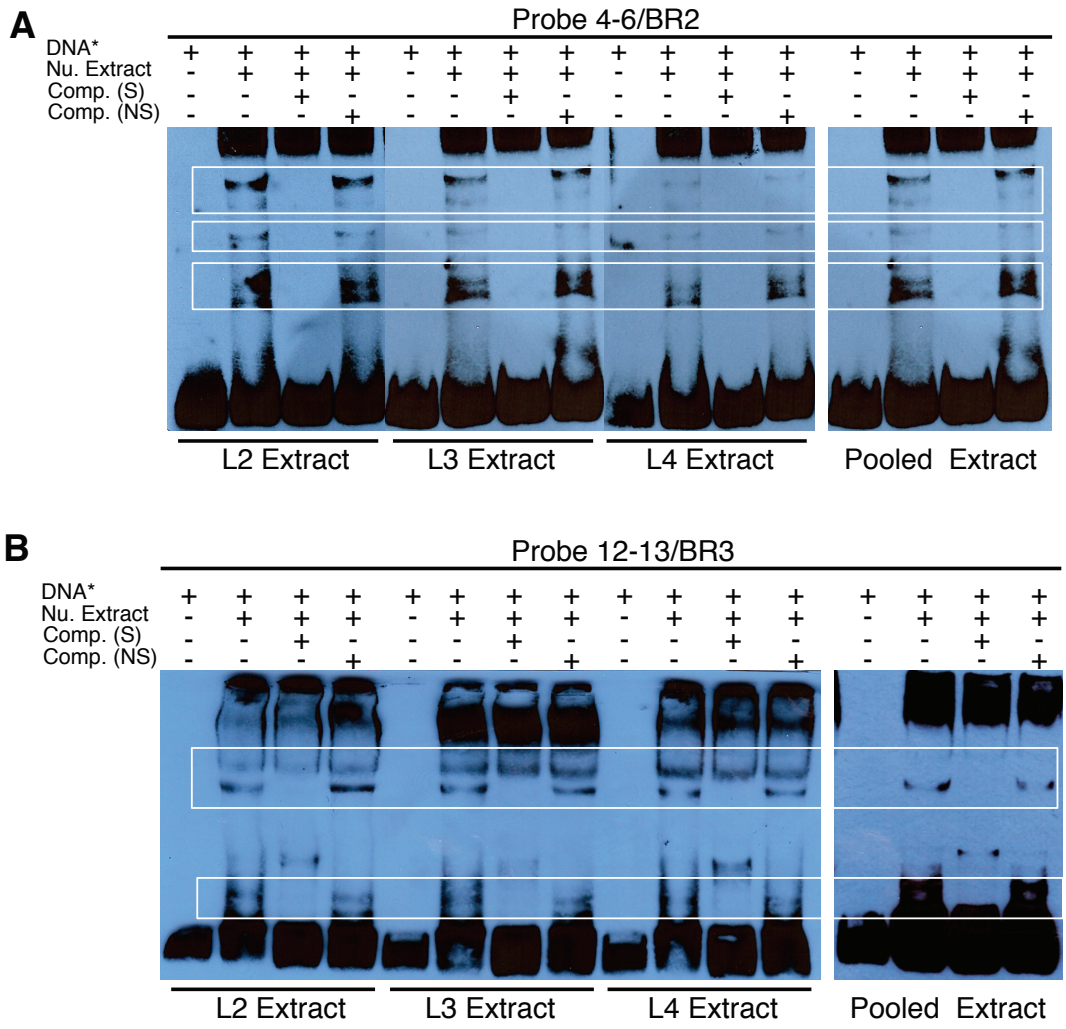


Figure 2.12: Stage-specific EMSAs of BR2 and BR3.

EMSA blots of the 65 bp probe 4-6 (A) and 100 bp probe 12-13 (B) using stage-specific nuclear extracts (L2, L3 and L4) and pooled extracts (L2+L3+L4). The white frames show shifted DNA probes that are bound by protein with specificity, as determined by competition assays. DNA*: biotin labeled DNA probe; Nu. Extract: nuclear extract from *C. elegans* larval cells; Comp. (S): specific competitor DNA, unlabeled probe DNA added in 50-fold excess of the biotinylated probe; Comp. (NS): non-specific competitor DNA, unlabeled 200 bp *myo-2* probe added in 50-fold excess of the biotinylated probe. (+) indicates addition and (-) indicates omission.

Component	Role	Function
HEPES / Tris-HCl	Buffer	Maintain pH at ~7.2
Glycerol	Neutral solute	Stabilizes labile proteins
Spermidine	Poly-cation	Eukaryotic transcription co-factor
BSA	Carrier protein	Minimizes non-specific loss of binding proteins
NP-40/Tergitol	Non-ionic detergent	Maximizes protein solubility
DTT	Reducing agent	Reducing conditions favor binding
PolydI.dC	Non-specific competitor	Reduces non-specific binding
K^+	Mono-valent cation	Required for optimal TF binding
Mg^{++}	Di-valent cation	Required for optimal TF binding

Table 2.1: Common components and variant components of EMSA binding buffer.

Source	Trial	*C. elegans* Protein	Protein identity
EMSA band	Trial 1, Trial 2	ACT-1	Actin
EMSA band	Trial 1	COL-178	Collagen
EMSA band	Trial 1	HSP-1	Heat Shock Protein
EMSA band	Trial 1, Trial 2	HSP-70	Heat Shock Protein
EMSA band	Trial 2	PCCA-1	Propionyl CoA Caboxylase, alpha subunit
EMSA band	Trial 2	RPL-32	Ribosomal Protein
EMSA band	Trial 2	RPL-35	Ribosomal Protein
EMSA band	Trial 2	RPL-7	Ribosomal Protein
EMSA band	Trial 2	RPS-16	Ribosomal Protein
EMSA band	Trial 1	RSP-0	Ribosomal Protein
EMSA band	Trial 1	TMP-1	Tropomyosin, muscle isoform I
Control	Trial 2	COL-178	Collagen
Control	Trial 1	EFT-3	Elongation Factor
Control	Trial 1	HSP-70	Heat Shock Protein
Control	Trial 2	RPS-10	Ribosomal Protein

Table 2.2: *C. elegans* proteins identified by LC/MS-MS analysis of protein-DNA complex formed by EMSA of 200 bp probe 5 encompassing BR2

CHAPTER 3 : Identification and Characterization of the *lin-42* Transcriptional Regulator *ztf-16* by RNAi Screen

SUMMARY

During *C. elegans* larval development, stage-specific and reiterative expression and function of *lin-42* isoforms control the timing of cell fate programs in a number of tissues, including the seam cells, which are epidermal stem cells. However, the transcriptional regulators that control the dynamic temporal expression patterns of *lin-42* are unknown. We utilized a reverse genetic approach, based on screening candidate genes by RNA interference (RNAi), in order to identify transcription factors that control *lin-42* expression during seam cell larval development. The RNAi screen is based on characterized genetic interactions between *lin-42* and *let-7* microRNAs. Loss of function (lf) of *let-7* microRNAs results in retarded development and late larval lethality, which are suppressed by *lin-42* loss-of-function. We thus knocked-down expression of candidate transcription factors and assayed for suppression or enhancement of the lethality phenotype in *let-7 (lf)* mutants in order to identify potential transcriptional activators or repressors of *lin-42*. To identify candidate transcription factor genes for the RNAi screen, we conducted bioinformatics analysis of putative transcription factor binding sites (BR1-3) that were identified in our previous EMSA screen (Chapter 2). We screened 43 candidate transcription factors genes and identified *ztf-16* and *nhr-23* as being potential transcriptional activators of *lin-42*. We confirmed that these two genes regulate *lin-42* transcription by using quantitative real time PCR to measure *lin-42* transcript levels on RNAi-knock down of *ztf-16* or *nhr-23*, and in *ztf-16(ok3028)lf* mutant worms. We went on to further characterize the effect of loss of function

of the zinc-finger transcription factor ZTF-16 on seam cell development, specifically seam cell division and differentiation, and found that *ztf-16(lf)* phenocopies *lin-42(lf),* consistent with its proposed function as a *lin-42* transcriptional regulator.

BACKGROUND AND RATIONALE

Period is a core regulatory protein in the circadian timing pathway that regulates daily patterns of physiology and behavior, and is widely conserved across animal phyla. Period also functions in regulating cell fate, specifically adult stem cell fate, and mutations in mammalian period genes result in oncogenesis [56, 75]. While the circadian function of Period is well characterized at the molecular and cellular levels, the developmental roles of Period are poorly understood. We are studying *C. elegans lin-42*, the nematode ortholog of *period*, in order to define the molecular mechanisms of *period* regulation and function in adult stem cell pathways.

As described in Chapter 1, correct spatio-temporal expression of *lin-42* is important for regulating cell fate choices throughout larval development in *C. elegans*. *lin-42* acts broadly in the worm to regulate temporal developmental events in the gonad, vulva, and sex myoblasts [44]. Additionally, *lin-42* controls terminal differentiation of the epidermal seam cells at the end of the larval stage that marks the transition to the adult. In *lin-42(lf)* mutants, this terminal differentiation event occurs precociously in the L3 or early L4 stage instead of at the end of larval development [14, 44]. *lin-42* regulation of seam cell division patterns during earlier larval stages is revealed in *lin-42(lf);lin-14(lf)* double mutants that exhibit omission of the L2-stage specific proliferative division [12, 18]. Furthermore, *lin-42* levels need to be repressed at L2 in order for the worm to enter into the developmentally quiescent dauer stage, thereby sensitizing *lin-42(lf)* mutants to dauer formation [46]. However, *lin-42* function is not restricted to

developmental alone. It has also been shown to have a physiological function in circadian pathways wherein *lin-42* mutations result in aberrant circadian rhythms of locomotion during the adult stage [65, 68]. The clear temporal separation of *lin-42*/*period* developmental timing function during larval stages versus circadian timing function during adult stage allows us to study *lin-42*/*period* developmental roles that are distinct from physiological roles mediated through the circadian pathway.

Unlike most other heterochronic genes that function during larval development, *lin-42* expression is not restricted to a specific stage but oscillates through the four larval stages with periodic peaks and troughs [13]. (**Figure 3.5A**) *lin-42* is expressed as three alternatively spliced isoforms that are regulated by different promoters and 3' UTR sequences. These isoforms have been shown to have different spatio-temporal expression profiles during larval versus adult stages suggesting stage-specific and pathway-specific transcriptional regulation. (Chapter 1) Besides the differential expression profiles, the isoforms are functionally non-redundant. *lin-42A* is primarily required to regulate the reiterative pattern of larval lethargus and ecdysis, *lin-42B* is primarily required to regulate larval stage-specific seam cell division and differentiation and *lin-42C* is the only isoform that is substantially expressed during adult stage, suggesting it is the isoform that regulates circadian rhythms of behavior in *C. elegans* [68], (Banerjee lab unpublished). (**Chapter 1, Table 1.1**) The different *lin-42* mRNA expression patterns and the phenotypic differences among *lin-42* mutants together suggest that there are differences in transcriptional and post-transcriptional regulation among the isoforms, and that *lin-42* transcriptional control may involve both stage-specific as well as multi-stage regulators.

Overview of Experimental Approach

We designed a reverse genetic screen, based on RNA interference (RNAi) of candidate genes, in order to identify transcriptional regulators of *lin-42*. We used the fact that *lin-42* loss or reduction of function suppresses the lethality phenotype of *let-7* loss of function as the basis of the RNAi screen. Thus, reduction of function by RNAi of a *lin-42* transcriptional activator should similarly suppress the *let-7(lf)* lethality phenotype, while RNAi of a *lin-42* transcriptional repressor should instead enhance the lethality phenotype. (**Figures 3.1B, 3.2A**) Candidate transcription factor genes to be tested in the RNAi screen were identified by bioinformatics analysis of the *lin-42B/C* promoter regions (BR1, BR2 and BR3) identified as potential transcription factor binding sites in the previous EMSA screen. (Chapter 2) The regulation of *lin-42* by a candidate transcription factor identified by the RNAi screen was then tested by assaying *lin-42* transcript levels on RNAi knockdown of the candidate factor.

let-7 family microRNAs and their mutant phenotypes

let-7 is a 21-nucleotide microRNA that has heterochronic developmental function during *C. elegans* larval stages [16]. *let-7* expression during early larval stages is inhibited by HBL-1, until de-repression takes place at the end of L2 stage [39]. (**Figures 1.4, 3.1A**) A low level of *let-7* expression is detected at the early L3 stage, with maximal expression occurring during L4 stage (**Figure 3.1A**) [16]. Loss-of-function of *let-7* results in retarded heterochronic phenotypes at the end of the L4 stage and larval worms fail to transition into the adult stage. *let-7(lf)* mutant worms exhibit reiteration of the L4 larval molt and the hypodermal seam cells fail to terminally differentiate [16]. Over-expression of *let-7* on the other hand causes the hypodermal seam cells to precociously exit the cell cycle at the end of the L3 stage. *let-7(n2853)* is a *let-7(lf)* strain that

carries a point mutation that causes a temperature-sensitive, adult lethal vulval bursting phenotype [16]. Like other microRNAs, *let-7* binds to complementary sequences in the 3' UTR of its target gene transcripts and down-regulates the expression of these gene products by inhibiting translation [40, 134]. *let-7* has multiple target genes, including the heterochronic genes *lin-41, lin-42* and *daf-12*. (**Figure 3.1A)** These *let-7* target genes all have *let-7* binding sites in their 3' UTRs, and loss of function of these genes suppresses lethality due to *let-7(lf)*, indicating that lethality is caused by inappropriate continued expression or over-expression of the *let-7* target genes in the absence *of let-7* regulatory function [16, 17, 37, 135].

In C. elegans there are three *let-7* paralogs, *mir-48, mir-84* and *mir-241*, which are referred to as the *let-7* family microRNAs *(let-7-fam miRs)* [136, 137]. The *let-7-fam* microRNAs are expressed as early as the L1 stage, with expression peaking in the L2 stage [14]. (**Figure 3.1A**) The *let-7-fam* microRNAs function redundantly. Single mutant strains of *mir-48*, *mir-84* or *mir-241* are wildtype and the double mutant strain (*mir-48; mir-84)* display a very slight retarded phenotype [14]. However, double mutants *mir-48; mir-241* and triple mutants *mir-48 mir-241; mir-84* display stronger retarded phenotypes in the seam cell lineage causing them to repeat the L2 proliferative division at the L3 stage, with incomplete alae formation at the L4 molt. The strain VT1066 is a triple deletion of all three *let-7-fam* microRNAs and exhibits about 60 to 70% lethality at the L4 molt [14]. The target of the *let-7-fam* microRNAs is the heterochronic gene *hbl-1*, whose expression is repressed at the end of the L2 stage to allow de-repression of *let-7* and allow progression to the later larval stages [14, 138]. (**Figure 3.1A**) Similar to the relationship between *let-7* and its target genes, *hbl-1* loss or reduction of function suppresses the retarded development and lethality phenotypes of VT1066 (*let-7-fam lf* strain) [14].

Hence, the *let-7* family of microRNAs function stage-specifically during early and late larval development. The *let-7-fam* microRNAs (*mir-48, mir-84, mir-241*) function at late L2 during the transition from early to late larval development, and *let-7* functions at late L4 during the transition from larval to adult stages. (**Figure 3.1A**) *lin-42* functions during both early and late larval development in conjunction with *let-7* target genes *hbl-1* and *lin-41* to specify the L2 stage and L4 stage cell fates, respectively. The loss-of-function phenotypes of both *let-7* and *let-7-fam* microRNAs are suppressed by *lin-42(lf),* thus placing *lin-42* downstream of *let-7* microRNAs [13, 17]. (**Figure 3.1A**)

Logic of the RNAi screen

To identify transcriptional regulators of *lin-42*, we designed an RNAi screen based on the genetic interactions between *lin-42* and the *let-7* microRNAs. Loss-of-function of *lin-42* suppresses *let-7* loss-of-function, indicating that *lin-42* levels are elevated in *let-7(lf)* mutants. We reasoned that RNAi knockdown of a *lin-42* transcriptional activator would also suppress lethality in a *let-7(lf)* mutant as *lin-42* expression would be depressed in these RNAi animals. (**Figure 3.1B**) Conversely, RNAi knockdown of a *lin-42* transcriptional repressor would enhance lethality in a *let-7(lf)* mutant as *lin-42* expression would be further elevated in these RNAi animals. (**Figure 3.1B)** There are three possible outcomes of RNAi knockdown of a candidate transcription factor, each one indicative of not only whether the candidate gene encodes a putative transcriptional regulator, but also of the nature of regulation, i.e. whether the factor is a transcriptional activator or repressor. The possible outcomes are: i) No significant change in *let-7(lf)* lethality, indicating that the tested gene does not encode a *lin-42* transcriptional regulator, ii) A significant enhancement of *let-7(lf)* lethality, indicating that the tested gene may encode a

repressor of *lin-42* expression or function, iii) A significant suppression of *let-7(lf)* lethality, indicating that the tested gene may encode an activator of *lin-42* expression or function.

The RNAi screen was carried out in two different *let-7(lf)* strains - VT1066, triple deletion mutant for the *let-7-fam* microRNAs and the temperature-sensitive *let-7(n2853)* strain. (**Figure 3.2A**) At 20°C, both VT1066 and *let-7(n2853)* strains show approximately 70% penetrance of lethality at the L4 to adult molt due to herniation through the vulva (vulval bursting phenotype) [14, 16]. Using the two different *let-7(lf)* strains allowed us to enable identification of the stages at which the possible *lin-42* regulators function, i.e. whether the factors function reiteratively during multiple stages or whether the factor is stage-specific. (**Figure 3.2B**) For example, if a candidate gene is a suppressor of VT1066 lethality and not a suppressor of *let-7(n2853)* lethality, then this result indicates that the encoded factor is probably a L2 stage-specific transcription factor. If instead, the candidate gene is a suppressor of only *let-7(n2853)* lethality, then the encoded factor may be a L4-specific transcription factor. If a candidate gene is a suppressor of both VT1066 lethality and *let-7(n2853)* lethality, then the encoded factor may act reiteratively during each of the larval stages to regulate the oscillating expression of *lin-42*. (**Figure 3.2B**)

The *let-7* microRNAs have multiple targets, and thus a candidate suppressor or enhancer resulting from the RNAi screen need not necessarily be a regulator of *lin-42*. Therefore, as the final step of the RNAi screen, we used isoform-specific qRT-PCR to directly measure *lin-42* transcript levels following RNAi knockdown of a candidate transcription factor gene in wildtype genetic background. (**Figure 3.2A**) We predicted that RNAi of lethality suppressors (i.e. putative transcriptional activators) would result in loss or lower levels of *lin-42* expression, while RNAi

of lethality enhancers (i.e. putative transcriptional repressors) would result in elevation of *lin-42* expression.

Bioinformatic analysis to identify candidate transcription factor genes for RNAi screen

The completely sequenced *C. elegans* genome is predicted to contain approximately 19,735 protein-coding genes and 934 of those are predicted to encode transcription factors [139, 140]. We used Phylogenetic Footprinting to identify candidate transcription factors that may bind to the *lin-42B/C* promoter. Phylogenetic Footprinting identifies sequences that have been conserved over evolutionary time among related species [141]. Coding sequences are conserved if the amino acids they encode are part of a protein with essential or important function. Only sequence variations that do not significantly affect protein structure and function are tolerated. However, there are fewer selective constraints on non-coding sequences, which tend to be much more divergent than coding sequences even among very closely related species. Thus, non-coding sequence conservation often indicates that the sequence has structural or regulatory function, such as an enhancer element that is recognized and bound by a specific type of transcription factor.

The UCSC Genome Browser program shows the level of sequence conservation in 1.7 kb of *lin-42B/C* 5' non-coding sequences between *C. elegans* and five other nematode species - *C. remanei, C. briggsae, C. brenneri, C. japonica* and *P. pacificus*. (**Figures 2.5, 3.3A**) *C. briggsae* and *C. remanei* are the most closely related of the *Caenorhabditis* species, with *C. elegans* and *C. brenneri* as sister species to the *C. brigssae* and *C. remanei* clade. *C. japonica is a Caenorhabditis* species that is the most divergent from and distantly related to *C. elegans*. *P. pacificus* is a satellite model nematode also commonly used in *C. elegans* comparative biology,

but is the most distantly related to *C. elegans*, as is evident from the low conservation of both coding and non-coding sequence between *C. elegans* and *P. pacificus* [142]. (**Figures 2.5, 3.3A**)

There is a high degree of sequence conservation for *lin-42* exons among *C. elegans*, *C. briggsae* and *C. remanei*. We thus used these three species for Phylogenetic Footprinting of BR1, BR2 and BR3 sequences of *lin-42B/C* 5' non-coding region. We previously defined these BRs as potential transcription factor binding sites using EMSA. (Chapter 2) We looked for DNA motifs within BR1, BR2 and BR3 that were conserved in *C. elegans, C. briggsae* and *C. remanei*, identified mammalian and *Drosophila* transcription factors that bind similar sequences in their native genomes and finally identified *C. elegans* genes that encode homologs of these transcription factors. These *C. elegans* transcription factor genes were tested as potential *lin-42* regulators in the *let-7(lf)* lethality suppressor/enhancer RNAi screen.

MATERIALS AND METHODS

C. elegans strains, staging and harvests: *N2* (wildtype), *ztf-16(ok3028), sma-2(ok3109), VT1066 and wIs78(ajm-1::gfp/MH27::GFP and scm- 1::gfp)* obtained from the CGC (*C. elegans* Genetics Center) were propagated at 25°C on NGM (Nematode Growth Medium) plates containing *E. coli* OP50. *let-7(n2853)* was propagated at the permissive temperature of 15°C. Staging for RNAi/qRT-PCR was carried out at 25°C starting with synchronized L1s from an egg prep. The worms were harvested by washing off the plates with M9 buffer and a further three times to remove bacteria. The worms were flash-frozen in liquid nitrogen and the pellets stored at -80°C. Staging for microscopy was carried out at 25°C using synchronized L1 populations on NGM plates seeded with OP50 bacteria. The staging times varied from strain to strain to compensate for any growth delays and depending on the experiment. Staging interval for the

RNAi experiments was every 3 hours over the time course of larval development. Staging intervals for microscopy was every 2 hours over the course of larval development.

RNAi constructs and screens: RNAi constructs for the candidate transcription factors were made using 0.6-1.1 kb of spliced coding sequence cloned into the multiple cloning site of the RNAi vector pL4440 using the TA cloning protocol (Ahringer Lab). Sequences of PCR primers used for making RNAi constructs are provided in **Appendix F**.

RNAi assay: Medium used for RNAi assays was NGM (Nematode Growth Medium) supplemented with 1mM IPTG to induce expression and antibiotic Ampicillin at 100μg/mL. The assays were carried out in (4 x 3) 12-well tissue culture plates such that three genes could be screened on a single plate in triplicate along with control (EV) also in triplicate. The plates were seeded with 100μL of dsRNA expressing HT115 bacteria and left at room temperature until completely soaked up. 20-30 synchronized L1s were plated into each well from an egg prep, fed on the bacteria and scored for lethality at 20°C after ~72h for *let-7* and *VT1066* and ~55h for *N2*. For statistical purposes, each plate was done in duplicates such that at least six well samples were scored for each strain per gene.

qRT-PCR: Total RNA was prepared from the frozen worm pellets using Trizol Ambion TRI-Reagent (#AM9738). The RNA was quantified using the Ribogreen assay and 1μg of total RNA was used as template for synthesis of cDNA using reverse transcriptase (M-MLV RT-Promega). *lin-42A, B* and *C* isoform expression during L1, L2 L3, L4 and early adult stages in worms was relative to expression of *ama-1* in each sample. Expression of the three *lin-42* isoforms was

determined using *ama-1*- specific and *lin-42* isoform-specific qRT-PCR primers and Taqman probes (ABI). Primer and probes sequences are provided in Appendix A.

Crosses and Microscopy: The *ztf-16* mutant strain *ok1916;wIs78* and *ok3028;wIs78* were constructed by crossing wildtype *wIs78* males to the *ok1916* and *ok3028* hermaphrodites. For the *ok1916* mutant, a GFP expressing F1 male was identified and crossed back with the *ok1916* hermaphrodites. GFP expressing hermaphrodite progeny were picked and selfed until the GFP was homozygozed. For the *ok3028* strain, 25 GFP expressing hermaphrodites were identified from the F1 progeny, picked onto individual plates and allowed to lay eggs. The mother worm was then genotyped for the deletion using designed primers. In the event of a bagged or burst mother worm, six progeny were randomly picked and genotyped for the deletion by PCR. The subsequent GFP expressing mutant strains were used for visualizing seam cells by Fluorescence Confocal Microscopy. The *ok3028* genotyping primer sequences are provided in **Appendix G**.

RESULTS

Phylogenetic Footprinting of *lin-42B/C* 5' non-coding sequences and *in silico* screens for candidate *C. elegans* transcription factor genes

To identify *cis*-regulatory elements that control *lin-42* transcription, we identified DNA motifs that are conserved among the *lin-42B/C* 5' non-coding sequences from *C. elegans*, *C. briggsae* and *C. remanei*. Using Genomatix software and databases of known vertebrate and invertebrate transcription factors and transcription factor binding sites, we specifically searched for conserved DNA motifs within the BR1, BR2 and BR3 sequences that we had previously defined by EMSA as potential sites for transcription factors. (Chapter 2) Using Genomatix, we

identified mammalian and *Drosophila* transcription factors that bind within these conserved motifs, and then used NCBI (National Center for Biotechnology Information) and WormBase databases and BLAST (basic local alignment search tool) to find the *C. elegans* genes that encode homologs of these transcription factors. (**Figure 3.3B, Appendix E**)

The results of Phylogenetic Footprinting and *in silico* screens are summarized in figure 3.3B. We identified a total of 30 potential transcription factor binding sites on positive and negative strands: 16 conserved DNA motifs in BR1, 6 conserved motifs in BR2 and 8 conserved motifs in BR3. These 30 sites are predicted to be bound by 37 *Drosophila* and vertebrate transcription factors that belong to 10 different transcription factor families. We identified a total of 127 *C. elegans* homologs of these 37 *Drosophila* and vertebrate transcription factors, but only 119 out of the 127 actually encode transcription factors or have DNA binding domains and may thus function as transcriptional regulators (**Appendix E**). Using WormBase, we found the expression profiles of the 127 *C. elegans* transcription factor homologs. We then sorted the genes into three broad categories: 1) gene is expressed during larval development and in cell/tissue types where *lin-42* is expressed and functions, i.e. hypodermis, specifically the seam cells, vulval cells, and somatic gonad, 2) gene is only expressed during embryonic or adult stages, or in a small subset of cells where *lin-42* is not expressed, e.g. only expressed in one type of neuron and 3) gene expression pattern is not known. The 43 transcription factor genes in category 1, i.e. genes whose expression patterns overlap with *lin-42* expression pattern, were assayed in the *let-7(lf)* lethality suppressor/enhancer RNAi screen. (**Table 3.1**)

Identification of three candidates *nhr-23*, *nhr-25* and *ztf-16* as possible *lin-42* regulators

Candidate *C. elegans* transcription factor genes, identified by bioinformatics analyses of *lin-42B/C* non-coding sequences, were tested as potential enhancers or suppressors of *let-7(lf)* lethality in the RNAi screens. The candidates genes were first put through a primary screen where gene expression was knocked-down by RNAi in wildtype background to check if reduction of gene expression resulted in lethality or in *lin-42(lf)* phenotypes like *egl* (egg-laying defect) or *pvl* (protruding vulva defect). Candidate genes were then knocked-down by RNAi in *let-7(n2853)* strain, followed by RNAi in VT1066 strain (triple deletion mutant for *let-7-fam* microRNAs), and we assayed the level of late-larval/young adult lethality by vulval bursting in the RNAi populations. Both *let-7(n2853)lf* and VT1066 strains show approximately 70% lethality at 20°C, and we used known genetic interactions with *let-7* microRNAs to determine the ranges of lethality/survival that would be used to evaluate a candidate gene as a *let-7(lf)* suppressor, enhancer or non-interactor. For example, we used RNAi of known suppressors of *let-7(lf)* lethality, such as *lin-42*, *hbl-1*, as positive controls for lethality suppression. Thus, if RNAi of a candidate gene resulted in suppression of *let-7(lf)* lethality in the same range, or higher, as these positive control genes, we evaluated the candidate gene as a potential transcriptional activator of *lin-42*. In addition to using RNAi on 'empty vector' as a negative control, we used RNAi of *aha-1, ahr-1, cky-1* and *hif-1, C. elegans* homologs of the Clock and Cycle/Bmal transcription factors that regulate *period*, as negative controls. We (and other labs) had previously determined that these genes not do interact with *lin-42* or other heterochronic genes during larval development, although RNAi of *hif-1* and *cky-1* does result in larval lethality in both wildtype and *let-7(lf)* backgrounds. Our criterion for designating a candidate gene an

enhancer of *let-7(lf)* lethality was that RNAi of the gene result in at least a 50% drop in *let-7(lf)* survival relative to the negative controls.

From the RNAi screens in *let-7(n2853)* and VT1066 backgrounds, we identified three suppressors and at least four enhancers of *let-7(lf)*. (**Figure 3.4**, **Table 3.2**) RNAi of *nhr-23*, a nuclear hormone receptor transcription factor gene, suppressed lethality in both *let-7* and VT1066 screens. RNAi of *ztf-16*, a zinc-finger transcription factor gene, suppressed *let-7(n2853)* lethality but not VT1066 lethality, suggesting that *ztf-16* function may be restricted to late larval stages. *blmp-1*, another zinc finger transcription factor gene, had the opposite phenotype – RNAi of *blmp-1* suppressed VT1066 lethality but not *let-7(n2853)* lethality, suggesting that *blmp-1* function may be restricted to early larval stages. *pax-3*, *fzr-1*, *klf-3* and *zfp-1* were identified as strong enhancers of *let-7(lf)* lethality, and thus potential native regulators of *lin-42*. We decided to continue further analysis of *nhr-23* and *ztf-16* based on what is known of function and spatio-temporal expression patterns of these genes, specifically function in development and expression in same tissues as *lin-42*. Based on published reports of *nhr-23* function overlapping with *nhr-25* [54], we additionally decided to include this nuclear hormone receptor transcription factor in further experiments along with *ztf-16* and *nhr-23*.

Effect of RNAi of *nhr-23*, *nhr-25* and *ztf-16 in* wildtype worms on *lin-42* isoform expression

Once candidate genes that are putative *lin-42* positive regulators had been identified by the *let-7* lethality screens, the next step was to check if the suppression of lethality was in fact as a result of altered or reduced *lin-42* expression. To assay expression of *lin-42* isoform specific transcripts, we used RNAi to knock-down expression of the candidate gene (*ztf-16*, *nhr-23* or *nhr-25*) in developmentally synchronized populations of wildtype/N2 worms and harvested

populations at regular intervals over the time course of larval development. We extracted RNA from these larval populations, and assayed *lin-42* transcript levels by qRT-PCR using isoform-specific primers and probes.

In wildtype worms, *lin-42A* and *lin-42B* isoforms show cycling expression during larval stages while *lin-42C* isoform is expressed at basal levels during larval stages with no apparent cycling. (**Figure 3.5A**) In *C. elegans* worms where *ztf-16* expression was knocked down by RNAi, we observed down-regulation of *lin-42A* and *lin-42B* transcript levels only during only the L4 stage, but no disruption of cycling expression pattern (**Figure 3.5B**) This expression data showing a decrease in *lin-42A* and *B* expression during late larval development is consistent with the RNAi screen results, in which *ztf-16(RNAi*) suppressed lethality in *let-7(n2853)* but not in VT1066, and strongly indicates that ZTF-16 regulates *lin-42* transcriptional expression.

Knockdown of *nhr-23* expression affected both the timing and level of expression of *lin-42A* and *B* isoforms in complex ways, although *lin-42C* expression was essentially the same as in wildtype. (**Figure 3.6B**) Both *lin-42A* and *lin-42B* transcript showed a cyclic expression pattern in *nhr-23(RNAi)* animals but the period of the oscillations was altered. The L2 stage expression peaks in *nhr-23(RNAi)* animals occurred at about 17 hrs as in wildtype, but expression in L2 persisted for longer and both L3 and L4 expression cycles were delayed, occurring at 29 hrs and 35 hrs in *nhr-23(RNAi)* instead of at 23 hrs and 29 hrs in wildtype. Moreover, in *nhr-23(RNAi)* worms *lin-42A* and *B* expression did not fully cycle between 23 and 38 hours as the decline in expression after the 29 hour peak was only 2 to 3-fold and did not fall to zero, compared to a 6 to 9-fold decrease in which transcript levels fell to zero in wildtype animals. (compare **Figures 3.6A and 3.6B**) The levels of *lin-42A* and *B* transcripts were also affected in *nhr-23*(RNAi) worms – both overall transcript levels were lower after 23 hours, with *lin-42B* levels affected

more than *lin-42A* levels. Despite the complexity of the alteration to *lin-42* expression in *nhr-23(RNAi)* animals, these results showing down-regulation of *lin-42A* and *lin-42B* transcripts after L2 stage are consistent with the RNAi screen results, in which *nhr-23(RNAi)* suppresses lethality of both *let-7(n2853)* and VT1066.

Knockdown of *nhr-25* expression had only a subtle effect on *lin-42A* and *B* expression, and as in the previous cases, *lin-42C* expression was the same as in wildtype. (**Figure 3.6C**) In *nhr-25(RNAi)* worms, both *lin-42A* and *B* transcripts showed cycling expression during L2, L3 and L4 stages, although *lin-42A* cycles became desynchronized relative to *lin-42B* and were delayed by approximately 3 hrs after L2 stage. Additionally, both *lin-42A* and *B* transcript levels were depressed in L4 sage compared to their peak levels in L2 and L3 stages. This depression in *lin-42* expression during the L4 stage is consistent with *nhr-25(RNAi)* suppression of lethality in *let-7(n2853)*. *nhr-25(RNAi)* also suppresses lethality in VT1066 although *lin-42* levels are relatively normal in *nhr-25(RNAi)* worms during early larval development. The suppression of VT1066 lethality by *nhr-25(RNAi)* during L2 and L3 stages must thus be due to mechanisms that do not involve *lin-42* expression. For example, NHR-25 may instead affect *hbl-1* expression during late L2 stage.

We decided to focus on investigating *ztf-16* as a transcriptional activator of *lin-42* and to discontinue analysis of *nhr-23* and *nhr-25* for the following reasons. The spatial expression pattern of *ztf-16* is similar to that of *lin-42*, that is, *ztf-16* is expressed in the seam cells and somatic gonad during larval development [143]. The temporal expression patterns of *ztf-16* and *lin-42* are also similar: Whole genome microarray expression data of wild type *C. elegans* shows that *ztf-16* is cyclically expressed with peaks of expression that precede *lin-42* induction. (**Figure 3.7A**) The *ztf-16* spatio-temporal expression pattern suggests that the cyclic expression pattern of

lin-42 may be driven by periodic activation by oscillating levels of ZTF-16. Furthermore, *ztf-16* is a paralog of *hbl-1*, a heterochronic transcriptional repressor that controls temporal expression of *let-7* microRNA [143]. In contrast to *ztf-16*, *nhr-23* mRNA expression shows peaks exactly coinciding with that of *lin-42,* thus, making it an unlikely direct positive regulator of *lin-42*. (**Figure 3.7B**) *nhr-25* expression shows no significant cycling during the larval stages, and in combination with the subtle effect of *nhr-25(RNAi)* on *lin-42* transcript levels, makes *nhr-25* a weak candidate for a *lin-42* transcriptional regulator (**Figure 3.7C**)

Expression levels of *lin-42* isoforms in *ztf-16 (ok3028)* mutants

RNAi knockdown of a gene by feeding has variable efficiency. Based on the altered levels of *lin-42* isoforms in *ztf-16(RNAi)* worms, we predicted a stronger or more dramatic reduction in expression levels of *lin-42* isoforms in a *ztf-16* deletion mutant strain. To test this, we obtained two deletion strains, *ztf-16(ok1916)* and *ztf-16(ok3028)*. The *ok1916* strain is a 1005 bp deletion mutant with a very mild, almost wildtype phenotype compared to the more severe *ok3028* strain which is a 571 bp deletion that disrupts more of the C2H2 Zn-finger domains that are critical for function [144]. We thus carried out isoform specific qRT-PCR in the *ztf-16(ok3028)* mutants over the time-course of larval development (spanning late L1 through adult stages). In these mutants, *lin-42A* and *lin-42B* isoform levels were altered to a greater extent than with RNAi of *ztf-16* in the wildtype worms. (**Figure 3.5C**) In *ztf-16(ok3028)*, none of the *lin-42* isoform transcripts show cycling expression in either early or late development, and both *lin-42A* and *lin-42B* levels are depressed by 2 to 6 fold compared to expression levels in wildtype. This indicates that *ztf-16* is required, but is not sufficient, for *lin-42* expression.

SC division analysis in *wIs78* (wildtype) versus *ztf-16 (ok3028);wIs78*

ztf-16(ok3028) development is overtly wildtype, at least at the gross morphological level, despite the fact that *lin-42* expression is severely depressed in these mutant worms. To see what effect, if any, reduced levels of *lin-42* expression have on various aspects of *lin-42*-mediated development, we characterized seam cell division and differentiation in *ztf-16(ok3028)* mutants.

Newly hatched wildtype *C. elegans* larvae have two sets of lateral seam cells in the hypodermis (H0, H1, H2, V1-V6 and T). (**Figure 3.8A**) Each of these cells, except for H0, undergoes larval stage-specific patterns of asymmetric divisions. During each asymmetric seam cell division, the anterior daughter cell differentiates and fuses with the hyp7 syncytium, while the posterior cell does not differentiate and continues to divide until the end of the larval stage, when it terminally differentiates and fuses with other seam cells to form the seam syncytium [9]. Exit from the cell cycle, seam cell fusion and subsequent alae secretion are terminal differentiation events that mark the transition from larva to the adult stage, and are cellular events that are regulated by heterochronic genes, such as *lin-42*. *lin-42(lf)* mutant strains like *n1089, mg152* and *ok2385* exhibit heterochronic phenotypes like precocious seam cell fusion and alae secretion. We predicted that *ztf-16(ok3028)* mutant worms, with depressed expression of *lin-42*, would show similar precocious heterochronic phenotypes.

To analyze seam cell division and fusion, we created a *ztf-16(ok3028)* strain expressing GFP in the seam cell nuclei and seam cell adherens junctions (*ztf-16(ok3028;wIs78)*). Developmentally synchronized populations of the seam cell GFP reporter strains in wildtype genetic background (*wIs78)* and *ztf-16(lf)* background (*ztf-16(ok3028;wIs78))* were scored for seam cell number every 2 hours over the time-course of larval development. Since *ztf-*

16(ok3028) worms development is delayed relative to wildtype by approximately 4 to 6 hours, we extended the time line for sampling seam cell number in *ztf-16(ok3028)* from 36 hours to 40 hours. We also monitored somatic gonad development in the same worm populations so that we could more directly compare late larval stages between wildtype and *ztf-16(ok3028)* worms. In *C. elegans,* two arms of the somatic gonad, located in the mid-body where the vulval opening will form, begin to extend anteriorly and posteriorly during L1, L2 and L3 stages. During late L3 stage, the extending tips of the gonad arms flex dorsally (designated gonad migration stage 1.1) and then extend back towards the mid-body during the L4 stage (gonad migration stages 1.25, 1.5. 1.75) until they meet at mid-body above the region where the vulva has formed at the young adult stage (gonad migration stage 2.0). (**Figure 3.9**) Thus gonad migration stages of 1.1, 1.25, 1.5, 1.75 and 2.0 correspond to developmental stages late-L3, early-L4, mid-L4, late-L4 and young adult respectively. (**Figure 3.9**) The seam cell division pattern in *ztf-16(ok3028);wIs78* compared to wildtype *wIs78* shows that early development is equivalent in the two strains – the timing and extent of L2 division patterns, and entry into L3 and L3 divisions look the same in the two strains (**Figure 3.8B**). However, in *ztf-16(ok3028);wIs78,* seam cells take twice as long to complete the L3 divisions and progress to the L4 stage divisions. If the seam cell divisions in the L4 stages of the two strains are compared using the extent of gonad migration instead of hours, then we find that the timing of the L4 division is again comparable – in both strains, seam cells take about 6 hours to complete the L4 divisions. However, fewer seam cells undergo division at L4 in *ztf-16(ok3028);wIs78* worms, as indicated by the lower number of seam cell nuclei observed compared to *wIs78*. (**Figure 3.8B**) This suggests that some seam cells in *ztf-16(ok3028);wIs78* may be precociously exiting the cell cycle and undergoing terminal

differentiation. To test this hypothesis we measured seam cell terminal differentiation phenotypes in *ztf-16(lf)* worms.

SC fusion and alae analysis in *wIs78* (wildtype) versus *ztf-16(ok3028);wIs78*

Terminal differentiation of seam cells at the end of larval development in *C. elegans* involves exit from the cell cycle, fusion of seam cells along their lateral axes to form the seam syncytium and secretion of adult-specific collagens that form alae. In relation to other developmental events such as gonad arm migration (GM) that occur during L4 stage, seam cell fusion in wildtype worms can occur as early as mid L4 stage (less than 20% of the population has partial seam cell fusion at GM 1.5) but the majority of fusion occurs during late L4 stage (more than 90% of the population showed complete seam cell fusion at GM 1.75). (**Figure 3.10B, left panel**). In contrast, in *ztf-16(ok3028);wIs78* about 65% of the population had undergone complete seam cell fusion by mid L4 stage (GM 1.5).

The developmental transition of *C. elegans* larvae into the reproductively active adult stage is marked by terminal differentiation of the seam cells and formation of the syncytium, followed by secretion of the adult-specific cuticle known as alae. (**Figure 3.10B**) These events occur precociously in *lin-42 (lf)* mutants. Hence, we predicted that in the absence of a likely positive regulator of *lin-42*, the mutant *ztf-16 (ok3028)* strain would exhibit these precocious phenotypes. Our earlier Seam cell number analysis pointed to a mild precocious seam cell differentiation pattern and to corroborate our hypothesis we carried out analysis of seam cell terminal fusion and alae formation in both the wildtype control worms *wIs78* and the *ztf-16 (ok3028);wIs78* strain. For this analysis we staged the worms by an egg prep and fed them on OP50 at 25°C and scored them for developmental age, seam cell number, seam cell fusion and

alae formation from the early to mid- L3 stage (gonad migration of 1.0) through the adult stage (gonad migration of 2.0). Developmental age of the worm was determined by the gonad migration. The previous findings showed that in the *ztf-16 (ok3028)* mutants, fewer seam cells undergo division in the L4 stage and they exit the cell cycle much earlier than the wildtype worms at the same developmental stage- indicative of precocious seam cell development. (**Figure 3.8B**) This precocious seam cell fusion phenotype is similar to that seen in *lin-42(n1089)* and *lin-42(mg152),* and suggests that adult alae formation may also be precocious in the *ztf-16(lf)* mutant [44]. We found that this was indeed the case. In wildtype worms, full expression of alae was observed as early as late L4 stage (less than 15% of the population had any alae at GM 1.75), with most worms showing expression of complete alae that extended the entire length of the body at the young adult stage (GM 2.0). In *ztf-16 (ok3028)* worms, we observed complete alae formation in mid L4 stage worms (approximately 8% of the population had complete alae at GM 1.5). (**Figure 3.11**) The seam cell division pattern data, seam cell fusion data and alae formation data together indicate that the decrease in *lin-42A/B* levels in *ztf-16(ok3028)* mutants results in precocious seam cell differentiation but does not affect the overall pattern of stage-specific seam cell divisions.

The relationship between *ztf-16* and *sma-2*

In order to determine how *ztf-16* expression and function is regulated, we looked for known genetic or molecular interactors of *ztf-16* using NCBI Aceview databases [145, 146]. We found SMA-2 listed as a physical interactor with ZTF-16 in Yeast two hybrid assay. *sma-2* is a component of the transforming growth factor-β (TGF-β) pathway, and is expressed in seam cells, similar to *ztf-16* [147, 148]. The TGF-β pathway regulates the dauer development pathway

of which *lin-42* is a core component. These pieces of evidence led us to hypothesize that *sma-2* modulates *ztf-16* function and thus, *sma-2(lf)* would also affect *lin-42* expression. To test this hypothesis we assayed for *lin-42* mRNA levels in a *sma-2* deletion mutant, *sma-2*(*ok3109)*. The *sma-2* mutant worms were staged and harvested over the time-course of late larval development and *lin-42* isoform transcripts were assayed by qRT-PCR. Disappointingly, we found that *lin-42* expression in *sma-2*(*ok3109)* mutant animals is essentially wildtype. (**Figure 3.11B**) Furthermore, microarray expression data reveals that *sma-2* is expressed at a constant low level during larval development. (**Figure 3.11C**) Thus, neither *sma-2* expression pattern nor the effect of loss of *sma-2* on *lin-42* expression indicates that *sma-2* regulates *ztf-16* to affect *lin-42* expression or function.

DISCUSSION

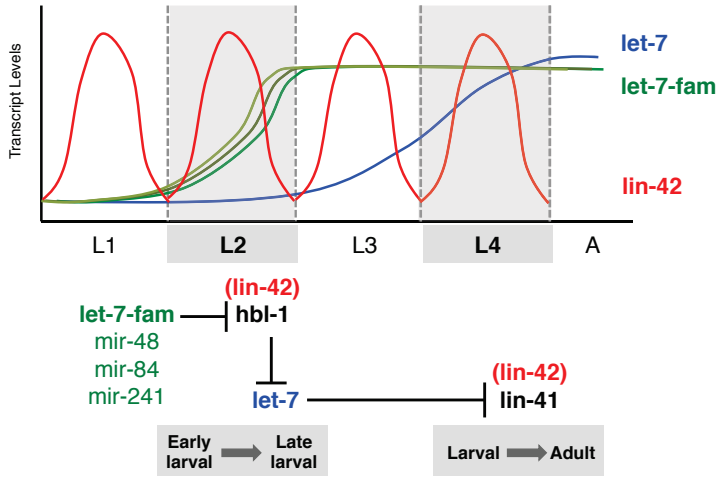
C. elegans ztf-16 (R08E3.4) is a C2H2 zinc-finger transcription factor that is a member of the Hunchback/Ikaros-like (HIL) family that are important regulators of development, specifically stem cell development, such as hematopoiesis [143, 149]. *ztf-16* encodes proteins with up to eight zinc fingers, similar to *Drosophila hunchback* and the Mammalian *Ikaros* family members, which have a specific arrangement of C2H2 zinc fingers - four at the N-terminal responsible for DNA-binding activity and two at the C-terminal that mediate protein-protein interactions [143, 150]. Yeast two-hybrid screens had previously identified the gene product of *R08E3.4* as a putative LIN-42-interacting protein [151]. *ztf-16* was also identified in a Mos1 transposon mutagenesis screen to identify suppressors of the temperature sensitive lethal phenotype of *let-7 (n2853)* [152]. Recent studies have shown that *ztf-16 (R08E3.4)* is involved in gonad development and in glial remodeling in the sensory organs of the worm [143, 153].

Bioinformatics analyses identified *ztf-16* along with over a hundred other candidate transcription factors as homologs of Drosophila and vertebrate transcription factors that bind phylogenetically conserved DNA motifs (BR1, 2, 3) found in the *lin-42B/C* promoter region. Forty-three of the candidate genes, including *ztf-16*, were tested in the *let-7(lf)* lethality suppression/enhancement RNAi screen. *nhr-23, nhr-25* and *ztf-16* were short-listed as candidates that suppressed *let-7(lf)* lethality, and were then tested further for a direct influence on *lin-42* mRNA levels using qRT-PCR. Subsequent review of the literature and our own preliminary data led to *ztf-16* emerging as our top candidate for a possible regulator of *lin-42* expression. Microarray data shows the expression patterns of *lin-42* and *ztf-16* from time-course microarray of N2-wildtype *C. elegans* over larval development. *ztf-16* transcripts are also cyclically expressed prior to *lin-42* induction during multiple larval stages consistent with what is expected of a potential positive regulator. Spatial expression pattern of ZTF-16 is much like that of LIN-42. It is also expressed in the seam cells of the hypodermis in larval worms and in the distal tip cells [144]. Furthermore, *ztf-16* is a paralog of *hbl-1* also a zinc-finger transcription factor which functions with *lin-42* during the L2 stages in the heterochronic pathway [143]. Our subsequent analysis of N2 worms on *ztf-16* RNAi showed de-regulated *lin-42* isoform levels by qRT-PCR. These findings were corroborated in the mutant deletion strain *ztf-16(ok3028)*. Further validation from seam cell and alae analyses supported our hypothesis that a true positive regulator should phenocopy *lin-42(lf)*. All these data together indicate that *ztf-16* is a likely positive regulator of *lin-42* with obvious effects on *lin-42* isoform expression levels.

However, the question remains as to whether *ztf-16* is a direct or indirect regulator of *lin-42*. From the data, we know that although there is an obvious abolishment of the cycling of the *lin-42A* and *lin-42B* isoforms, expression was not completely repressed. This may be because of

redundancy of function among *ztf-16* isoforms, some of which are not affected by the *ztf-16(ok3028)* mutation. Alternatively, ZTF-16 may function with other transcriptional regulators to control *lin-42* expression. The *lin-42* EMSA screen (Chapter 2) identified three distinct binding regions within 1700 bp of the upstream non-coding region of *lin-42B/C* open reading frame. Of these three regions, we found two putative ZTF-16 GATA binding sites within BR1 and BR2 that are perfectly conserved among *C. elegans*, *C. briggsae* and *C. remanei*. (**Figure 3.3C**) There are 11 other potential ZTF-16 GATA-binding sites in the 1700 bp of *lin-42B/C* non-coding sequence but none of these is conserved among all three nematode species. We showed that *ztf-16(lf)* mutant phenocopies *lin-42(lf)* in causing precocious seam cell terminal differentiation. *lin-42* loss-of-function also results in aberrant dauer, lethargus and ecdysis phenotypes, and the next step in analysis of *ztf-16* regulation of *lin-42* would be to test whether similar phenotypes are also observed *ztf-16(lf)* mutants [46, 47].

A



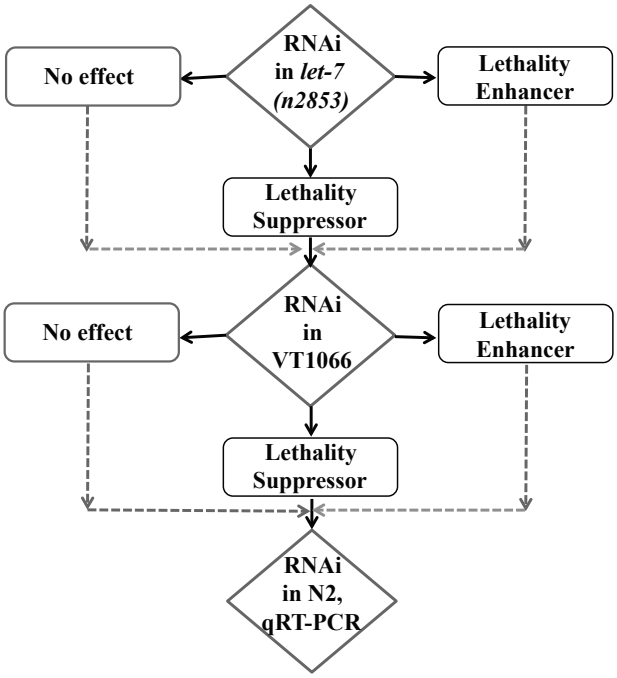
B

Genetic Background	RNAi	Expression at L2 --> L3 or L4 --> A		Phenotype
		let-7 miRs	*lin-42*	
Wildtype	--	↑	↓	Wildtype
let-7(lf)	--	↓	↑	~70% Lethality @ 20°C
let-7(lf)	*lin-42*	↓	↓	Suppression: < 70% lethality @ 20°C
let-7(lf)	TF-A	↓	↓↓	Suppression: < 70% lethality @ 20°C
let-7(lf)	TF-R	↓	↑↑	Enhancement: > 70% lethality @ 20°C

Figure 3.1: Expression patterns of *let-7* family microRNAs during larval development and logic of the *let-7(lf)* lethality suppression/enhancement RNAi screen.

(A) Schematic of stage-specific expression of *let-7* family microRNAs and *lin-42*. *miR-48, miR-84* and *miR-241* (*let-7-fam*) are expressed starting in early larval development (green) and function during late L2 (grey) to enable progression to the next developmental stage. *let-7* microRNA expression (blue) is inhibited during early larval stages and functions during late L4 (grey) to enable terminal differentiation and progression to the adult stage. The cyclic expression pattern of *lin-42* is shown in red. Figure based on [14, 17, 39] .
(B) The logic of the *let-7(lf)* lethality suppression/enhancement RNAi screen. The phenotypes of *let-7(lf)* strains (i.e. *let-7(n2853)* and VT1066), as well as expected phenotypes in the event of knocking down a potential positive/negative regulator of *lin-42* are indicated. TF(A): transcriptional activator of *lin-42*; TF(R): transcriptional repressor of *lin-42*. Upward pointing arrow indicates up-regulation of expression/function, downward pointing arrow indicates down-regulation of expression/function, and number of arrows indicates degree of up or down-regulation.

A



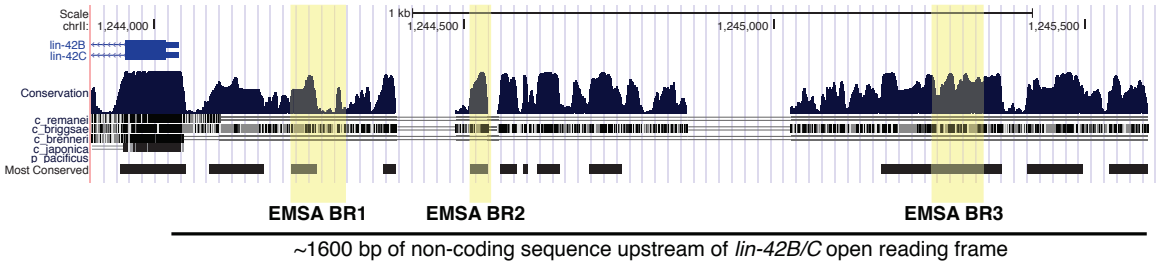
B

let-7(n2853) lf	VT1066 *(miRs-48,-241,-84) lf*	Interpretation
No effect	No effect	TF does not regulate lin-42
No effect	Suppress or Enhance	TF is stage-specific, acts at L2, not L4
Suppress or Enhance	No effect	TF is stage-specific, acts at L4, not L2
Suppress	Suppress	TF is not stage-specific, acts at L2 & L4- Same regulatory role at both stages
Enhance	Enhance	
Suppress	Enhance	TF is not stage-specific, acts at L2 & L4- Different regulatory role at both stages
Enhance	Suppress	

Figure 3.2: Flowchart of *let-7(lf)* lethality suppression/enhancement RNAi screen and interpretation of possible results.

(A) Flowchart showing sequence of RNAi assays *let-7(n2853)* and VT1066 and qRT-PCR of candidate suppressor or enhancer at the end of the screen to validate regulation og *lin-42* expression.
(B) Interpretation of the possible results of the RNAi screens. Refer to text for detailed description.

A



B

Binding Region	# Conserved TFBS motifs	TF Families	Vertebrate & *Dm* TFs	*C. elegans* TFs
BR1	16	10	37	119
BR2	4			
BR3	8			

C

EMSA Binding Region	Matrix Family	Family Information	Matrix	Vertebrate TFs	*C. elegans* Homolog	*C. elegans lin-42B/C* Promoter		
						Anchor position	Strand	Sequence
BR1	Vertebrate GATA	GATA-sequence binding factors	GATA.01	Gata1, Trps	ZTF-16	338	+	aactGATAaaatg
BR2	Vertebrate GATA	GATA-sequence binding factors	GATA1.05	Gata1	ZTF-16	559	+	tataGATAaagtg

Figure 3.3: Phylogenetic Footprinting and results of *in silico* screens of *lin-42B/C* 5' non-coding sequences.

(A) UCSC genome browser view of alignment and conservation of *lin-42B/C* 5' non-coding sequences between *C. elegans* and related nematode species. The blue peaks, black and grey bars are representative of the degree of sequence conservation across the closely and distantly related worm species. Areas of highest sequence conservation are represented by the highest blue peaks and the black boxes at the bottom. The identified sequences that gave specific binding in Electrophoretic mobility shift assays (EMSA, Chapter 2) are shown highlighted in yellow and designated as Binding Regions BR1, BR2 and BR3.
(B) Summary of *in silico* screens for *C. elegans* transcription factors (TFs) that may bind to phylogenetically conserved transcription factor binding sites (TFBS) within BR1, 2 and 3. Using Genomatix software and databases, we identified a total of 28 potential TFBSs within BRs 1, 2 and 3. These TFBSs are predicted to or are known to be bound by 37 different vertebrate and *Drosophila melanogaster* (Dm) transcription factors that belong to 10 different transcription factor families. Using NCBI BLAST and WormBase, we identified 119 *C. elegans* homologs of these 37 vertebrate and *Drosophila* transcription factors.
(C) Sequence and location of the putative ZTF-16 GATA binding sites in BR1 and BR2 regions identified as potential sites of transcription factor binding in EMSA screen (Chapter 2).

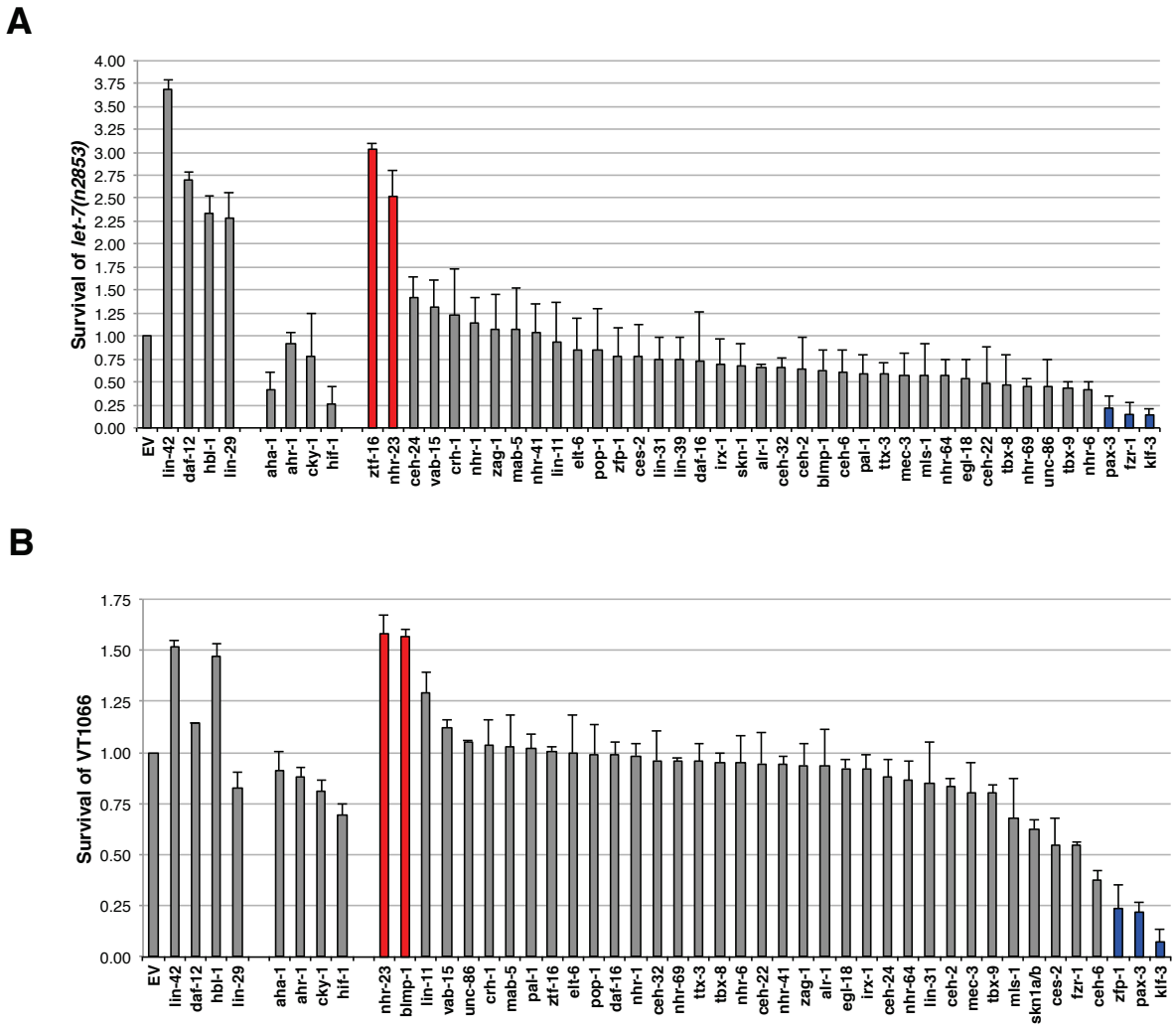


Figure 3.4: Results of the *let-7(lf)* lethality suppression/enhancement RNA interference screen in *let-7(n2853)* and VT1066 mutant genetic backgrounds.

(A) Survival of *let-7(n2853)* on RNAi of candidate transcription factors relative to survival on mock RNAi of empty vector (EV).
(B) Survival of VT1066 on RNAi of candidate transcription factors relative to survival on RNAi of empty vector (EV).
The first five bars on both graphs represent survival on Empty Vector or mock RNAi and relative survival on the positive control genes *lin-42*, *daf-12*, *hbl-1* and *lin-29*, which are known to suppress lethality of *let-7 (n2853)* and VT1066. The next four bars on both graphs represent relative survival on RNAi of the homologs of circadian transcription factors *clock*, *cycle* and *bmal*: *aha-1*,*ahr-1*, *cky-1* and *hif-1*. The third set of bars on both graphs represent the relative survival of *let-7(n2853)* and VT1066 on RNAi of candidate transcription factors tested compared to that on EV(RNAi). The error bars represent the standard deviation. The bars indicated in red are putative positive regulators and the ones in blue are putative negative regulators of *lin-42*.

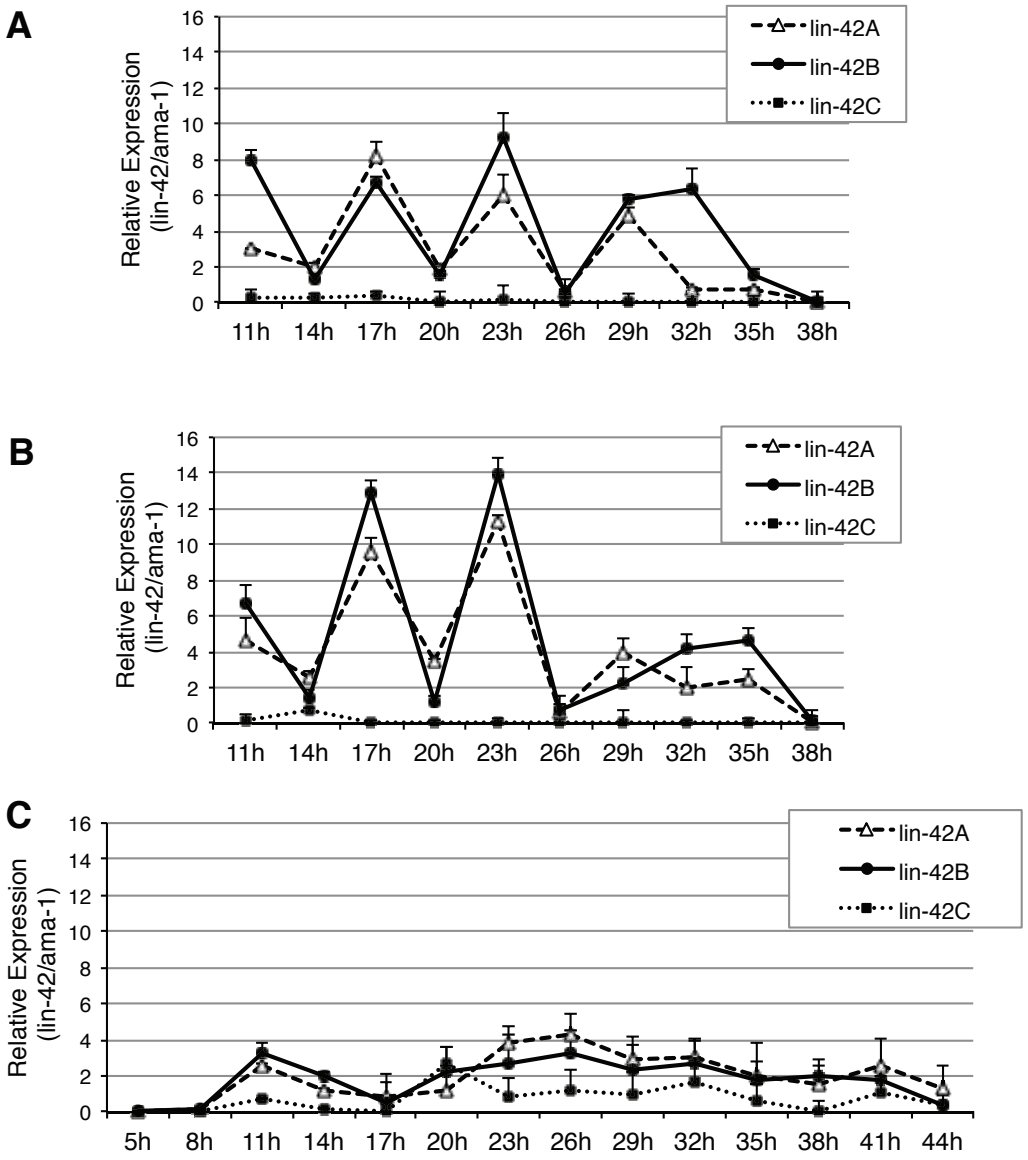


Figure 3.5: Expression of *lin-42* isoforms in wildtype and *ztf-16(lf)* genetic backgrounds.

(A) *lin-42 A,B,C* isoform expression in N2-wildtype worms on mock RNAi (RNAi of empty vector)**.**
(B) *lin-42 A,B,C* isoform expression in *ztf-16(RNAi)* worms, i.e. N2-wildtype worms on RNAi of *ztf-16*.
(C) *lin-42 A,B,C* isoform expression in *ztf-16(ok3028)* mutant worms.
Transcript levels were measured by quantitative real-time PCR (qRT-PCR) using *lin-42* isoform specific primers and probes (Appendix A). Transcript values are represented relative to the expression of the internal reference gene *ama-1* in the same sample. The x-axis represents hours post addition of synchronized L1 worms to RNAi bacteria at 25°C. Within a single time-course experiment, qRT-PCR reactions were run in triplicate for each *lin-42* isoform or *ama-1* at each time point. The error bars represent standard deviation.

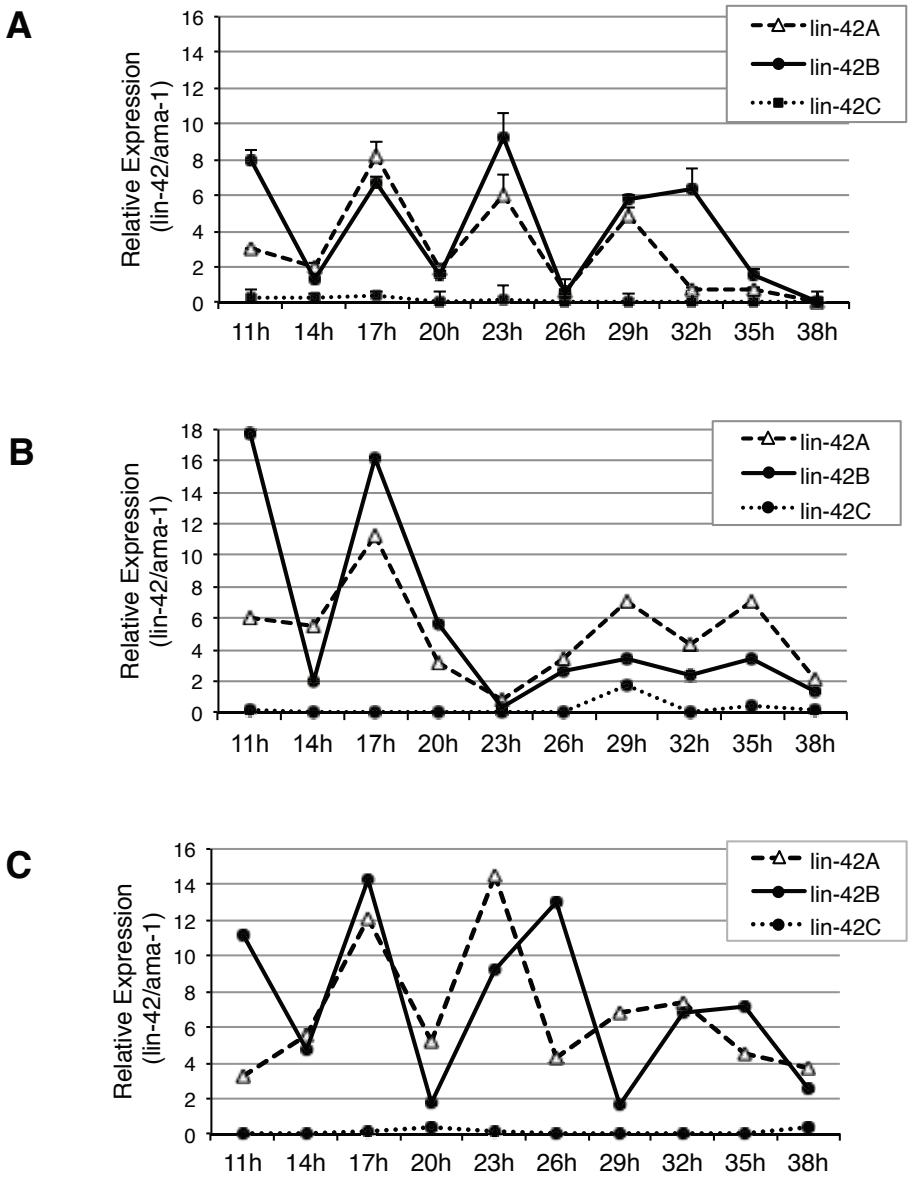


Figure 3.6: *lin-42* isoform-specific expression over larval stages L2, L3 and L4 in N2/wildtype worms on RNAi of EV, *nhr-23* and *nhr-25*.

(A) *lin-42 A,B,C* isoform expression in N2-wildtype worms on mock RNAi (RNAi of empty vector)**.**
(B) *lin-42 A,B,C* isoform expression in *nhr-23(RNAi)* worms.
(C) *lin-42 A,B,C* isoform expression in *nhr-25(RNAi)* worms.
Transcript levels were measured by quantitative real-time PCR (qRT-PCR) using *lin-42* isoform specific primers and probes (Appendix A). Transcript values are represented relative to the expression of the internal reference gene *ama-1* in the same sample. The x-axis represents hours post addition of synchronized L1 worms to RNAi bacteria at 25°C. Within a single time-course experiment, qRT-PCR reactions were run in triplicate for each *lin-42* isoform or *ama-1* at each time point. The error bars represent standard deviation.

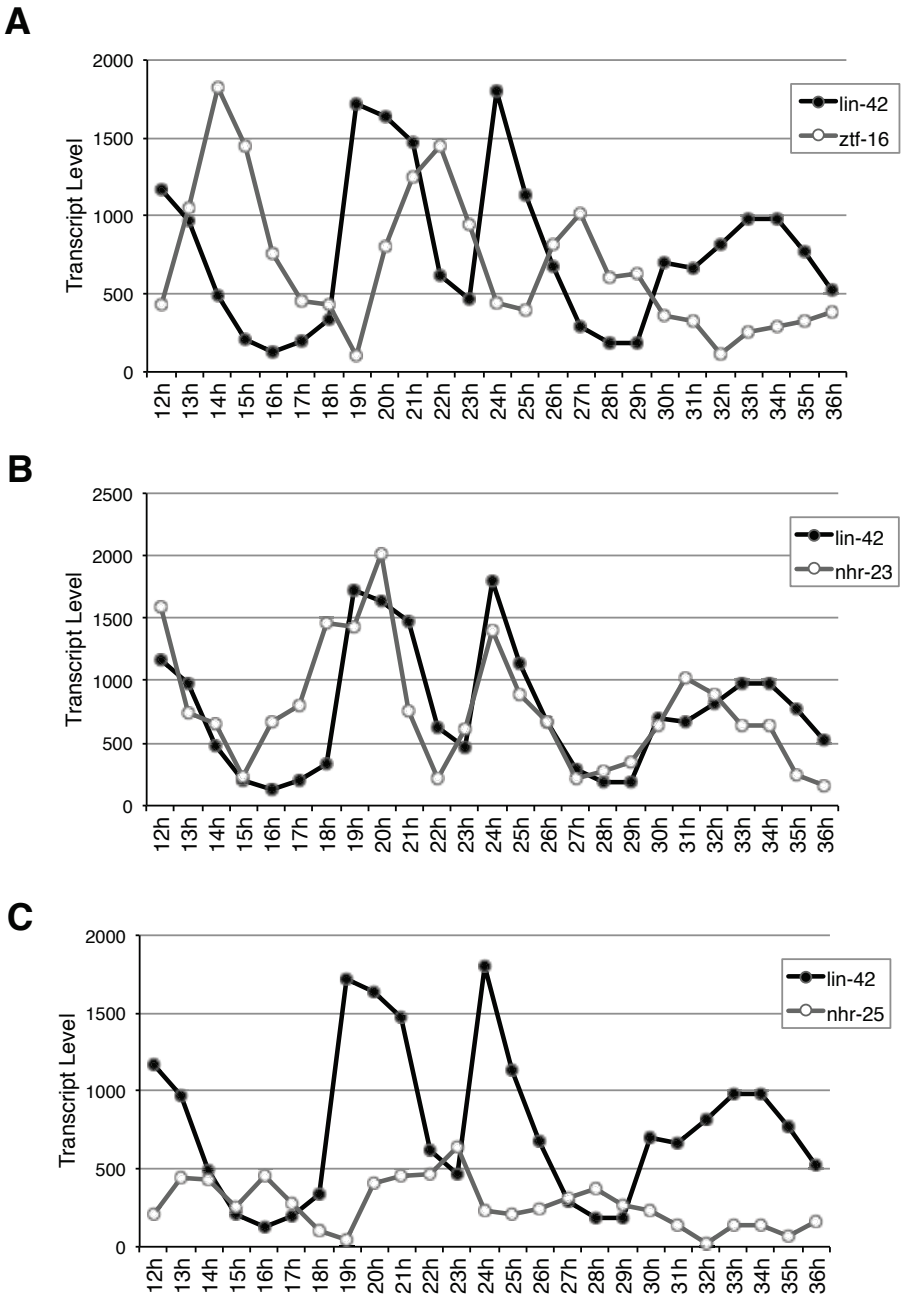


Figure 3.7: Microarray expression data of mRNA levels of *ztf-16, nhr-23* and *nhr-25* compared to that of *lin-42* in N2 wildtype worms over time-course of larval development.

(A) *ztf-16* mRNA expression compared to that of *lin-42*.
(B) *nhr-23* mRNA expression compared to that of *lin-42*.
(C) *nhr-25* mRNA expression compared to that of *lin-42*.
The x-axis represents time in hours (h) post addition of synchronized L1 worms to OP50 bacterial food at 25°C. Microarray data provided by DB (Banerjee Lab unpublished).

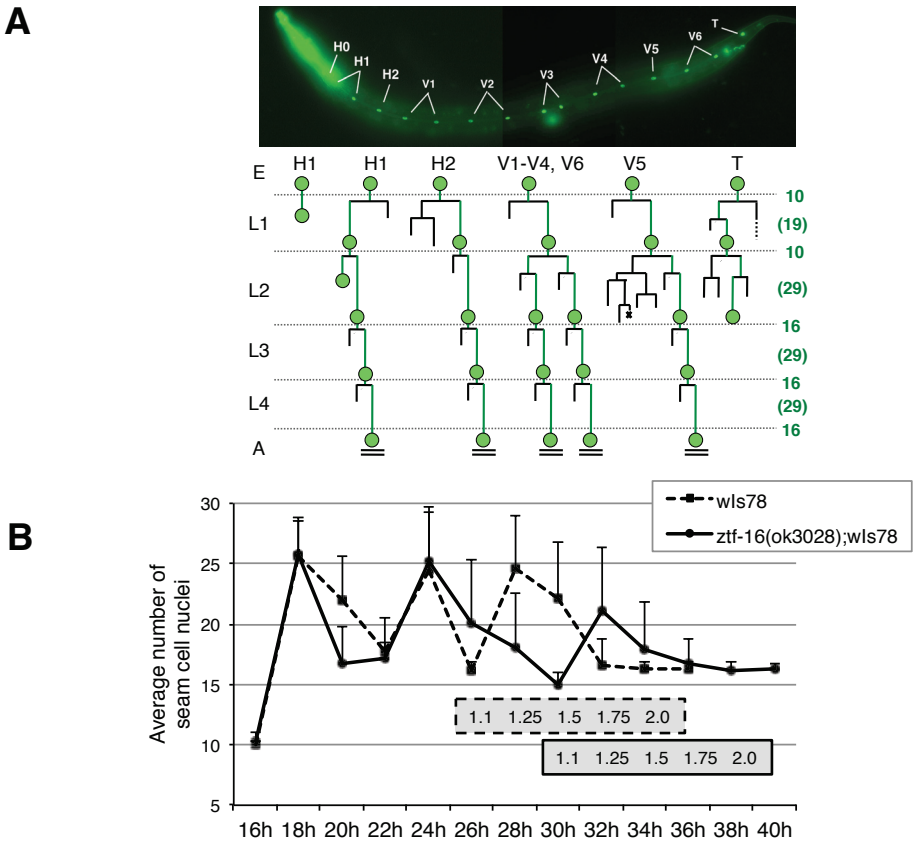


Figure 3.8: Seam cell division patterns during larval development in wildtype and *ztf-16(ok3028)* mutants.

(A) A schematic drawing of the wildtype seam cell division patterns from L1 to L4 stages that eventually give rise to 16 seam cells in the adult (A). Embryos (E) hatch into L1 larvae that have 10 seam cells (H0-H2, V1-V6 and T). Stage-specific divisions of seam cells give rise to a fixed number of seam cells at the end of each larval stage (indicated by the numbers in green). The numbers in parentheses (green) indicate the total number of cells within the seam after division of seam cells but before migration of differentiated seam cell daughters out of the seam. The vertical bars in green represent the seam cell division pattern. The vertical black bars represent the daughter cells that fuse with the hyp7 and the green bars ending in the green circles represent the posterior daughter cell that retains proliferative capacity. The black horizontal double bars represent terminal differentiation and secretion of alae.
(B) The seam cell division pattern in wildtype (*wIs78*) versus *ztf-16(ok3028);wIs78* worms. For each worm strain at each time-point, we counted the GFP marked seam cell nuclei in a single worm, for 25 to 30 worms from a synchronized population. Each time point count is the average of at least three independent time-course experiments. The x-axis refers to the time in hours post addition of synchronized L1 larvae to OP50 bacterial food at 25°C. The error bars represent standard deviation. The extent of L4 stage gonad migration for each strain is indicated in the grey boxes placed approximately in line with the times of analysis corresponding to the developmental stages of the worm (box with dotted lines: *wIs78* and box with solid lines: *ztf-16(ok3028);wIs78*).

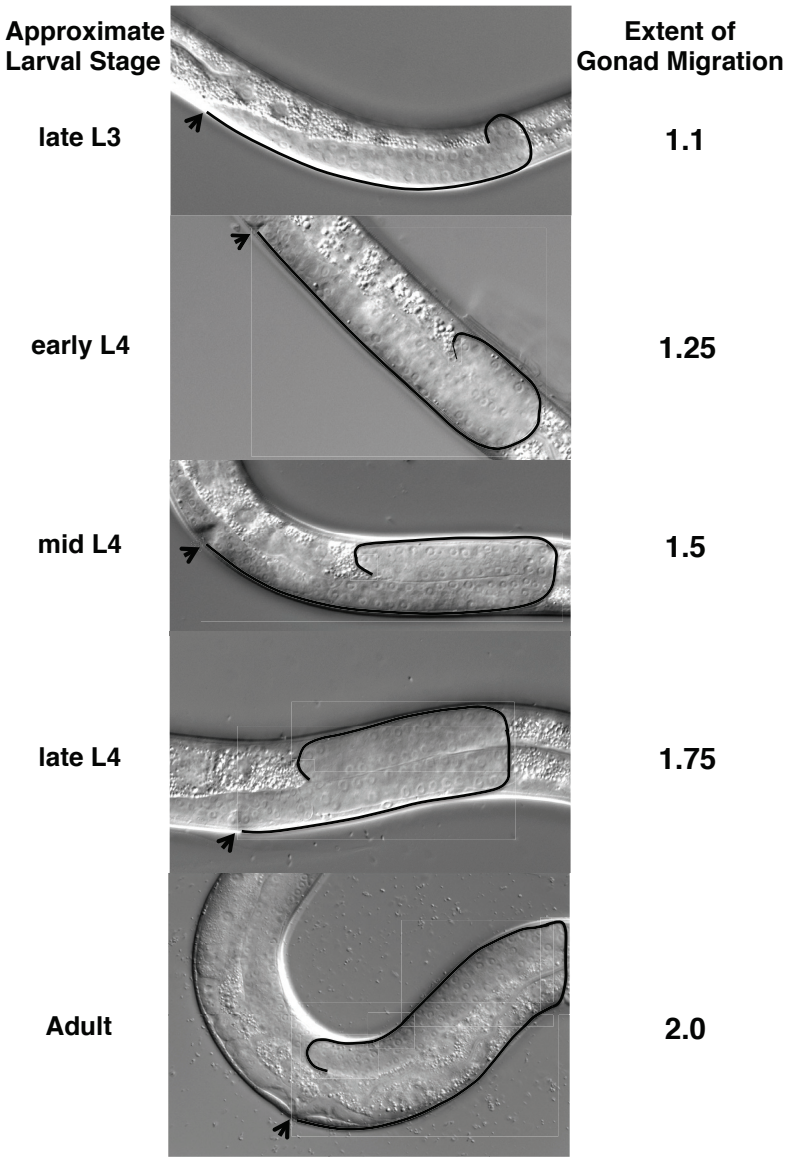


Figure 3.9: Extent of gonad migration and the corresponding developmental age of the larval worm.

The black line outlines a single gonadal arm of the worm and shows the extent of migration of the gonad arm tip relative to the position of the vulva (black arrowhead). The extent of gonad migration (GM) is designated by the numbers 1.1, 1.25, 1.5, 1.75 and 2.0 corresponding to the stages late L3, early L4, mid L4, late L4 and adult, respectively.

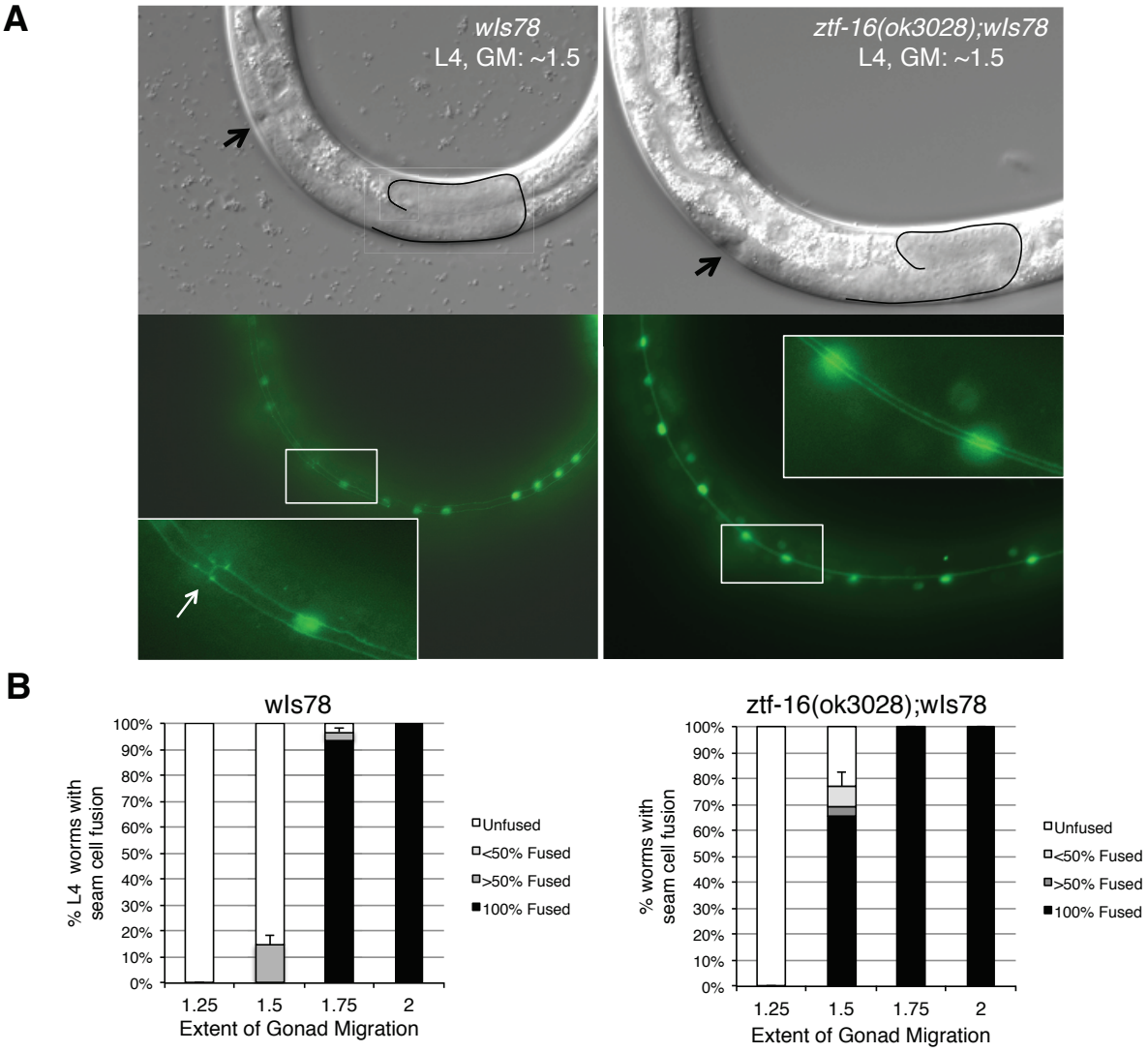


Figure 3.10: Seam cell fusion in wildtype and *ztf-16(ok3028)* mutants.

(A) Top Left: DIC image of a L4 stage wildtype (*wIs78*) worm showing gonad migration of ~1.5. Bottom Left: GFP image of the same worm (*wIs78*) showing fluorescent seam cell nuclei and intact lateral cell junctions (white arrow in the enlarged section). Top Right: DIC image of L4 stage *ztf-16 (ok3028);wIs78* mutant worm showing gonad migration of ~1.5. Bottom Right: GFP image of the same worm (*ztf-16 (ok3028);wIs78*) showing fluorescent seam cell nuclei and completely fused cell junctions along the lateral axis (enlarged section of the image). The black line outlines the gonadal arm of the worms and shows the extent of the gonadal arm end relative to the position of the vulva (black arrow).
(B) Left panel: Seam cell fusion in wildtype *wIs78* worms from early-L4 to young adult stage. Right panel: Seam cell fusion in *ztf-16(ok3028)* worms from early-L4 to young adult stage.
For each strain, the developmental age, i.e. early L4 to young adult, was determined by the extent of gonad migration (1.25, 1.5, 1.75 or 2.0). At each stage for both strains, 25 to 30 worms were scored for the extent of lateral seam cell fusion and for formation of alae. Error bars represent standard deviation.

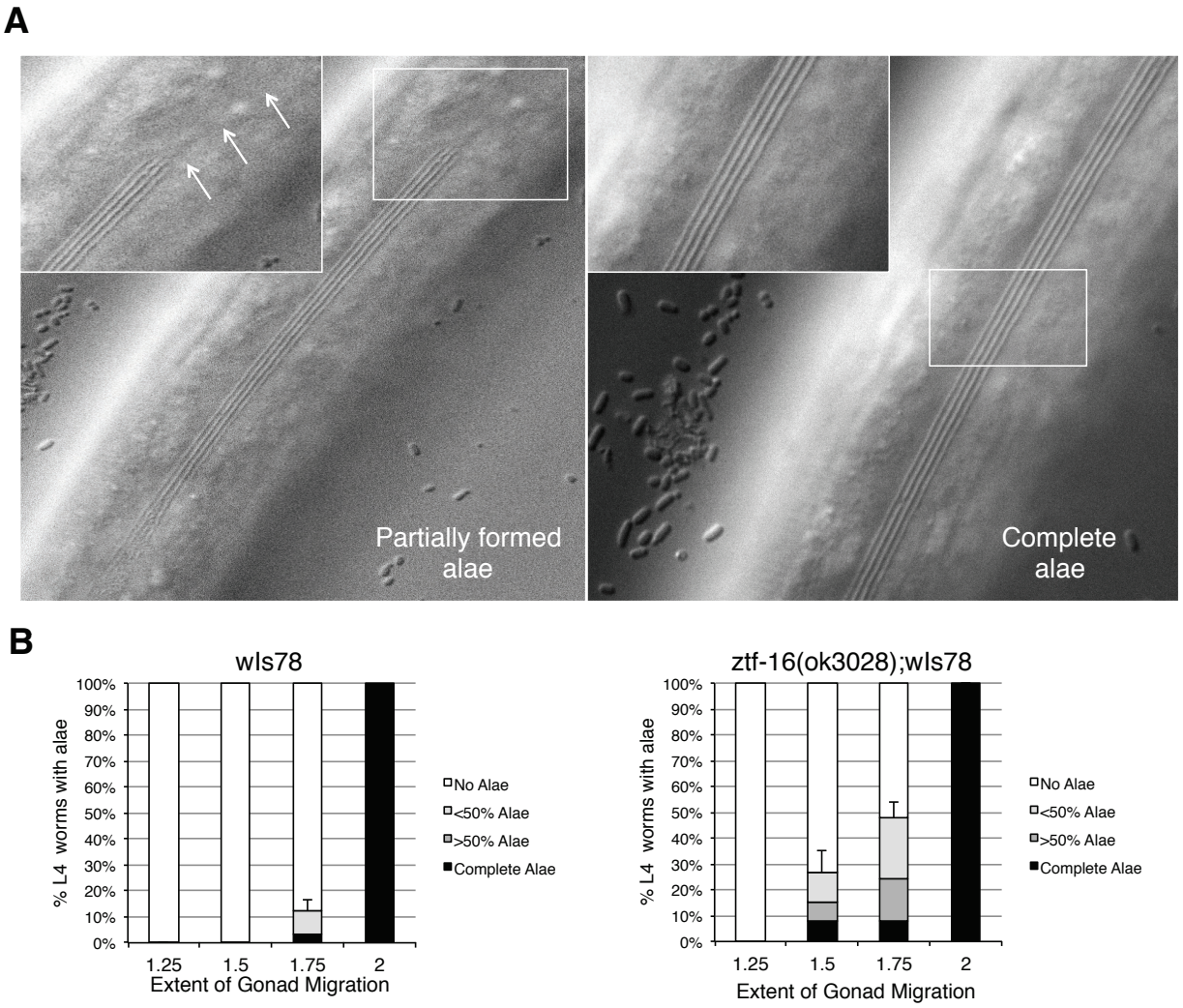


Figure 3.11: Alae analysis across early-L4 to the adult stages in *ztf-16(ok3028)* versus wildtype *(wIs78*) worms.

(A) Left panel: DIC image of a worm showing partially formed alae. White arrows indicate the areas where alae are absent. Right panel: DIC image of a worm showing completely formed alae. **(B)** Left panel: Alae formation in wildtype *wIs78* worms from early L4 to young adult stage. Right panel: Alae analysis in mutant *ztf-16(ok3028)* worms from early L4 to young adult stage. For each strain, the developmental age, i.e. early L4 to young adult, was determined by the extent of gonad migration (1.25, 1.5, 1.75 or 2.0). At each stage for both strains, 25 to 30 worms were scored for the extent of lateral seam cell fusion and for formation of alae. Error bars represent standard deviation.

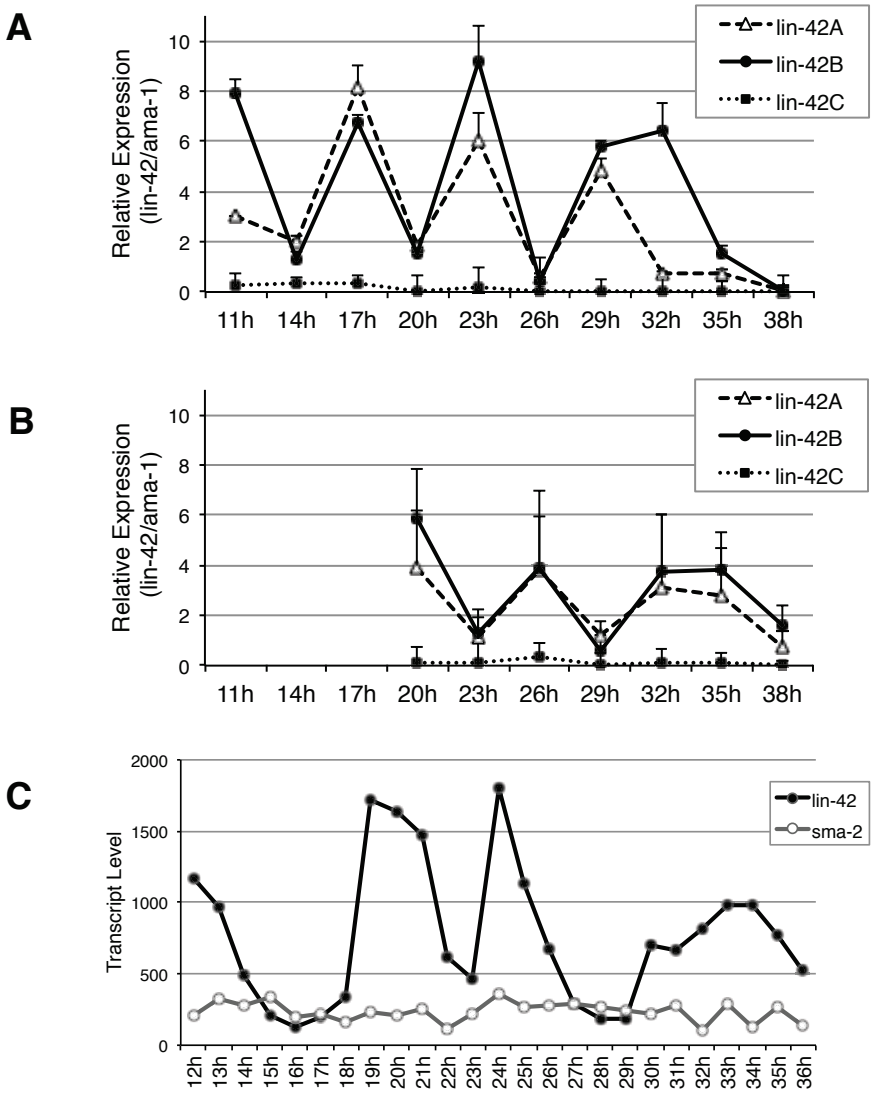


Figure 3.12: *lin-42A, B and C* isoform levels in *sma-2(ok3019)* mutant and *sma-2* expression pattern over larval development.

(A) *lin-42 A,B,C* isoform expression in N2-wildtype worms on mock RNAi.
(B) *lin-42 A,B,C* isoform expression in *sma-2(ok3109)* deletion mutant. Transcript levels were measured by quantitative real-time PCR (qRT-PCR) using *lin-42* isoform specific primers and probes (Appendix A). Transcript values are represented relative to the expression of the internal reference gene *ama-1* in the same sample. The x-axis represents hours post addition of synchronized L1 worms to RNAi bacteria at 25°C. Within a single time-course experiment, qRT-PCR reactions were run in triplicate for each *lin-42* isoform or *ama-1* at each time point. The error bars represent standard deviation.
(C) Microarray expression data for *sma-2* and *lin-42* transcripts in N2 wildtype worms over time-course of larval development. The x-axis represents time in hours (h) post addition of synchronized L1 worms to OP50 bacterial food at 25°C. Microarray data provided by DB (Banerjee Lab unpublished).

Candidate #	Binding Region	TF Family	Gene	Gene #	Gene Product Description	Post Embryonic Expression
1	BR1	HIFF	**aha-1**	C25A1.11	bHLH transcription factor	Hypodermal, intestinal, pharygeal
2	BR1	HIFF	**ahr-1**	C41G7.5	Aryl-hydrocarbon receptor nuclear translocator	Neural
3	BR1	HOMF, NKX6	**alr-1**	R08B4.2	Homeodomain transcription factor	Hypodermal
4	BR1	DHOM, NKX6	**ceh-2**	C27A12.5	Homeodomain transcription factor	Vulval, muscle, pharyngeal
5	BR1	HOMF	**ceh-22**	F29F11.5	Hom of Transcription factor tinman/NKX2-3	Intestine, pharynx, neural(VNC)
6	BR1	DHOM	**ceh-24**	F55B12.1	Homeodomain transcription factor	Vuval, muscle
7	BR1	NKX6	**ceh-6**	K02B12.1	Hom of OCT1 TF, POU and HOX domains	Hypodermal
8	BR1	HIFF	**cky-1**	C15C8.2	bHLH-PAS TF	Pharyngeal
9	BR1	EBOX	**crh-1**	Y41C4A.4	cAMP response element binding protein, CREB homolog	Constitutive
10	BR1	PCBE	**fzr-1**	ZK1307.6	WD40 repeat DNA binding protein	Broad expression
11	BR1	HIFF	**hif-1**	F38A6.3	bHLH TF	Broad expression
12	BR1	DHOM, HOMF	**lin-39**	C07H6.7	hom zerknullt, Dm deformed, sex combs reduced	Ventral epidermal cells, VNC
13	BR1	DHOM, HOMF	**mab-5**	C08C3.3	Homeodomain transcription factor	Hypodermal, Q lineagel
14	BR1	EBOX	**mls-1**	H14A12.4	Tbx1 subfamily of T-box transcription factors	Vulva, uterine, M lineage
15	BR1	NKX6	**pal-1**	C38D4.6	Homeodomain transcription factor, Caudal homolog	Hypodermal, neural, intestine
16	BR1	HOMF	**pax-3**	F27E5.2	Homeodomain transcription factor	Hypodermal, neural
17	BR1	EBOX	**tbx-8**	T07C4.2	TBX2 and related T-box transcription factor	Hypodermal
18	BR1	EBOX	**tbx-9**	T07C4.6	TBX2 and related T-box transcription factor	Hypodermal
19	BR1	HOMF	**vab-15**	R07B1.1	Homeodomain transcription factor	Hypiodermal SC, VNC
20	BR1, BR2	GATA	**egl-18**	F55A8.1	GATA transcription factor	Seam cells, neural
21	BR1, BR2	E2FF, GATA	**elt-6**	F52C12.5	GATA transcription factor	Seam cells, neural
22	BR1, BR2	GATA	**zfp-1**	F54F2.2	Leucine Zipper, Zn finger transcription factor	Seam Cells
23	BR1, BR2	GATA	**ztf-16**	R08E3.4	Leucine Zipper, Zn finger transcription factor	Seam Cells
24	BR1, BR2, BR3	HOMF	**pop-1**	W10C8.2	HMG-box transcription factor, TCF/LEF homolog	Seam cells, vulva, somatic gonad
25	BR1, BR2, BR3	NKX6	**skn-1**	T19E7.2	bZIP transcription factor	Hypodermal, neural, intestine
26	BR1, BR3	NKX6	**ces-2**	ZK909.4	bZIP transcription factor	Hypodermal, neural, muscle
27	BR2	PLZF	**blmp-1**	F25D7.3	Zn finger transcription factor	Seam cells, neural
28	BR2	DHOM	**ceh-32**	W05E10.3	Homeodomain transcription factor	Hypodermal, neural
29	BR2	SP1F	**klf-3**	F54H5.4	Zn finger transcription factor	Hypodermal
30	BR2	HOMF	**nhr-23**	C01H6.5	Nuclear hormone receptor transcription factor	Seam cells
31	BR2, BR3	NKX6	**daf-16**	R13H8.1	Forkhead transcription factor, FOXO homolog	Hypodermal, vulva, intestine, neural
32	BR2, BR3	DHOM, NKX6	**irx-1**	C36F7.1	Homeodomain transcription factor	Hypodermal, neural
33	BR2, BR3	OCT1	**lin-11**	ZC247.3	LIM homeodomain transcription factor	Vulval, muscle, neural
34	BR2, BR3	NKX6	**lin-31**	K10G6.1	Winged helix-like transcription factor	Vulval
35	BR2, BR3	OCT1	**mec-3**	F01D4.6	Homeodomain transcription factor	Neural, hypodermal
36	BR2, BR3	HOMF	**nhr-1**	R09G11.2	Nuclear hormone receptor transcription factor	Hypodermal
37	BR2, BR3	HOMF	**nhr-41**	Y104H12A.1	Nuclear hormone receptor transcription factor	Seam cells
38	BR2, BR3	HOMF	**nhr-6**	C48D5.1	Nuclear hormone receptor transcription factor	Neural, spermatheca
39	BR2, BR3	HOMF	**nhr-64**	C45E1.1	Nuclear hormone receptor transcription factor	Hypodermis, neurons, pharynx, gut
40	BR2, BR3	HOMF	**nhr-69**	T23H4.2	Nuclear hormone receptor transcription factor	Hypodermis, gut, uterus
41	BR2, BR3	OCT1	**ttx-3**	C40H5.5	Homeodomain transcription factor	Hypodermal, neural
42	BR2, BR3	OCT1	**unc-86**	C30A5.7	POU-domain transcription factor	Hypodermal, neural
43	BR2, BR3	OCT1	**zag-1**	F28F9.1	Homeodomain transcription factor	Neural, intestine

Table 3.1: List of 43 candidate *C. elegans* transcription factor genes that were tested in *let-7(lf)* lethality suppression/enhancement RNAi screen.

Candidate TF	Lethality in genetic background			Putative function	Stage-specificity
	let-7(n2853)	VT1066	N2/wildtype		
nhr-23	Suppressor	Suppressor	~50%	Positive regulator	All stages
nhr-25	Suppressor	Suppressor	Not tested	Positive regulator	All stages
ztf-16	Suppressor	No effect	None	Positive regulator	Late larval
blmp-1	No effect	Suppressor	None	Positive regulator	Early-larval
pax-3	Enhancer	Enhancer	None	Negative regulator	All stages
fzr-1	Enhancer	Enhancer	Not tested	Negative regulator	All stages
klf-3	Enhancer	Enhancer	~30%	Negative regulator	All stages
zfp-1	No effect	Enhancer	Not tested	Negative regulator	Early-larval

Table 3.2: Summary of *let-7(lf)* lethality suppression/enhancement RNAi screen results.

CHAPTER 4 : Conclusions & Future Directions

The role and mechanism of function of the oscillating circadian timing gene *period* has been well elucidated at both the molecular and cellular levels. However, recent studies have also shown that the constitutive or non-oscillating expression of *period* is related to regulation of fate choices in stem cells, indicative of a non-circadian function of *period* during development [56, 57, 76]. Conversely, how *period* functions in this capacity, as a developmental timer is not clearly understood. *lin-42* is the *C. elegans* ortholog of *period*. It is a heterochronic gene that regulates various post-embryonic stage-specific events in the worm, including the development of the epidermal stem cells. The work presented here is an effort to characterize the transcriptional regulation of *lin-42*, the ortholog of *period*, in a simpler organism *C. elegans*, as a model to understand how *period* might be functioning in development to regulate stem cell fate choices in higher organisms.

To identify both the *cis*-regulatory DNA elements as well as the *trans*-regulatory protein factors that bind to them to regulate the oscillating expression pattern of *lin-42*, we developed and optimized an Electrophoretic mobility shift assay (EMSA) in *C. elegans* using nuclear extracts (Chapter 2). Using the EMSA, we tested approximately 1700 bp of 5' non-coding sequence of the *lin-42B/C* promoter and defined three regions of specific and distinct binding activity that we designated as Binding regions 1, 2 and 3 (BR1-3). We hypothesized that BRs 1 to 3 represented proximal and/or distal enhancer elements and that *lin-42*-specific transcription factors bind to these DNA regions to control transcription of *lin-42* isoforms. In support of this hypothesis, BRs 1 to 3 lie in non-coding regions that are among the most conserved across *C. elegans* and closely as well as distantly related nematode species, which suggests that these non-

coding sequences have regulatory function. Although, mass spectrometric analysis of the complex that bound to BR2 yielded no significant or relevant results, in the future, incorporating transcription factor enrichment steps (like Oligonucleotide or promoter trapping reviewed in [130, 131]) prior to MS analysis will increase the likelihood of identifying relevant protein/proteins. Analyses of transcriptional reporters with deletions or additions of these binding regions can further test the regulatory significance of these sequences on *lin-42* temporal and spatial expression patterns.

We used the results from Chapter 2 to inform a parallel RNA interference based approach to identify gene-specific transcriptional regulators of *lin-42* (Chapter 3). This approach yielded ZTF-16 as a potential transcriptional activator of *lin-42*. Direct effects on *lin-42* expression patterns were evaluated in a *ztf-16* deletion mutant by isoform-specific qRT-PCR, which revealed that transcription of *lin-42A* and *lin-42B* isoforms was significantly reduced, although not completely abolished. These results led us to conclude that ZTF-16 is necessary but not sufficient for *lin-42* expression. Furthermore, whole genome microarray expression analysis revealed that *ztf-16* transcripts are also expressed in a cyclic fashion (like *lin-42*) with peaks of expression preceding those of *lin-42*. This mRNA profile is consistent with what we would expect of a transcriptional activator of *lin-42*. This evidence leads us to conjecture that ZTF-16 may be acting in conjunction with other protein factor/s to regulate the cyclic temporal expression pattern of *lin-42A* and *lin-42B* isoforms. This hypothesis is supported by the stage-specific EMSA results that indicate reiterative complex formation throughout the four larval stages. Finally, phenotypic analyses of the *ztf-16* deletion mutant revealed precocious heterochronic defects in the seam cell division pattern, seam cell fusion and alae formation, consistent with *lin-42* loss-of-function phenotypes, These data support the hypothesis that ZTF-

16 regulates *lin-42* expression and function. *lin-42* loss-of-function also results in aberrant dauer, lethargus and ecdysis phenotypes. An interesting next step in the analysis of *ztf-16* regulation of *lin-42* would be to test whether similar phenotypes are also observed *ztf-16(lf)* mutants [46, 47].

The two parallel, complementary approaches that we have taken to identify transcriptional regulators of *lin-42* merge to corroborate the data yielded from each. Bioinformatics analysis of Binding Regions 1 and 2 (Chapter 2) has revealed a putative ZTF-16 GATA binding site in each that is perfectly conserved among *C. elegans, C. briggsae* and *C. remanei*. We could determine whether ZTF-16 is one of the proteins that bind to the BR2 GATA site by using mass spectrometry of the EMSA complex that forms with BR1 or BR2 as DNA probe. We could also test for direct binding of ZTF-16 to this site by One Hybrid assays. The regulatory significance of the GATA sites on *lin-42* expression could be tested using transcriptional reporter analysis. By carrying out deletion and sufficiency analysis of BR1 or BR2 in *lin-42* transcriptional reporters, we can assay the effect of loss or ectopic gain of these DNA motifs on transcriptional regulation *in vivo*. We know from microarray expression data that the *ztf-16* message shows oscillating expression during larval development with expression peaks that precede induction of *lin-42* expression. We can confirm this expression pattern using qRT-PCR of *ztf-16* in synchronized larval populations. We can also test whether ZTF-16 protein shows a similar oscillatory pattern using a translational ZTF-16::GFP reporter. Cyclic expression of ZTF-16 protein suggests that cyclic *lin-42* expression pattern may be driven by ZTF-16, and poses the key question – How are *ztf-16* expression and function regulated?

REFERENCES

1. Pasquinelli AE, R.G., *Control of developmental timing by micrornas and their targets*. Annu Rev Cell Dev Biol. , 2002(18): p. 495-513.
2. Eric G, M., *Heterochronic Genes and the Nature of Developmental Time*. Current Biology, 2007. **17**(11): p. R425-R434.
3. Ambros, V., *Control of developmental timing in Caenorhabditis elegans*. Curr Opin Genet Dev. , 2000 **10**(4): p. 428-33.
4. Berezikov, E., C.I. Bargmann, and R.H. Plasterk, *Homologous gene targeting in Caenorhabditis elegans by biolistic transformation*. Nucleic Acids Res, 2004. **32**(4): p. e40.
5. Mello, C. and A. Fire, *DNA transformation*. Methods Cell Biol, 1995. **48**: p. 451-82.
6. Mello, C.C., et al., *Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences*. EMBO J, 1991. **10**(12): p. 3959-70.
7. Praitis, V., et al., *Creation of low-copy integrated transgenic lines in Caenorhabditis elegans*. Genetics, 2001. **157**(3): p. 1217-26.
8. Altun, Z.F.a.H., D.H., *Handbook of C. elegans Anatomy*. *In WormAtlas*. 2008.
9. Sulston, J.E. and H.R. Horvitz, *Post-embryonic cell lineages of the nematode, Caenorhabditis elegans*. Developmental Biology, 1977. **56**(1): p. 110-156.
10. Joshi, P.M., et al., *Caenorhabditis elegans as a model for stem cell biology*. Dev Dyn, 2010. **239**(5): p. 1539-54.
11. Abrahante, J.E., E.A. Miller, and A.E. Rougvie, *Identification of heterochronic mutants in Caenorhabditis elegans. Temporal misexpression of a collagen::green fluorescent protein fusion gene*. Genetics, 1998. **149**(3): p. 1335-51.
12. Ambros, V. and H.R. Horvitz, *Heterochronic mutants of the nematode Caenorhabditis elegans*. Science, 1984. **226**(4673): p. 409-16.
13. Jeon, M., et al., *Similarity of the C. elegans developmental timing protein LIN-42 to circadian rhythm proteins*. Science, 1999. **286**(5442): p. 1141-6.
14. Abbott, A.L., et al., *The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans*. Dev Cell, 2005. **9**(3): p. 403-14.
15. Moss, E.G., R.C. Lee, and V. Ambros, *The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA*. Cell, 1997. **88**(5): p. 637-46.
16. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans*. Nature, 2000. **403**(6772): p. 901-6.
17. Slack, F.J., et al., *The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor*. Mol Cell, 2000. **5**(4): p. 659-69.
18. Liu, Z.C. and V. Ambros, *Heterochronic genes control the stage-specific initiation and expression of the dauer larva developmental program in Caenorhabditis elegans*. Genes Dev, 1989. **3**(12B): p. 2039-49.

19. Liu, Z., S. Kirch, and V. Ambros, *The Caenorhabditis elegans heterochronic gene pathway controls stage-specific transcription of collagen genes*. Development, 1995. **121**(8): p. 2471-8.
20. Euling, S. and V. Ambros, *Heterochronic genes control cell cycle progress and developmental competence of C. elegans vulva precursor cells*. Cell, 1996. **84**(5): p. 667-76.
21. Hallam, S.J. and Y. Jin, *lin-14 regulates the timing of synaptic remodelling in Caenorhabditis elegans*. Nature, 1998. **395**(6697): p. 78-82.
22. Olsson-Carter, K. and F.J. Slack, *A Developmental Timing Switch Promotes Axon Outgrowth Independent of Known Guidance Receptors*. PLoS Genet, 2010. **6**(8): p. e1001054.
23. Resnick, T.D., K.A. McCulloch, and A.E. Rougvie, *miRNAs give worms the time of their lives: Small RNAs and temporal control in Caenorhabditis elegans*. Developmental Dynamics, 2010. **239**(5): p. 1477-1489.
24. Banerjee, D. and F. Slack, *Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression*. Bioessays, 2002. **24**(2): p. 119-29.
25. Moss, E.G., *RNA interference: it's a small RNA world*. Curr Biol, 2001. **11**(19): p. R772-5.
26. Grosshans, H. and F.J. Slack, *Micro-RNAs: small is plentiful*. J Cell Biol, 2002. **156**(1): p. 17-21.
27. Hayes, G.D. and G. Ruvkun, *Misexpression of the Caenorhabditis elegans miRNA let-7 is sufficient to drive developmental programs*. Cold Spring Harb Symp Quant Biol, 2006. **71**: p. 21-7.
28. 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.
29. Seggerson, K., L. Tang, and E.G. Moss, *Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation*. Dev Biol, 2002. **243**(2): p. 215-25.
30. 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.
31. Ambros, V. and H.R. Horvitz, *The lin-14 locus of Caenorhabditis elegans controls the time of expression of specific postembryonic developmental events*. Genes Dev, 1987. **1**(4): p. 398-414.
32. Zhu, C., et al., *The lin-4 Gene Controls Fat Accumulation and Longevity in Caenorhabditis elegans*. Int J Mol Sci, 2010. **11**(12): p. 4814-25.
33. Johnson, R.W., et al., *The Caenorhabditis elegans heterochronic gene lin-14 coordinates temporal progression and maturation in the egg-laying system*. Developmental Dynamics, 2009. **238**(2): p. 394-404.
34. Morita, K. and M. Han, *Multiple mechanisms are involved in regulating the expression of the developmental timing regulator lin-28 in Caenorhabditis elegans*. EMBO J, 2006. **25**(24): p. 5794-804.

35. Pepper, A.S., et al., *The C. elegans heterochronic gene lin-46 affects developmental timing at two larval stages and encodes a relative of the scaffolding protein gephyrin*. Development, 2004. **131**(9): p. 2049-59.
36. Bethke, A., et al., *Nuclear hormone receptor regulation of microRNAs controls developmental progression*. Science, 2009. **324**(5923): p. 95-8.
37. Abrahante, J.E., et al., *The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs*. Dev Cell, 2003. **4**(5): p. 625-37.
38. Antebi, A., J.G. Culotti, and E.M. Hedgecock, *daf-12 regulates developmental age and the dauer alternative in Caenorhabditis elegans*. Development, 1998. **125**(7): p. 1191-205.
39. Roush, S.F. and F.J. Slack, *Transcription of the C. elegans let-7 microRNA is temporally regulated by one of its targets, hbl-1*. Dev Biol, 2009. **334**(2): p. 523-34.
40. 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.
41. Ambros, V., *A hierarchy of regulatory genes controls a larva-to-adult developmental switch in C. elegans*. Cell, 1989. **57**(1): p. 49-57.
42. Bettinger, J.C., K. Lee, and A.E. Rougvie, *Stage-specific accumulation of the terminal differentiation factor LIN-29 during Caenorhabditis elegans development*. Development, 1996. **122**(8): p. 2517-27.
43. Rougvie, A.E. and V. Ambros, *The heterochronic gene lin-29 encodes a zinc finger protein that controls a terminal differentiation event in Caenorhabditis elegans*. Development, 1995. **121**(8): p. 2491-500.
44. Tennessen, J.M., et al., *Novel heterochronic functions of the Caenorhabditis elegans period-related protein LIN-42*. Dev Biol, 2006. **289**(1): p. 30-43.
45. Liu, Z., *Genetic Control of Stage-Specific Developmental Events in C. elegans* 1990, Ph. D. Thesis, Harvard University.
46. Tennessen, J.M., K.J. Opperman, and A.E. Rougvie, *The C. elegans developmental timing protein LIN-42 regulates diapause in response to environmental cues*. Development, 2010. **137**(20): p. 3501-11.
47. Monsalve, G.C., C. Van Buskirk, and A.R. Frand, *LIN-42/PERIOD controls cyclical and developmental progression of C. elegans molts*. Curr Biol, 2011. **21**(24): p. 2033-45.
48. Cassada, R.C. and R.L. Russell, *The dauerlarva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans*. Dev Biol, 1975. **46**(2): p. 326-42.
49. Hu, P.J., *Dauer*. WormBook, 2007: p. 1-19.
50. Antebi, A., et al., *daf-12 encodes a nuclear receptor that regulates the dauer diapause and developmental age in C. elegans*. Genes Dev, 2000. **14**(12): p. 1512-27.
51. Ludewig, A.H., et al., *A novel nuclear receptor/coregulator complex controls C. elegans lipid metabolism, larval development, and aging*. Genes Dev, 2004. **18**(17): p. 2120-33.
52. Motola, D.L., et al., *Identification of ligands for DAF-12 that govern dauer formation and reproduction in C. elegans*. Cell, 2006. **124**(6): p. 1209-23.
53. Karp, X. and V. Ambros, *The developmental timing regulator HBL-1 modulates the dauer formation decision in Caenorhabditis elegans*. Genetics, 2011. **187**(1): p. 345-53.

54. Hayes, G.D., A.R. Frand, and G. Ruvkun, *The mir-84 and let-7 paralogous microRNA genes of Caenorhabditis elegans direct the cessation of molting via the conserved nuclear hormone receptors NHR-23 and NHR-25*. Development, 2006. **133**(23): p. 4631-41.
55. Raizen, D.M., et al., *Lethargus is a Caenorhabditis elegans sleep-like state*. Nature, 2008. **451**(7178): p. 569-572.
56. Borgs, L., et al., *Period 2 regulates neural stem/progenitor cell proliferation in the adult hippocampus*. BMC Neurosci, 2009. **10**: p. 30.
57. Wang, C.Y., et al., *Increased vascular senescence and impaired endothelial progenitor cell function mediated by mutation of circadian gene Per2*. Circulation, 2008. **118**(21): p. 2166-73.
58. Xia, H.C., et al., *Deregulated expression of the Per1 and Per2 in human gliomas*. Can J Neurol Sci, 2010. **37**(3): p. 365-70.
59. Yang, X., et al., *The circadian clock gene Per1 suppresses cancer cell proliferation and tumor growth at specific times of day*. Chronobiol Int, 2009. **26**(7): p. 1323-39.
60. Fu, Y.H., *Oscillating per-cision*. PLoS Biol, 2008. **6**(7): p. e192.
61. Panda, S., J.B. Hogenesch, and S.A. Kay, *Circadian rhythms from flies to human*. Nature, 2002. **417**(6886): p. 329-35.
62. Gallego, M. and D.M. Virshup, *Post-translational modifications regulate the ticking of the circadian clock*. Nat Rev Mol Cell Biol, 2007. **8**(2): p. 139-48.
63. Hardin, P.E., *Essential and expendable features of the circadian timekeeping mechanism*. Curr Opin Neurobiol, 2006. **16**(6): p. 686-92.
64. Kippert, F., D.S. Saunders, and M.L. Blaxter, *Caenorhabditis elegans has a circadian clock*. Curr Biol, 2002. **12**(2): p. R47-9.
65. Saigusa, T., et al., *Circadian behavioural rhythm in Caenorhabditis elegans*. Curr Biol, 2002. **12**(2): p. R46-7.
66. Simonetta, S.H., et al., *Circadian stress tolerance in adult Caenorhabditis elegans*. J Comp Physiol A Neuroethol Sens Neural Behav Physiol, 2008. **194**(9): p. 821-8.
67. van der Linden, A.M., et al., *Genome-wide analysis of light- and temperature-entrained circadian transcripts in Caenorhabditis elegans*. PLoS Biol, 2010. **8**(10): p. e1000503.
68. Simonetta, S.H., et al., *Timing of locomotor activity circadian rhythms in Caenorhabditis elegans*. PLoS One, 2009. **4**(10): p. e7571.
69. Banerjee, D., et al., *Developmental timing in C. elegans is regulated by kin-20 and tim-1, homologs of core circadian clock genes*. Dev Cell, 2005. **8**(2): p. 287-95.
70. Gery, S. and H.P. Koeffler, *The role of circadian regulation in cancer*. Cold Spring Harb Symp Quant Biol, 2007. **72**: p. 459-64.
71. Gery, S., et al., *The circadian gene per1 plays an important role in cell growth and DNA damage control in human cancer cells*. Mol Cell, 2006. **22**(3): p. 375-82.
72. Lee, C.C., *The circadian clock and tumor suppression by mammalian period genes*. Methods Enzymol, 2005. **393**: p. 852-61.
73. Lee, C.C., *Tumor suppression by the mammalian Period genes*. Cancer Causes Control, 2006. **17**(4): p. 525-30.
74. Miyazaki, K., et al., *Tumor growth suppression in vivo by overexpression of the circadian component, PER2*. Genes Cells, 2010. **15**(4): p. 351-8.
75. Chen, S.T., et al., *Deregulated expression of the PER1, PER2 and PER3 genes in breast cancers*. Carcinogenesis, 2005. **26**(7): p. 1241-6.

76. Costa, M.J., et al., *Circadian rhythm gene period 3 is an inhibitor of the adipocyte cell fate*. J Biol Chem, 2011. **286**(11): p. 9063-70.
77. Weaver, D.R., *The suprachiasmatic nucleus: a 25-year retrospective*. J Biol Rhythms, 1998. **13**(2): p. 100-12.
78. Rana, S. and S. Mahmood, *Circadian rhythm and its role in malignancy*. Journal of Circadian Rhythms, 2010. **8**(1): p. 3.
79. Matsuo, T., et al., *Control mechanism of the circadian clock for timing of cell division in vivo*. Science, 2003. **302**(5643): p. 255-9.
80. Riley, T., et al., *Transcriptional control of human p53-regulated genes*. Nat Rev Mol Cell Biol, 2008. **9**(5): p. 402-412.
81. Hunt, T. and P. Sassone-Corsi, *Riding tandem: circadian clocks and the cell cycle*. Cell, 2007. **129**(3): p. 461-4.
82. Fu, L., et al., *The circadian gene Period2 plays an important role in tumor suppression and DNA damage response in vivo*. Cell, 2002. **111**(1): p. 41-50.
83. Sancar, A., et al., *Circadian clock control of the cellular response to DNA damage*. FEBS Lett, 2010. **584**(12): p. 2618-25.
84. Nyberg, K.A., et al., *Toward maintaining the genome: DNA damage and replication checkpoints*. Annu Rev Genet, 2002. **36**: p. 617-56.
85. Sato, F., et al., *PERIOD1 is an anti-apoptotic factor in human pancreatic and hepatic cancer cells*. J Biochem, 2009. **146**(6): p. 833-8.
86. Wood, P.A., et al., *Period 2 mutation accelerates ApcMin/+ tumorigenesis*. Mol Cancer Res, 2008. **6**(11): p. 1786-93.
87. Moriya, T., et al., *Correlative association between circadian expression of mousePer2 gene and the proliferation of the neural stem cells*. Neuroscience, 2007. **146**(2): p. 494-8.
88. Hanyu, R., et al., *Per-1 is a specific clock gene regulated by parathyroid hormone (PTH) signaling in osteoblasts and is functional for the transcriptional events induced by PTH*. J Cell Biochem, 2011. **112**(2): p. 433-8.
89. Fu, L., et al., *The molecular clock mediates leptin-regulated bone formation*. Cell, 2005. **122**(5): p. 803-15.
90. Mendez-Ferrer, S., et al., *Haematopoietic stem cell release is regulated by circadian oscillations*. Nature, 2008. **452**(7186): p. 442-7.
91. Tsinkalovsky, O., et al., *Circadian variations in clock gene expression of human bone marrow CD34+ cells*. J Biol Rhythms, 2007. **22**(2): p. 140-50.
92. Tsinkalovsky, O., et al., *Circadian expression of clock genes in purified hematopoietic stem cells is developmentally regulated in mouse bone marrow*. Exp Hematol, 2006. **34**(9): p. 1249-61.
93. Banerjee, D., et al., *kin-19/casein kinase Ialpha has dual functions in regulating asymmetric division and terminal differentiation in C. elegans epidermal stem cells*. Cell Cycle, 2010. **9**(23): p. 4748-65.
94. Jorgensen, E.M. and S.E. Mango, *The art and design of genetic screens: caenorhabditis elegans*. Nat Rev Genet, 2002. **3**(5): p. 356-69.
95. Sommer, R.J. and A. Ogawa, *Hormone signaling and phenotypic plasticity in nematode development and evolution*. Curr Biol, 2011. **21**(18): p. R758-66.
96. Bell-Pedersen, D., et al., *Circadian rhythms from multiple oscillators: lessons from diverse organisms*. Nat Rev Genet, 2005. **6**(7): p. 544-56.

97. Ronni, T., et al., *Common interaction surfaces of the toll-like receptor 4 cytoplasmic domain stimulate multiple nuclear targets*. Mol Cell Biol, 2003. **23**(7): p. 2543-55.
98. Hahn, S., *Structure and mechanism of the RNA polymerase II transcription machinery*. Nat Struct Mol Biol, 2004. **11**(5): p. 394-403.
99. Roeder, R.G., *Transcriptional regulation and the role of diverse coactivators in animal cells*. FEBS Lett, 2005. **579**(4): p. 909-15.
100. Orphanides, G. and D. Reinberg, *A unified theory of gene expression*. Cell, 2002. **108**(4): p. 439-51.
101. Yang, C., et al., *Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters*. Gene, 2007. **389**(1): p. 52-65.
102. Okkema, P.G.a. and M. Krause, http://www.wormbook.org. , *Transcriptional regulation* in *WormBook, ed*., W. The C. elegans Research Community, Editor 2005.
103. Okkema, P.G. and M. Krause, *Transcriptional regulation*. WormBook, 2005: p. 1-40.
104. Blackwell, T.K. and A.K. Walker, *Transcription mechanisms*. WormBook, 2006: p. 1-16.
105. Russnak, R.H. and E.P. Candido, *Locus encoding a family of small heat shock genes in Caenorhabditis elegans: two genes duplicated to form a 3.8-kilobase inverted repeat*. Mol Cell Biol, 1985. **5**(6): p. 1268-78.
106. MacMorris, M., et al., *Regulation of vitellogenin gene expression in transgenic Caenorhabditis elegans: short sequences required for activation of the vit-2 promoter*. Mol Cell Biol, 1992. **12**(4): p. 1652-62.
107. Stringham, E.G., et al., *Temporal and spatial expression patterns of the small heat shock (hsp16) genes in transgenic Caenorhabditis elegans*. Mol Biol Cell, 1992. **3**(2): p. 221-33.
108. Thellmann, M., J. Hatzold, and B. Conradt, *The Snail-like CES-1 protein of C. elegans can block the expression of the BH3-only cell-death activator gene egl-1 by antagonizing the function of bHLH proteins*. Development, 2003. **130**(17): p. 4057-71.
109. Wagmaister, J.A., et al., *Identification of cis-regulatory elements from the C. elegans Hox gene lin-39 required for embryonic expression and for regulation by the transcription factors LIN-1, LIN-31 and LIN-39*. Dev Biol, 2006. **297**(2): p. 550-65.
110. Wagmaister, J.A., J.E. Gleason, and D.M. Eisenmann, *Transcriptional upregulation of the C. elegans Hox gene lin-39 during vulval cell fate specification*. Mech Dev, 2006. **123**(2): p. 135-50.
111. Okkema, P.G., et al., *Sequence requirements for myosin gene expression and regulation in Caenorhabditis elegans*. Genetics, 1993. **135**(2): p. 385-404.
112. Kostrouchova, M., et al., *CHR3: a Caenorhabditis elegans orphan nuclear hormone receptor required for proper epidermal development and molting*. Development, 1998. **125**(9): p. 1617-26.
113. Leblanc, B. and T. Moss, *DNase I footprinting*. Methods Mol Biol, 2001. **148**: p. 31-8.
114. Hampshire, A.J., et al., *Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands*. Methods, 2007. **42**(2): p. 128-40.
115. Drewett, V., et al., *DNA-bound transcription factor complexes analysed by mass-spectrometry: binding of novel proteins to the human c-fos SRE and related sequences*. Nucleic Acids Res, 2001. **29**(2): p. 479-87.

116. Inouye, C., et al., *Isolation of a cDNA encoding a metal response element binding protein using a novel expression cloning procedure: the one hybrid system*. DNA Cell Biol, 1994. **13**(7): p. 731-42.
117. Fried, M.G., *Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay*. Electrophoresis, 1989. **10**(5-6): p. 366-76.
118. Hellman, L.M. and M.G. Fried, *Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions*. Nat Protoc, 2007. **2**(8): p. 1849-61.
119. Thatcher, J.D., C. Haun, and P.G. Okkema, *The DAF-3 Smad binds DNA and represses gene expression in the Caenorhabditis elegans pharynx*. Development, 1999. **126**(1): p. 97-107.
120. Koh, K., Y. Bernstein, and M.V. Sundaram, *The nT1 translocation separates vulval regulatory elements from the egl-18 and elt-6 GATA factor genes*. Dev Biol, 2004. **267**(1): p. 252-63.
121. Wang, C., et al., *Identification of USF2 as a key regulator of Runx2 expression in mouse pluripotent mesenchymal D1 cells*. Mol Cell Biochem, 2006. **292**(1-2): p. 79-88.
122. Wang, C., et al., *Identification of a novel transcription factor, GAGATA-binding protein, involved in androgen-mediated expression of prostate-specific antigen*. J Biol Chem, 2003. **278**(34): p. 32423-30.
123. Xia, T., et al., *Characterization of the promoter and the transcription factors for the mouse UDP-Gal:betaGlcNAc beta1,3-galactosyltransferase gene*. Gene, 2003. **309**(2): p. 117-23.
124. Dick, D.A., *The distribution of sodium, potassium and chloride in the nucleus and cytoplasm of Bufo bufo oocytes measured by electron microprobe analysis*. J Physiol, 1978. **284**: p. 37-53.
125. Hudson, J.M., M. Crowe, and M.G. Fried, *Effects of anions on the binding of the cAMP receptor protein to the lactose promoter*. Eur J Biochem, 1993. **212**(2): p. 539-48.
126. Wysocki, V.H., et al., *Mass spectrometry of peptides and proteins*. Methods, 2005. **35**(3): p. 211-22.
127. Nalefski, E.A., et al., *Single-Molecule Detection of Transcription Factor Binding to DNA in Real Time:,Äâ Specificity, Equilibrium, and Kinetic Parameters*. Biochemistry, 2006. **45**(46): p. 13794-13806.
128. Narlikar, G.J., H.Y. Fan, and R.E. Kingston, *Cooperation between complexes that regulate chromatin structure and transcription*. Cell, 2002. **108**(4): p. 475-87.
129. Segal, E., et al., *A genomic code for nucleosome positioning*. Nature, 2006. **442**(7104): p. 772-8.
130. Moxley, R.A. and H.W. Jarrett, *Oligonucleotide trapping method for transcription factor purification systematic optimization using electrophoretic mobility shift assay*. J Chromatogr A, 2005. **1070**(1-2): p. 23-34.
131. Jiang, D., R.A. Moxley, and H.W. Jarrett, *Promoter trapping of c-jun promoter-binding transcription factors*. J Chromatogr A, 2006. **1133**(1-2): p. 83-94.
132. Jiang, S., M.R. Galindo, and H.W. Jarrett, *Purification and identification of a transcription factor, USF-2, binding to E-box element in the promoter of human telomerase reverse transcriptase (hTERT)*. Proteomics, 2010. **10**(2): p. 203-11.
133. Smale, S.T. and J.T. Kadonaga, *The RNA polymerase II core promoter*. Annu Rev Biochem, 2003. **72**: p. 449-79.

134. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function*. Cell, 2004. **116**(2): p. 281-97.
135. Lin, S.Y., et al., *The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target*. Dev Cell, 2003. **4**(5): p. 639-50.
136. Lau, N.C., et al., *An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans*. Science, 2001. **294**(5543): p. 858-62.
137. Lim, L.P., et al., *The microRNAs of Caenorhabditis elegans*. Genes Dev, 2003. **17**(8): p. 991-1008.
138. Li, M., et al., *Regulatory mutations of mir-48, a C. elegans let-7 family MicroRNA, cause developmental timing defects*. Dev Cell, 2005. **9**(3): p. 415-22.
139. Reece-Hoyes, J.S., et al., *A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks*. Genome Biol, 2005. **6**(13): p. R110.
140. Chen, N., et al., *WormBase: a comprehensive data resource for Caenorhabditis biology and genomics*. Nucleic Acids Res, 2005. **33**(Database issue): p. D383-9.
141. Rister, J. and C. Desplan, *Deciphering the genome's regulatory code: the many languages of DNA*. Bioessays, 2010. **32**(5): p. 381-4.
142. Kiontke, K. and D.H. Fitch, *The phylogenetic relationships of Caenorhabditis and other rhabditids*. WormBook, 2005: p. 1-11.
143. Large, E.E. and L.D. Mathies, *hunchback and Ikaros-like zinc finger genes control reproductive system development in Caenorhabditis elegans*. Dev Biol, 2010. **339**(1): p. 51-64.
144. *WormBase*. 2008 [cited 2009 May 7]; Available from: www.wormbase.org.
145. *AceView*. Available from: http://www.ncbi.nlm.nih.gov/ieb/research/acembly/.
146. Thierry-Mieg, D. and J. Thierry-Mieg, *AceView: a comprehensive cDNA-supported gene and transcripts annotation*. Genome Biol, 2006. **7 Suppl 1**: p. S12 1-14.
147. Hunt-Newbury, R., et al., *High-throughput in vivo analysis of gene expression in Caenorhabditis elegans*. PLoS Biol, 2007. **5**(9): p. e237.
148. McKay, S.J., et al., *Gene expression profiling of cells, tissues, and developmental stages of the nematode C. elegans*. Cold Spring Harb Symp Quant Biol, 2003. **68**: p. 159-69.
149. Smale, S.T. and K. Dorshkind, *Hematopoiesis flies high with Ikaros*. Nat Immunol, 2006. **7**(4): p. 367-9.
150. McCarty, A.S., et al., *Selective dimerization of a C2H2 zinc finger subfamily*. Mol Cell, 2003. **11**(2): p. 459-70.
151. Gardner, H.F., Volk, M., Jeon, M., Rougvie, A.E., *lin-42 Controls Developmental Timing in C. elegans. International Worm Meeting 729*. 2001.
152. Reinert, K.L., Slack, F.J., *A Mos1 transposon mutagenesis screen for suppressors of the let-7 microRNA. East Coast Worm Meeting 217*. 2004.
153. Procko, C., Y. Lu, and S. Shaham, *Sensory Organ Remodeling in Caenorhabditis elegans Requires the Zinc-finger Protein ZTF-16*. Genetics, 2012.
154. Hardin, P.E., *Molecular genetic analysis of circadian timekeeping in Drosophila*. Adv Genet, 2011. **74**: p. 141-73.
155. Borgs, L., et al., *Period 2 regulates neural stem/progenitor cell proliferation in the adult hippocampus*. BMC Neurosci, 2009. **10**(1): p. 30.
156. Janich, P., et al., *The circadian molecular clock creates epidermal stem cell heterogeneity*. Nature, 2011. **480**(7376): p. 209-214.

157. Costa, M.J., et al., *Circadian rhythm gene period 3 is an inhibitor of the adipocyte cell fate*. Journal of Biological Chemistry, 2011: p. 1-18.
158. Tennessen, J.M., K.J. Opperman, and A.E. Rougvie, *The C. elegans developmental timing protein LIN-42 regulates diapause in response to environmental cues*. Development (Cambridge, England), 2010.
159. Ambros, V., *MicroRNAs and developmental timing*. Curr Opin Genet Dev, 2011. **21**(4): p. 511-7.
160. Melton, C. and R. Blelloch, *MicroRNA Regulation of Embryonic Stem Cell Self-Renewal and Differentiation*. Adv Exp Med Biol, 2010. **695**: p. 105-17.
161. Sulston, J.E. and H.R. Horvitz, *Post-embryonic cell lineages of the nematode Caenorhabditis elegans*. Dev. Biol, 1977. **56**: p. 110-156.
162. Arasu, P., B. Wightman, and G. Ruvkun, *Temporal regulation of lin-14 by the antagonistic action of two other heterochronic genes, lin-4 and lin-28*. Genes & Development, 1991. **5**(10): p. 1825-1833.
163. Seggerson, K., *Two Genetic Circuits Repress the Caenorhabditis elegans Heterochronic Gene lin-28 after Translation Initiation*. Developmental Biology, 2002. **243**(2): p. 215-225.
164. Abbott, A., et al., *The let-7 MicroRNA Family Members mir-48, mir-84, and mir-241 Function Together to Regulate Developmental Timing in Caenorhabditis elegans*. Dev Cell, 2005. **9**(3): p. 403-414.
165. Liu, Z., *Genetic Control of Stage-Specific Developmental Events in C. elegans*. Ph.D. Thesis, Harvard University., 1990.
166. Tennessen, J., et al., *Novel heterochronic functions of the Caenorhabditis elegans period-related protein LIN-42*. Developmental Biology, 2006. **289**(1): p. 30-43.
167. Evans (ed.), T.C., *Transformation and microinjection*, in *WormBook*, T.C.e.R. Community, Editor, WormBook.
168. Vella, M., K. Reinert, and F. Slack, *Architecture of a Validated MicroRNA::Target Interaction*. Chemistry & Biology, 2004. **11**(12): p. 1619-1623.
169. 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 Research, 2008. **68**(20): p. 8535-8540.
170. Kloosterman, W. and R. Plasterk, *The Diverse Functions of MicroRNAs in Animal Development and Disease*. Dev Cell, 2006. **11**(4): p. 441-450.
171. Lytle, J.R., T.A. Yario, and J.A. Steitz, *Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(23): p. 9667-9672.
172. Bartel, D.P., *MicroRNAs: target recognition and regulatory functions*. Cell, 2009. **136**(2): p. 215-33.
173. Mangone, M., et al., *The landscape of C. elegans 3'UTRs*. Science, 2010. **329**(5990): p. 432-5.
174. Lee, R., *An Extensive Class of Small RNAs in Caenorhabditis elegans*. Science, 2001. **294**(5543): p. 862-864.
175. Roush, S. and F. Slack, *The let-7 family of microRNAs*. Trends in Cell Biology, 2008. **18**(10): p. 505-516.

176. Vella, M., *The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR*. Genes & Development, 2004. **18**(2): p. 132-137.
177. Long, D., et al., *Potent effect of target structure on microRNA function*. Nat Struct Mol Biol, 2007. **14**(4): p. 287-294.
178. Honda, S. and H.F. Epstein, *Modulation of muscle gene expression in Caenorhabditis elegans: differential levels of transcripts, mRNAs, and polypeptides for thick filament proteins during nematode development*. Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(3): p. 876-80.
179. Kaufman, E.J. and E.A. Miska, *The microRNAs of Caenorhabditis elegans*. Seminars in Cell & Developmental Biology, 2010. **21**(7): p. 728-737.
180. Bagga, S., et al., *Regulation by and miRNAs Results in Target mRNA Degradation*. Cell, 2005. **122**(4): p. 553-563.
181. Jing, Q., et al., *Involvement of microRNA in AU-rich element-mediated mRNA instability*. Cell, 2005. **120**(5): p. 623-34.
182. Lim, L.P., et al., *Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs*. Nature, 2005. **433**(7027): p. 769-73.
183. Giraldez, A.J., et al., *Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs*. Science, 2006. **312**(5770): p. 75-9.
184. Baek, D., et al., *The impact of microRNAs on protein output*. Nature, 2008. **455**(7209): p. 64-71.
185. Esquela-Kerscher, A., et al., *Post-embryonic expression ofC. elegans microRNAs belonging to thelin-4 andlet-7 families in the hypodermis and the reproductive system*. Developmental dynamics : an official publication of the American Association of Anatomists, 2005. **234**(4): p. 868-877.
186. Hoogewijs, D., et al., *Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in C. elegans*. BMC Mol Biol, 2008. **9**(1): p. 9.
187. 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.
188. McIntosh, B.E., J.B. Hogenesch, and C.A. Bradfield, *Mammalian Per-Arnt-Sim proteins in environmental adaptation*. Annual review of physiology, 2010. **72**: p. 625-45.
189. Lagos-Quintana, M., *Identification of Novel Genes Coding for Small Expressed RNAs*. Science, 2001. **294**(5543): p. 853-858.
190. Nottrott, S., M. Simard, and J. Richter, *Human let-7a miRNA blocks protein production on actively translating polyribosomes*. Nat Struct Mol Biol, 2006. **13**(12): p. 1108-1114.
191. Yekta, S., C.J. Tabin, and D.P. Bartel, *MicroRNAs in the Hox network: an apparent link to posterior prevalence*. Nat Rev Genet, 2008. **9**(10): p. 789-96.
192. Friedman, R.C., et al., *Most mammalian mRNAs are conserved targets of microRNAs*. Genome Res, 2009. **19**(1): p. 92-105.
193. Seitz, H., *Redefining microRNA targets*. Curr Biol, 2009. **19**(10): p. 870-3.
194. Thomas, M., J. Lieberman, and A. Lal, *Desperately seeking microRNA targets*. Nat Struct Mol Biol, 2010. **17**(10): p. 1169-74.
195. Brennecke, J., et al., *Principles of microRNA-target recognition*. PLoS Biol, 2005. **3**(3): p. e85.
196. 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.

197. Grimson, A., et al., *MicroRNA targeting specificity in mammals: determinants beyond seed pairing*. Mol Cell, 2007. **27**(1): p. 91-105.
198. Sun, G., H. Li, and J.J. Rossi, *Sequence context outside the target region influences the effectiveness of miR-223 target sites in the RhoB 3'UTR*. Nucleic Acids Res, 2010. **38**(1): p. 239-52.
199. Abbott, A.L., *Uncovering New Functions for MicroRNAs in Caenorhabditis elegans*. Current biology : CB, 2011. **21**(17): p. R668-R671.
200. Hornstein, E. and N. Shomron, *Canalization of development by microRNAs*. Nat Genet, 2006. **38**(6s): p. S20-S24.
201. Johnson, S.M., et al., *RAS is regulated by the let-7 microRNA family*. Cell, 2005. **120**(5): p. 635-47.
202. Martinez, N.J., et al., *A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity*. Genes & Development, 2008. **22**(18): p. 2535-2549.
203. Yi, R. and E. Fuchs, *MicroRNAs and their roles in mammalian stem cells*. Journal of Cell Science, 2011. **124**(11): p. 1775-1783.
204. Rybak, A., et al., *The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2*. Nat Cell Biol, 2009. **11**(12): p. 1411-20.
205. Newman, M.A. and S.M. Hammond, *Emerging paradigms of regulated microRNA processing*. Genes & Development, 2010. **24**(11): p. 1086-1092.
206. Rybak, A., et al., *A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment*. Nat Cell Biol, 2008: p. 19.
207. Viswanathan, S., G. Daley, and R. Gregory, *Selective Blockade of MicroRNA Processing by Lin28*. Science, 2008. **320**(5872): p. 97-100.
208. Hammell, C.M., X. Karp, and V. Ambros, *A feedback circuit involving let-7-family miRNAs and DAF-12 integrates environmental signals and developmental timing in Caenorhabditis elegans*. Proc Natl Acad Sci USA, 2009. **106**(44): p. 18668-73.
209. Roush, S.F. and F.J. Slack, *Transcription of the C. elegans let-7 microRNA is temporally regulated by one of its targets, hbl-1*. Developmental Biology, 2009: p. 1-12.
210. Nakada, D., B.P. Levi, and S.J. Morrison, *Integrating Physiological Regulation with Stem Cell and Tissue Homeostasis*. Neuron, 2011. **70**(4): p. 703-718.
211. Gimble, J.M., et al., *Circadian rhythms in adipose tissue: an update*. Curr Opin Clin Nutr Metab Care, 2011. **14**(6): p. 554-61.
212. Kadener, S., et al., *A role for microRNAs in the Drosophila circadian clock*. Genes & Development, 2009: p. 1-14.
213. Shende, V.R., et al., *Expression and rhythmic modulation of circulating microRNAs targeting the clock gene Bmal1 in mice*. PLoS One, 2011. **6**(7): p. e22586.
214. Nagel, R., L. Clijsters, and R. Agami, *The miRNA-192/194 cluster regulates thePeriodgene family and the circadian clock*. FEBS Journal, 2009. **276**(19): p. 5447-5455.
215. Tazawa, H., S. Kagawa, and T. Fujiwara, *MicroRNAs as potential target gene in cancer gene therapy of gastrointestinal tumors*. Expert Opin Biol Ther, 2011. **11**(2): p. 145-55.
216. Li, J.Y., et al., *Review: The role of microRNAs in kidney disease*. Nephrology (Carlton), 2010. **15**(6): p. 599-608.
217. McKenna, L.B., et al., *MicroRNAs control intestinal epithelial differentiation, architecture, and barrier function*. Gastroenterology, 2010. **139**(5): p. 1654-64, 1664 e1.

218. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei*. Nucleic Acids Res, 1983. **11**(5): p. 1475-89.

APPENDICES

APPENDIX A

Manuscript- *lin-42/Period* is Post-transcriptionally Regulated by *let-7* MicroRNAs During *C. elegans* Post-embryonic Stem Cell Development

Tracy James[a], Autumn Timpano[a], Lena Chin[b], Frank J. Slack[b] and Diya Banerjee[a]

[a] Dept. of Biological Sciences, Virginia Tech University, Blacksburg, VA 24061
[b] Dept. of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520

Corresponding author:
D. Banerjee: dibaner@vt.edu, Phone: 540-231-0787, Fax: 540-231-3414

Manuscript status:
Submitted to 'Developmental Biology', in review.

Distribution of effort
DB designed the experiments, conducted the preliminary experiments and wrote the manuscript.
TJ and AT conducted the experiments except for the HeLa luciferase assays.
LC under supervision of FJS conducted the HeLa luciferase assays.

ABSTRACT

Period, a core component of the circadian timing pathway, regulates both daily physiology and post-embryonic stem cell development in animals. In adult stem/progenitor cells, Period suppresses proliferation and maintains self-renewal or quiescence. *let-7* microRNAs (miRNAs) promote cellular differentiation and thus function in opposition to *period*. We sought to determine whether *let-7* controls expression of *period* during adult stem cell development.

We report that in *C. elegans*, maturation and terminal differentiation of post-embryonic epidermal stem/progenitor cells is dependent upon down-regulation of *lin-42*/*period* by *let-7* miRNAs. Using *lin-42* 3'UTR reporters and site-directed mutagenesis of putative *let-7* binding sites, we have determined that *let-7* acts non-redundantly with the *let-7* family miRNAs, *mir-48*, *mir-84* and *mir-241*, to down-regulate *lin-42*/*period* transcripts at two developmental time points - at larval stage 2 to enable transition from the early to the late larval developmental program, and at larval stage 4 to enable terminal differentiation and transition from the larval to the adult stage. Loss of function of *let-7* miRNAs leads to elevated *lin-42*/*period* levels that result in delayed epidermal stem/progenitor cell maturation and impaired terminal differentiation. However, we find that while loss of *let-7* miRNAs results in an increase in *lin-42* transcript levels, the cyclic pattern of *lin-42* expression is unaffected. Thus, *lin-42* cyclic expression appears to be controlled via mechanisms that are independent of *let-7*, but *let-7* miRNAs impose an additional layer of negative control that is essential at key developmental transition points.

We further show that human *let-7a* represses expression of all three *period* genes in human cell culture via targeting of *period* 3'UTRs. Our findings suggest that the regulatory relationship between *period* and *let-7* in controlling adult stem cell development is conserved between invertebrates and vertebrates. We hypothesize that *let-7*-mediated repression of *period* is a conserved mechanism for balancing adult stem cell quiescence and self-renewal with differentiation during tissue homeostasis and regeneration.

INTRODUCTION

Period genes are core components of the broadly conserved circadian timing pathway in eukaryotes that coordinates the outputs of physiological and behavioral pathways with diurnal cycles of light and temperature [154]. Period has also recently been found to regulate adult stem cell and progenitor fate in mammals [155-157]. However, in contrast to the detailed molecular characterization of *period* function in the physiological circadian clock, the regulation and function of *period* in developmental pathways is poorly characterized in most animals. A notable exception is *lin-42*, the *C. elegans* ortholog of *period*, which functions in the heterochronic pathway to control developmental transitions and stage-specific expression of cell fates during post-embryonic development [13, 158]. In addition to *lin-42*/*period*, the heterochronic pathway includes orthologs of genes that control mammalian stem cell fate such as *lin-28, lin-41* and *let-7* microRNAs (miRNAs) [159]. A regulatory theme that emerges from both *C. elegans* and vertebrate systems is that *lin-42*/*period* functions to maintain multipotency of stem and progenitor cells, while *let-7* miRNA activity promotes differentiation [155-157, 160]. In this study we have investigated the molecular regulatory relationship between *C. elegans lin-42*/*period* and *let-7* that can serve as a model for understanding the developmental functions of *period* in mammalian stem cell development.

The interactions between *lin-42* and other heterochronic genes in the temporal control of cell fate have been most extensively studied in the context of epidermal stem/progenitor cells called seam cells. During postembryonic development, the seam cells (H0-2, V1-6, T) undergo asymmetric division and self-renewal during each of the four larval stages (L1 to L4) [161]. The V-lineage seam cells also undergo one round of symmetric divisions during the L2 stage that

doubles the V-seam cell number. At the end of larval development, seam cells terminally differentiate by withdrawing from the cell cycle, fusing their cell membranes along the lateral axis to form the seam syncytium and secreting alae, which are raised collagen ridges thought to aid locomotion [161]. Heterochronic genes function in a 'late timer' that regulates seam cell terminal differentiation to occur at the end of the L4 stage, and also function in an 'early timer' that limits symmetric, proliferative V-seam cell divisions to the L2 stage [159]. Mutations in heterochronic genes result in precocious or retarded expression of these seam cell fates, and *lin-42*, which is expressed in a cyclic pattern during all four larval stages, functions in both late and early timers.

In the late timer, *lin-42* partial loss of function mutations result in precocious seam cell differentiation during L4, similar to *lin-41* mutations [13, 17]. The transcription factor gene, *lin-29* is epistatic to both *lin-41* and *lin-42*, and *lin-29* mutation results in retarded seam cell differentiation [43]. Loss of function of *let-7* miRNA also results in retarded seam cell differentiation, but this phenotype is suppressed by loss of function of either *lin-41* or *lin-42* , indicating that *lin-41* and *lin-42* function downstream of *let-7* [13, 17]. These genetic relationships indicate that in the late timer *lin-42* functions in parallel with *lin-41* downstream of *let-7* to maintain larval fates and inhibit adult fates, in part by inhibiting expression of LIN-29 transcription factor.

In the early timer, two parallel pathways converge on *hbl-1*, which functions to maintain L2 cell fates and prevent L3 fates, analogous to the function of *lin-41* at the L4 stage. One of the *hbl-1* regulatory pathways involves *lin-28*, which acts as a positive regulator of *hbl-1* activity by repressing *lin-46* [35]. LIN-28 is active during the L1 and early L2 stages, but as L2 progresses *lin-28* expression falls as a result of repression by increasing levels of *lin-4* miRNA, in a

complex positive regulatory loop involving *lin-14* [15, 162, 163]. Decreasing expression of *lin-28* during the L2 stage leads to derepression of *lin-46* and subsequent inhibition of HBL-1 function, thus allowing transition to the L3 stage. *hbl-1* repression is reinforced by a second *lin-28* independent pathway in which the *let-7* family *(let-7-fam)* miRNAs *mir-48*, *mir-84* and *mir-241* function redundantly to directly inhibit translation of *hbl-1* mRNAs [164]. Mutation of at least two of the *let-7-fam* miRNAs results in a retarded phenotype similar to *lin-46* mutation, due to persistence of *hbl-1* expression into the L3 stage. Expression of the *let-7-fam* miRNAs during the L1 and L2 stages is transcriptionally regulated by DAF-12 in complex with Dafachronic acid (DA) ligand [36]. Consistent with *daf-12* positive regulation of the *let-7-fam* miRNAs, *daf-12* null mutants and mutants that lack the DA ligand binding domain show a retarded heterochronic phenotype in which the L2 proliferative divisions are repeated in the L3 stage [38].

lin-42 partial loss of function does not result in heterochronic phenotypes in the early timer, but several lines of genetic evidence indicate that *lin-42* acts as a positive regulator of L2 stage-specific seam cell division pattern. First, the *lin-42(n1089)* mutation in combination with the *lin-14(n179)* temperature sensitive mutation results in the L2 proliferative division being replaced by the L3 division pattern at the permissive temperature, a phenotype that is not observed in *lin-14(n179)* mutants alone [165]. Second, *lin-42* loss of function suppresses the retarded seam cell division phenotype of *lin-46* mutation, and leads to precocious expression of L3 seam cells fates, albeit at a lower penetrance than the precocious phenotype of *lin-42* loss of function alone [35]. These mutant phenotypes indicate that, while *lin-42* is epistatic to *lin-46*, *lin-46* functions through other factors in addition to *lin-42*. Third, simultaneous loss of function of *lin-42* and *daf-12* results in mutual suppression of both the precocious seam cell differentiation phenotype caused by *lin-42* mutation and the retarded seam cell lineage phenotype caused by

daf-12 mutation [166]. Thus, *lin-42* function is necessary for the retarded expression of L2 seam cell division pattern in the L3 in *daf-12* mutants, but *daf-12* and *lin-42* do not function in a linear pathway to regulate the L2 stage-specific division pattern. Together, the genetic analyses indicate that *lin-42* functions in the early timer downstream of *lin-46* and *daf-12*, and in parallel with *hbl-1*, to positively regulate L2 stage-specific seam cell divisions.

In both early and late timers of the heterochronic pathway, *lin-42* functions downstream of *let-7* miRNAs and in conjunction with the validated *let-7* target genes *hbl-1* and *lin-41* to maintain stage-specific cell fates and repress transition to the next developmental stage. These genetic relationships suggest a model in which *lin-42* might be a direct regulatory target of *let-7* miRNAs. In this model, *let-7* miRNAs coordinately down-regulate *lin-42* and *hbl-1* at the L2 stage to allow transition to L3 stage, and coordinately down-regulate *lin-42* and *lin-41* at L4 stage to allow terminal differentiation. Alternatively, control of *lin-42* expression may be independent of *let-7* miRNAs. In this alternate model, the cyclic expression and down-regulation of *lin-42* are controlled by an unidentified pathway that functions in parallel to and independently of the heterochronic pathway. In this study we tested the two models of *lin-42* regulation by identifying potential *let-7* binding sites in *lin-42* transcripts and assaying for post-transcriptional regulation via these putative miRNA binding sites.

MATERIALS AND METHODS

C. elegans strains

Bristol N2 was used as the wild type strain. Strains JR1000/wIs78 (*ajm-1::gfp/MH27::GFP* and *scm-1::gfp*) and JR667/wIs51 (*unc-119(e2498)*::Tc1;wIs51) were used to visualize seam cell nuclei and seam cell junctions. Additional strains used in this study are as follows: Strains with single gene mutations- *lin-42(n1089)* II MT2257, *lin-42(mg152)* II, *lin-42(ok2385)* II RB1843, *let-7(n2853)* X MT7626, *mir-84(n4037)* X MT13651, *mir-241(n4316)* V MT13897, *lin-58(n4097)* V MT13650. Strains with multiple gene mutations- *lin-58&mir-241(nDf51)* V MT13669, *lin-58&mir241(nDf51)* V; *mir-84(n4037)* X VT1066 and *58&mir241(nDf51)* V; wIs78. All strains were propagated on NGM plates containing *E. coli* OP50 bacterial lawns. Synchronized starved L1 populations were obtained by hatching eggs, isolated from gravid adults, and starving larvae in M9 medium. All strains were cultured at 20°C, unless otherwise noted.

RNAi Analysis

RNAi construct for *lin-42B/C* was obtained from the Ahringher RNAi library. Additional RNAi constructs were created using the pL4440 vector, and were made for *lin-42A/B* and *lin-42C*. RNAi was induced by feeding, starting with L1 stage starved and synchronized animals. RNAi constructs were expressed from *E. coli* HT115(DE3) bacteria, which were grown on NGM media containing 1 mM IPTG to induce expression from the convergent T7 polymerase promoters on the pL4440 vector. All strains feeding on RNAi bacteria were cultured at 20°C unless otherwise noted.

Microscopy

Light and fluorescence microscopy were carried out on Zeiss AxioPlan2 and Zeiss AxioimagerZ1 microscopes and images were captured and analyzed using Zeiss AxioVision software and Hamatzu cameras.

***lacZ* post-transcriptional reporters**

Plasmid B29 (*col-10*p::*lacZ*::*unc-54* 3'UTR) was used as the parent vector from which *unc-54* 3'UTR sequence was removed by Sac II and Nco I digest. *lin-42A/B* 3'UTR (~1.0 kb) and *lin-42* 3'UTR (~1.4 kb) sequences were PCR amplified from N2 genomic DNA and cloned into pB29 in place of *unc-54* 3'UTR, generating pDB421 (*lin-42A/B* 3'UTR) and pDB422 (*lin-42C* 3'UTR) (see Table S1 for primer sequences). Post-transcriptional reporters with mutant LCSs were generated by site-directed mutagenesis of pDB421 using GeneTailor (Invitrogen) (see Table S1 for primer sequences). pDB421sc3LCS was generated by sequential mutagenesis: pDB421 was mutagenized at LCS1 to generate pDB421scLCS1, which was further mutagenized at LCS2 to generate pDB421scLCS1/2, which in turn was mutagenized at LCS3 to finally generate pDB421sc3LCS. The mutagenized 3'UTRs were sequenced to confirm the presence of the introduced LCS mutations and the lack of additional mutations. The mutagenized 3'UTRs were then re-cloned into pB29 lacking *unc-54* 3'UTR in order to avoid mutations introduced into the remainder of the plasmid during site-directed mutagenesis.

To create transgenic *C. elegans*, post-transcriptional reporter plasmids were linearized and injected into N2 animals at 50 ng/μl along with *myo-2p::gfp* at 20 ng/μl as a coinjection marker. Transgenic lines were identified by pharyngeal GFP expression from *myo-2p::gfp* and at

least three independent transgenic lines were identified and propagated for each post-transcriptional reporter. pDB421 line 3 was integrated using UV as previously described [167], and crossed into various *let-7* microRNA loss of function strains. Expression from the post-transcriptional reporter in each transgenic *C. elegans* lines was assayed by b-galactosidase staining as previously described [168].

***luciferase* reporters**

The 3'UTRs from *period 1*, *period 2* and *period 3* were PCR amplified from human genomic DNA isolated from normal, adult lung tissues (BioChain), and cloned into pGL3-Promoter vector (*SV40*p::*luciferase*) (Promega) (see Table S1 for primer sequences). The *period* 3'UTR *luciferase* reporters and antisense RNAs (anti-miRs) were transfected into HeLa cells with lipofectamine 2000 (Invitrogen), and *luciferase* expression assayed using the Dual Luciferase Reporter Assay (Promega) as previously described [169].

Quantitative real time PCR (qRT-PCR)

Synchronized populations of N2/wildtype or VT1066 were grown at 25°C on OP50, and harvested every 3 hours post-plating, starting at 9 hrs and extending to 48 hrs. Synchronized populations of *let-7(n2853)* were grown at 15°C on OP50 and harvested at time points corresponding to late L3 (determined by upward flexing of the gonad arms, ~60 hrs post plating of L1), late L4 (determined by 90% completion of gonad arm migration, ~92 hrs post plating of L1) and early adult (~6 hours post late L4 time point). Total RNA was extracted from each population using Trizol, and cDNA was synthesized from the RNA and used for qRT-PCR. *lin-42A, lin-42B, lin-42C* and *ama-1* specific primers and TaqMan probes (5'-6FAM/3'-TAMARA

labeled, custom synthesized by Applied Biosciences) were used for each qRT-PCR run on Biorad CFX96 RT-PCR detection system (see Table S1 for primer and probe sequences).

RESULTS

lin-42 3'UTRs have putative *let-7* microRNA binding sites (LCSs)

The *C. elegans lin-42* gene consists of 12 exons that are differentially spliced into at least three isoforms, designated *lin-42A*, *lin-42B* and *lin-42 C* (**Figure 1A**) [166]. *lin-42B* is the largest isoform and includes all *lin-42* exons except for exons 6 and 7. Exon 6 is unique to *lin-42C*, which also includes exons 1 to 5, in common with *lin-42B*. Exon 7 is unique to *lin-42A*, which also includes exons 8 to 12, in common with *lin-42B*. In order to determine whether *lin-42* isoforms are post-transcriptionally regulated by *let-7* miRNAs, we sought to identify potential *let-7* binding sites (*let-7* complementary sequences, LCSs) in *lin-42* transcripts.

While some animal miRNAs have been shown to bind at 5' untranslated regions (UTRs) and within introns, the majority of animal miRNAs, and all characterized *C. elegans* miRNAs, bind within the 3'UTRs of messenger RNAs (mRNAs) [170-172]. We thus focused our search for *let-7* binding sites within the *lin-42* 3'UTRs. *lin-42A* and *lin-42B* transcripts have in common a 927 base pair (bp) 3'UTR that is immediately downstream of the exon 12 stop codon. We inferred the length of the 3'UTR from *lin-42A/B* cDNAs and subsequently confirmed the 3'UTR length and sequence by 3'RACE (rapid amplification of cDNA ends) (data not shown). The known *lin-42C* 3'UTR consists of 156 bp of sequence immediately downstream of the exon 6 stop codon, and was inferred from *lin-42C* cDNA and subsequently confirmed by 3'RACE (data not shown). However, our 3'RACE assays also consistently identified two additional, longer

3'UTRs that terminate at approximately 720 bp and 1353 bp downstream of exon 6 (data not shown). *C. elegans* genome-wide analysis of 3'UTRs has revealed that 31% to 43% of genes have alternative 3'UTR isoforms, some of which vary in length due to alternative cleavage sites [173]. *lin-42C* appears to be one such gene with alternative 3'UTR lengths, and we searched for LCSs within 1.4 kilobases (kb) of 3' sequence in order encompass all three *lin-42C* 3'UTRs.

Numerous genomic, proteomic and computational studies of miRNA targets have converged on the consensus that productive regulatory interactions consist of a perfect match between the target mRNA and positions 2 to 8 (the seed sequence) of the mature miRNA [172]. MiRNAs with identical seed sequences are grouped into families and often regulate common targets. In *C. elegans*, the *let-7* miRNA family consists of *mir-84*, *mir-48* and *mir-241* (*let-7-fam*), in addition to the founding *let-7* miRNA, and share the 5'-GAGGUAG-3' seed sequence but differ in the remaining mature miRNA sequence [136, 174, 175]. We searched the *lin-42* 3'UTRs for sequences complementary to the *let-7* family seed sequence but did not any identify sites that would allow contiguous Watson-Crick base pairing between miRNA seed and mRNA, the hallmark of a canonical miRNA biding site. While most conserved miRNA targets in animals have canonical miRNA binding sites, there are examples of validated miRNA targets with atypical miRNA binding sites that do not have a perfect match with the cognate miRNA seed [172]. *C. elegans lin-41* is one such validated target of *let-7* miRNA with two atypical LCSs: in *lin-41* LCS1, base-pairing to the seed sequence is interrupted by an unmatched A nucleotide, while in LCS2, contiguous base-pairing is interrupted by a non-Watson-Crick G:U wobble base pair (**Figure 1B**) [176]. When we relaxed the stringency of the *lin-42* LCS search to include potential atypical *let-7* binding sites, we identified three putative LCSs in 0.9 kb of *lin-42A/B* 3'UTR (**Figure 1B**, A/B-LCSs 1-3) and seven putative LCSs in 1.4 kb of *lin-42C* 3'UTR

(**Figure 1B**, C-LCSs 1-7). We have only shown the match between the miRNA seed and mRNA because base pairing between the 3' miRNA sequences and the mRNA can vary considerably dependent of the non-conserved 3' sequences of miRNAs within a single miRNA family, and even within a single miR:mRNA pair [172]. Moreover, the efficacy of miRNA repression is poorly correlated with the stability of the miRNA:mRNA duplex [172], but strongly correlated with the secondary structures of the miRNA:miR-binding site and surrounding 3'UTR sequences [168, 177].

Transcript expression is down regulated via *lin-42A/B* 3'UTR but not *lin-42C* 3'UTR

To determine whether the *lin-42* LCSs are biologically relevant for expression, we first tested whether *lin-42* expression patterns are post-transcriptionally down regulated via *lin-42* 3'UTRs during *C. elegans* larval development. We used a reporter system consisting of the *col-10* promoter to drive expression of *lacZ* reporter gene in the epidermal seam cells, in which both *lin-42* and *let-7* miRNAs are normally expressed and interact during larval development [159].

We created post-transcriptional reporters for *lin-42* 3'UTRs by attaching1.2 kb of *lin-42A/B* 3'UTR or 1.4 kb of *lin-42C* 3'UTR sequences to the *col-10p::lacZ* transcriptional reporter construct. The 3'UTR of *unc-54*, a gene that is not regulated post-transcriptionally during larval development, was used as a negative control [178]. The post-transcriptional reporter constructs were injected into wildtype *C. elegans,* and maintained as extra-chromosomal arrays in independent transgenic lines.

lin-42 has an oscillating pattern of expression during larval development, with expression peaking during mid-larval stages and down regulation of expression during late-larval stages (**Figure 4A and 4B**) [13, 166]. Genetic lines of evidence suggest that *let-7*, which is expressed

during L3 and L4 stages, may repress *lin-42* expression during the late L4 stage, and the *let-7-fam* miRNAs may repress *lin-42* expression during late L2 stage. We tested this model of late L2 and late L4 repression of *lin-42* expression by *let-7* family miRNAs by assaying *lacZ* reporter expression in synchronized populations during mid and late larval stages for L2, L3 and L4 stages. In three independent transgenic lines carrying the *lin-42A/B* post-transcriptional reporter, we observed a statistically significant decrease in reporter expression from the mid L2 to the late L2 stages, and from the mid L4 to the late L4 stages, but not during the L3 stage (**Figure 2**). Neither the *unc-54*/control reporter lines, nor any of the three independent transgenic lines carrying *lin-42C* post-transcriptional reporter showed similar decreases in expression between mid and late L2 or L4 stages. These results indicate that *lin-42A* and *lin-42B* transcripts are post-transcriptionally down regulated via their 3'UTRs during late L2 and late L4 stages, when *let-7* family miRNAs function. Furthermore, our results indicate that *lin-42C* is not differentially regulated at the post-transcriptional level via its 3'UTR(s) at these times, despite the presence of multiple potential *let-7* binding sites.

For many of the *lin-42* 3'UTR reporter lines, we noticed a decrease in reporter expression between L1 and later stages, which could indicate post-transcriptional regulation by 3'UTR binding proteins or miRNAs, such as *lin-4*, that are expressed during the early larval stages [162, 179]. However, since a similar pattern of L1 to L2 down regulation was observed in the *unc-54* reporter lines, it is more likely that the L1 to L2 down regulation is an experimental artifact or reflects stage-specific transcriptional rather than post-transcriptional regulation. We thus focused further analysis on testing regulation of *lin-42A/B* expression by *let-7* miRNAs during L2 and L4 stages.

Post-transcriptional down regulation via *lin-42A/B* 3'UTR is dependent on *let-7* family miRNAs

To test whether down regulation of *lin-42A/B* 3'UTR reporter expression during L2 and L4 stages was dependent on the presence of *let-7* miRNAs, we assayed reporter expression in genetic backgrounds that lacked *let-7* miRNAs. We crossed *lin-42A/B 3'UTR* reporter line 3 (in N2/ genetic background) with the following *let-7* loss-of-function mutant strains: 1) *let-7(n2853),* 2) *mir-84(n4037),* 3) *lin-58(n4097)* deletion strain for *mir-48*, 4) *mir-241(n4316)* and 5) VT1066 (triple deletion strain for *mir-48*, *mir-241* and *mir-84*). We isolated genetic lines that carried the *lin-42A/B 3'UTR* reporter, as indicated by positive b-galactosidase activity in the seam cells, and were homozygous for the *let-7* miRNA mutation or deletion, as indicated by expression of retarded development and lethality phenotypes associated with *let-7* miRNA loss-of-function) [16, 17, 164]. For each line, we assayed reporter expression in developmentally synchronized larval populations (**Figure 3A**).

For *lin-42A/B 3'UTR* reporters in *let-7(n2853)lf* genetic background, we observed that there was no longer a decrease in reporter activity during the late L4 compared to activity of the same reporter in wildtype genetic background (**Figure 3A**). However, a decrease in reporter expression at the late L2 stage was observed for reporters in both wildtype and *let-7 lf* genetic backgrounds. Conversely, the decrease in reporter expression observed at the late L2 stage in wildtype genetic background was abolished in the absence of any or all three of the *let-7-fam* miRs, but a decrease in reporter expression at the late L4 stage was observed for reporters in both wildtype and *let-7-fam* miRNA loss-of-function genetic backgrounds (**Figure 3A**). These results indicate that the *let-7* and *let-7-fam* miRNAs function non-redundantly to post-transcriptionally down regulate expression of *lin-42A* and *lin-42B* transcripts via their 3'UTRs at specific

developmental stages – *let-7* targets *lin-42* at the late L4 stage, while *mir-48*, *mir-241* and *mir-84* act redundantly to target *lin-42* at the late L2 stage.

Post-transcriptional down regulation via *lin-42A/B* 3'UTR is dependent on LCSs

To determine whether down regulation of *lin-42A/B* 3'UTR reporter expression was mediated by binding of *let-7* miRNAs to LCSs, we mutated the potential *let-7* binding sites in the reporter and assayed reporter expression in wildtype genetic background. We used site-directed mutagenesis to alter the LCS sequence complementary to the *let-7* seed sequence, such that *let-7* miRNAs would no longer bind efficiently to the 'scrambled' LCSs (scLCS1-3) (**Figure 1B**). For example, *lin-42A/B* LCS1 is predicted to pair with 6 out of 7 bases in the *let-7* family seed sequence, with 4 contiguous Watson-Crick base pairs, but scLCS1 can pair with only 3 bases in the seed sequence and does not form any contiguous Watson-Crick base pairs. We created three *lin-42A/B* 3'UTR reporters in which either LCS1 or LCS2 or LCS3 was mutated (scLCS1, scLCS2, scLCS3), and a fourth reporter, in which all three LCSs were mutated (sc3LCS). These post-transcriptional reporters were individually injected into N2/wildtype worms, and at least 3 independent transgenic lines for each reporter were isolated and assayed for reporter expression in developmentally synchronized larval populations.

We found that sc3LCS reporter lines did not show a decrease in reporter expression at either the late L2 stage or the late L4 stage, in contrast to *lin-42A/B* 3'UTR reporter lines with intact LCSs (**Figure 3B**). This result indicates that 3'UTR-mediated expression down regulation of the reporter transcript is dependent on the presence of intact *lin-42A/B* 3'UTR LCSs, and specifically on the part of the LCSs that base pairs with the *let-7* family seed sequence. Results similar to the sc3LCS reporter were found with scLCS1 and scLCS2 reporters. However, loss of

LCS3 did not have any effect on L2 and L4 stage down regulation of reporter expression in wildtype genetic background (**Figure 3B**), and is consistent with LCS3 being the weakest of the three potential *let-7* binding sites in *lin-42A/B* 3'UTR (**Figure 1B**). These results suggest that LCS1 and LCS2 function redundantly as sites for *let-7* miRNA binding, resulting in expression down regulation of the transcript, while LCS3 does not interact functionally with either the *let-7* sister miRNAs at L2 stage or with *let-7* at L4 stage. The results of our experiments in *let-7* loss of function genetic backgrounds and the results of our experiments with mutated LCSs together indicate that *let-7* miRNAs bind directly to *lin-42A/B* 3'UTR LCS1 and LCS2 to mediate post-transcriptional down regulation.

lin-42A and *lin-42B* transcript expression is elevated in the absence of either *let-7* or the L2 *let-7* miRNAs

Based on *lin-4* miRNA:*lin-14* interactions, regulation by miRNAs was initially though to inhibit protein translation but not to affect target mRNA stability [30]. However, a number of studies in *C. elegans* and mammalian systems have shown that message destabilization/degradation is a common feature of regulation by miRNAs [180-184]. The results of our experiments with post-transcriptional reporters indicated that *let-7* miRNAs bind to *lin-42A/B* 3'UTRs to repress reporter protein expression at the late L2 and late L4 stages. To determine whether native *lin-42* expression is similarly regulated by *let-7* miRs, and to test whether miRNA regulation results in mRNA degradation, we measured *lin-42* mRNA levels over the time course of larval development in wildtype and *let-7* loss-of-function mutant strains.

Previous studies have reported a generalized, cyclic expression pattern for *lin-42* that did not distinguish among the *lin-42* isoforms. We used isoform specific primers and probes in

quantitative real-time PCR (qRT-PCR) assays to measure the mRNAs profiles of *lin-42* isoforms A, B and C (**Figure 1A** and **Figure 4**). The qRT-PCR primers for *lin-42A* anneal within exon 7, which is unique to *lin-42A*, and exon 8, which is common to *lin-42A* and *lin-42B*, and the qRT-PCR probe base pairs with a sequence that spans the junction of exons 7 and 8. Thus, this qRT-PCR primers and probe combination can only bind to and amplify from sequences specific to the *lin-42A* transcript. The qRT-PCR primers for *lin-42B* anneal to exons 5 and 8, and the qRT-PCR probe base pairs with a sequence that spans the junction of exons 5 and 8. Thus, this qRT-PCR primers and probe combination can only bind to and amplify from sequences specific to the *lin-42B* transcript, since exon 5 is found in *lin-42C* but not *lin-42A*, and exon 8 is found in *lin-42A* but not *lin-42C*. The qRT-PCR primers and probe for *lin-42C* bind to and amplify sequences from exon 6, which is unique to *lin-42C*. To test the *lin-42* isoform-specific qRT-PCR primer/probe sets we used *lin-42(ok2385)* strain, in which *lin-42* exons exons 7 to 12 are deleted. Thus, *lin-42(ok2385)* cannot express *lin-42A* and *lin-42B* transcripts, but *lin-42C* transcripts should be unaffected. We found that *lin-42A* and *lin-42B* qRT-PCR primers and probe efficiently and specifically amplified mRNAs from N2/wildtype and *lin-42(mg152)* strains, which express the *lin-42A* and *lin-42B* transcripts, but did not amplify mRNAs from *lin-42(ok2385)* (data not shown). In contrast, the *lin-42C* primer/probe set amplifies and detects transcripts from *lin-42(ok2385)*, but does not detect transcripts from *lin-42(n1089)*, in which *lin-42* exons 1 to 6 are deleted [166]. These results indicate that our *lin-42* isoform-specific qRT-PCR primer/probe sets can be used to reliably detect and measure *lin-42A*, *lin-42B* and *lin-42C* isoforms.

let-7 miRNA expression begins during the L3 stage but does not appear to build to functional levels for post-transcriptional down regulation of its target mRNAs until the late L4 stage [16, 185] . Therefore, based on the *let-7* expression pattern and our findings that *lin-42*

post-transcriptional reporters are down regulated during the late L4 stage, we measured *lin-42* transcripts in developmentally synchronized populations of wildtype and *let-7(n2853)* strain animals at late L3, late L4 and early adult stages. For each time point in a particular strain, we measured mRNA levels of *lin-42A*, *B* and *C* isoforms by qRT-PCR. In each sample, we also measured the mRNA levels of *ama-1*, which has been shown to have a constant and steady level of expression during larval development and can thus be used as an internal reference gene [186]. We first calculated the level of expression of each *lin-42* isoform relative to *ama-1* expression in the same sample and then calculated the ratio of *lin-42* isoform expression in the *let-7(n2853)* to N2/wildtype populations. We found that *lin-42A* and *lin-42B* mRNAs were expressed at comparable levels during late L3 stage in wildtype and *let-7(n2853)* populations, but increased 1.8 to 2 fold by late L4 stage in *let-7(n2853)* compared to wildtype, and remained elevated during early adult stage (**Figure 4A**). In contrast, *lin-42C* mRNAs were expressed at comparable levels during L3, L4 and early adult stages in wildtype and *let-7(n2853)* populations. These results are consistent with the findings from the *lin-42* post-transcriptional reporter assays and indicate that regulation by *let-7* miRNA leads to destabilization of *lin-42A* and *lin-42B* mRNA transcripts at the late L4 stage, but does not have any effect on *lin-42C* transcript levels.

Having demonstrated that *let-7* miRNA down regulates *lin-42A* and *lin-42B* mRNA levels during late L4 stage, we predicted that the *let-7-fam* miRNAs *mir-84*, *mi-48* and *mir-241* would similarly regulate *lin-42* expression during the late L2 stage. Thus, we expected that loss of all three *let-7-fam* miRNAs, such as in the VT1066 strain, would result in elevated levels of *lin-42A* and *lin-42B* mRNAs at the late L2 stage. However, it was not known whether elevated levels of *lin-42* would persist into the L3 and L4 stages, and whether loss of repression would affect the oscillating pattern of *lin-42* expression. Expression of the *let-7-fam* miRNAs begins

during the L1 and L2 stages, reaches maximal levels during the late L2 stage and continues at these levels during L3 and L4 stages [164, 185]. In order to encompass all the developmental stages during which *let-7* miRNAs and *lin-42* mRNAs could interact, we assayed mRNA levels of *lin-42* isoforms in developmentally synchronized populations of wildtype and VT1066 strain worms from early L2 to young adult stages. For both wildtype and VT1066 populations, we noted the time at which animals undergoing L1, L2, L3 and L4 molts was first observed (molting times, MTs), thus allowing us to determine the duration and boundaries of each larval stage. We measured mRNA levels in population samples taken at 3-hour intervals, starting at early L2, approximately 1 hour after first observation of the L1 molt, and continuing until young adult stage, approximately 4 hours after first observation of the L4 molt. Using the molting times as guides, we noticed that while the length of a particular larval stage was approximately the same in both wildtype and VT1066 populations, development of the VT1066 population was delayed by approximately 4 to 6 hours relative to the wildtype population. For example, in wildtype the L2 molt occurred at approximately 22 hours and the L3 molt occurred at approximately 32 hours post plating of L1 worms, while in VT1066 the L2 molt occurred at approximately 26 hours and the L3 molt occurred at approximately 37 hours post plating. Therefore, we shifted the VT1066 *lin-42* isoform expression profiles back by 6 hours (minus 2 times points) in order to bring the VT1066 larval stages in register with the wildtype stages. Our results show that *lin-42A* and *lin-42B* mRNA levels during late L2 stage were elevated in VT1066 by approximately 2.0 to 2.5 fold relative to wildtype (**Figure 4B, C**). However, *lin-42C* mRNA levels were approximately equivalent in VT1066 and wildtype (**Figure 4D**). These results indicate that, similar to *let-7* regulation of *lin-42* isoforms, the *let-7-fam* miRNAs function to down regulate *lin-42A* and *lin-42B* mRNAs at the late L2 stage, but do not affect the stability of the *lin-42C* mRNA.

Our temporal expression profiling experiments revealed three additional notable features of *lin-42* isoform expression and regulation. First, both *lin-42A* and *lin-42B* isoform mRNAs in wildtype and VT1066 worms had a cyclic expression pattern similar to the previously reported generalized *lin-42* expression pattern [13, 166], but expression of *lin-42C* was not cyclic in either wildtype or VT1066 worms. Second, elevated levels of *lin-42A* or *lin-42B* during late L2 in VT1066 populations did not disrupt cyclic expression in the subsequent L3 and L4 stages. However, induction of *lin-42A* and *lin-42B* expression in the L3 stage of VT1066 worms was delayed relative to wildtype worms, with the peak of expression occurring during late L3 rather than mid L3. Third, elevated levels of *lin-42A* and *lin-42B* during the late L2 in VT1066 populations did not persist into the L3 and L4 stages. Instead, the amplitude of *lin-42A* and *lin-42B* transcript expression in VT1066 worms returned to wildtype levels for both L3 and L4 stages. These findings indicate that *mir-48*, *mir-84* and *mir-241* redundant regulation of *lin-42A* and *lin-42B* mRNAs is restricted to the late L2 stage, despite continued expression of the miRNAs during subsequent larval stages. Moreover, the continued cyclic expression of *lin-42A/B* in VT1066 worms indicates that, while the *let-7* miRNAs dampen the amplitude of *lin-42A/B* transcript expression, they are not the primary factors responsible for *lin-42* transcript down regulation at L2 and L4 stages. The cyclic expression pattern of *lin-42* is thus independent of *let-7* regulation.

Elevated expression of *lin-42* results in retarded seam cell development

The results of our experiments indicate that *let-7* family miRNAs partially down regulate *lin-42A* and *lin-42B* transcript abundance at late L2 and late L4 stages. What is the functional relevance of this repression? Previous studies have shown that loss of function of *let-7* results in

retarded seam cell terminal differentiation and adult stage lethality, while loss of function of *lin-42* results in precocious seam cell terminal differentiation during the L4 stage, and suppresses *let-7* loss of function phenotypes [13, 16, 166]. We found that *lin-42* has a similar genetic relationship with the *let-7-fam* miRNAs. Similar to *let-7* mutation, loss of function of at least two (nDf51 strain) or all three (VT1066 strain) of the *let-7-fam* miRNAs results in adult stage lethality (**Figure 5A**) and retarded seam cell terminal differentiation, indicated by delayed formation of adult-specific alae (**Figure 5B**) [164]. *lin-42* loss of function in *let-7-fam* mutant genetic background suppresses both adult lethality and delayed alae formation (**Figure 5A and B**), indicating that excess *lin-42* results in retarded seam cell terminal differentiation.

lin-42 loss of function also suppresses the retarded seam cell division pattern in the L3 stage caused by loss of function of *let-7-fam* miRNAs (**Figure 5C**). We tracked seam cell division patterns from L2 to early adult stage using a *C. elegans* strain (wIs78) in which seam cell nuclei are fluorescently labeled with GFP. During normal development, 13 H- and V-lineage seam cells divide once during mid L3 and mid L4 stages giving rise to 26 daughter cells, half of which retain seam cell identity and the nuclear GFP marker and the other half of which differentiate into Hyp7 epidermal cells and lose the nuclear GFP marker. Consistent with the normal seam cell division pattern, we observed an oscillating pattern of GFP-marked seam cell nuclei numbers during L3 and L4 stages, with nadirs of approximately 16 seam cells nuclei (13 dividing seam cells and 3 non-dividing seam cells) and apogees of 26 to 28 seam cell nuclei (**Figure 5C**). In *let-7-fam* loss of function genetic background (nDf51, *mir-48* and *mir-241* deletion), seam cells undergo supernumerary divisions during L3 stage giving rise to an average of 31 and as many as 34 daughter cells, instead of an observed wildtype average of 24 daughter cells. The supernumerary L3 stage seam cell divisions have been interpreted as a retarded

heterochronic phenotype in which the two rounds of seam cell division that are characteristic of the L2 stage are partially reiterated in the L3 stage, due to lack of suppression of *hbl-1* by *let-7-fam* miRNAs at the end of the L2 stage [164]. The L3 supernumerary seam cell divisions caused by mutation of *let-7-fam* miRNAs are completely suppressed and rescued to wildtype division pattern by RNAi knock-down of *lin-42* (**Figure 5C**), indicating that excess *lin-42* results in retarded progression from the L2 to the L3 seam cell developmental program.

In combination with our findings that *lin-42A* and *lin-42B* mRNA levels are elevated in *let-7* loss of function mutant, the results of the RNAi experiments suggest that elevated *lin-42* expression is in large part responsible for the *let-7* family loss of function mutant phenotypes, namely retarded developmental progression to L3 stage and retarded terminal differentiation. Therefore, the results of our experiments together indicate that normal regulation of seam stem cell maturation and terminal differentiation during L3 and L4 stages are dependent upon correct levels of *lin-42* expression, which are controlled redundantly by *let-7* sister miRNAs at the late L2 stage and by *let-7* at the late L4 stage.

***let-7a* represses human *period 1*, *period 2* and *period 3* expression in HeLa cells**

The sequence, developmental expression patterns and regulatory targets of *let-7* miRNAs are well conserved across metazoan species [187]. *Period* genes and their function in circadian control of physiology are also widely conserved [188]. Moreover, both *let-7* and *period* regulate stem cell fate, but appear to function in opposition: *period* loss of function results in inappropriate differentiation of stem and progenitor cells [155-157], while *let-7* loss of function correlates with maintenance of stem cell identity and enhances reprogramming of somatic cells into induced pluripotent stem cells [160]. We thus hypothesized that down regulation of *lin-42/period* by *let-7*, which is necessary for correct development and differentiation of the *C.*

elegans seam stem cells, may be a relationship that is conserved in mammalian cells. We therefore investigated whether *let-7* regulates *period* expression in human cell culture.

In *Homo sapiens* (Hs), there are three independent *period* genes (Hs *Period1*, Hs *Period 2,* Hs *Period 3*). None of the Hs *Period* 3'UTRs contain canonical *let-7* binding sites, but we identified numerous atypical LCSs by applying the same relaxed stringency criteria used to identify *lin-42* LCSs (**Figure 6A**). These LCSs could be binding sites for any of the eleven human *let-7* family members (Hs *let-7a* – Hs *let-7k*), which all have seed sequences that are identical to *C. elegans let-7*. To test for regulatory interaction between Hs *let-7* and Hs *Period*, we used a post-transcriptional reporter assay consisting of *Period* 3'UTR attached to a *SV40*p::*luciferase* transcriptional reporter. Expression from *Period* 3'UTR reporters was assayed in HeLa cells. *let-7a* is natively expressed in HeLa cells [189] and can function like *C. elegans let-7* to down-regulate heterologous transcripts with *C. elegans lin-41* 3'UTRs [190]. Thus, if *let-7a* mediates down-regulation by binding to atypical LCSs in Hs *Period* 3'UTRs, we expected to observe repression of expression from the *Period* 3'UTR reporters under normal cell culture conditions. However, if *let-7a* is depleted by addition of antisense RNA (anti-*let7a*), we predicted that we would observe de-repression of 3'UTR reporter expression. Indeed, for 3'UTR reporters of all three *period* genes, we observed a 1.6 fold to 2.5 fold increase in Luciferase expression on transfection with anti-*let7a* compared to transfection with a control antisense RNA (**Figure 6B**). These results indicate that Hs *let-7* is capable of repressing expression via the 3'UTRs of Hs *Period1*, Hs *Period2* or Hs *Period3*, although these assays do not distinguish between direct and indirect regulation. In combination with the reciprocal phenotypes of *Period* loss of function and *let-7* loss of function on stem cell fate determination, our results from *C. elegans* and human cell culture suggest that direct down regulation of *lin-42*/*period* expression

by *let-7* miRNAs may be a critical and conserved step in enabling the transition from pluri- or multipotency to differentiation during post-embryonic development.

DISCUSSION

lin-42/period and *let-7* miRNAs function in the *C. elegans* heterochronic pathway to control cell fate determination during post-embryonic development of epidermal stem/progenitor cells. We report that two *lin-42* isoforms, *lin-42A* and *lin-42B*, are direct targets of *let-7* miRNAs, which down-regulate *lin-42* transcripts via atypical *let-7* binding sites in the *lin-42* 3'UTRs. Our experiments show that retarded heterochronic phenotypes that result from *let-7* loss of function are largely due to excess *lin-42* expression, indicating that negative regulation of *lin-42* by *let-7* is necessary for normal developmental progression and terminal differentiation. We further show that in human cell culture *let-7a* represses expression via *period* gene 3'UTRs, suggesting that the regulatory relationship between *lin-42/period* and *let-7* may be conserved between worms and humans.

let-7 mediates biologically relevant repression of *lin-42* via binding to marginal LCSs that are not evolutionarily conserved

Animal miRNAs, such as *let-7*, bind to the majority of their targets with perfect Watson-Crick base-pairing between the miRNA seed region and mRNA. The mRNA sequences complementary to the miRNA seed sequence constitute canonical miRNA binding sites, which are generally conserved across related animal species [172]. *C. elegans lin-42A/B* LCSs 1 and 2 are examples of non-conserved and non-canonical miRNA binding sites that are nevertheless

used by *let-7* miRNAs to effect biologically relevant repression. In addition to *lin-42*, there are two examples of genes with non-canonical, atypical miRNA binding sites that have been validated by mutational analysis: *C. elegans lin-41*, which is regulated by *let-7*, and mouse *HoxB8*, which is regulated by miR-196 [168, 176, 191]. For both *lin-41* and *HoxB8,* imperfect complementarity between the 3'UTR and miRNA seed sequence is compensated by base-pairing with the 3' region of the miRNA. In *lin-41* LCSs, 3' compensatory base-pairing with *let-7* extends for ten contiguous bases and is necessary for *let-7*-mediated repression [168]. For *lin-42*, 3' compensatory base-pairing is not as extensive that of *lin-41* LCSs and both *lin-42* LCSs 1 and 2 are predicted to form a minimum of three and a maximum of eight contiguous base pairs with 3' sequences of *let-7* or *let-7* family miRs (**Supplementary Figure S1).** However, *let-7:lin-42* LCS interactions result in biologically relevant repression, indicating that undefined structural features of the *let-7:lin-42* duplex allow recruitment and function of the RISC silencing complex that prevents translation of miRNA-targeted mRNAs.

Genome-wide expression studies in animals indicate that canonical miRNA binding sites enable robust repression, while weaker, atypical sites mediate very limited levels of down-regulation [172, 192]. Consistent with these findings, the weakness of the structural interaction between *lin-42A/B* LCSs and *let-7* miRNAs correlates with the relatively modest, two-fold repression of *lin-42* by *let-7* miRs. Similarly, *lin-41* is modestly down-regulated via *let-7* binding compared to the relatively stronger, but still atypical, *lin-41* LCSs. Two-fold or lower changes in gene expression are well within the natural variations tolerated by most genetic networks and do not result in discernable phenotypic consequences for the majority of genes. However, characterization of validated miRNA targets in vertebrates and invertebrates has revealed that function of these genes is very sensitive to gene dosage [193, 194]. This is true of *lin-42*

regulation by *let-7*; modest and transient elevation of *lin-42A/B* transcripts at late L2 and late L4 stages results in abnormal seam cell divisions in the L3 stage, and delay of terminal differentiation at adult stage, indicating that seam cell fate determination pathways are sensitive to small changes in *lin-42* expression. Thus, although *let-7* miRNA binding to marginal *lin-42A/B* LCSs results in only two-fold down-regulation, this small change in *lin-42* expression is significant within the context of seam stem cell pathways and results in abnormal development.

lin-41 atypical LCS sequences are fully conserved between *C. elegans* and *C. briggsae* [176], but the *lin-42A/B* LCSs are not conserved. While *lin-42* exon sequences are strongly conserved among Caenorhabditis species, alignment of 3'UTRs reveals that there is very little sequence conservation, and what little there is does not encompass *lin-42A/B* nor *lin-42C* LCSs (data not shown). The lack of sequence conservation of *lin-42A/B* LCSs is in accord with studies that have found that non-canonical miRNA sites with 3'-compensatory base-pairing are only rarely under selective pressure to be conserved [192, 195, 196] and constitute only 1% of preferentially conserved sites in animals [172]. However, we identified three atypical LCSs in *C. briggsae* and *C. remanei lin-42A/B* 3'UTRs that are at positions roughly equivalent to those of *C. elegans lin-42A/B* LCSs 1 and 2 (**Supplementary Figure S2**). The presence of these atypical LCSs suggests that *lin-42* down-regulation by *let-7* may be functionally conserved in *C. briggsae* and *C. ramenai* even though LCS sequences are not. Our data indicate that *lin-42A/B* LCSs, which are not conserved at the sequence level, and which do not base-pair extensively with the 3' regions of *let-7* miRNAs, constitute the weakest known atypical miRNA targeting sites that are still functional. These *lin-42A/B* LCSs would not have been identified as potential miRNA regulatory sites using criteria that emphasize perfect base-pairing within the miRNA seed and sequence conservation of the miRNA binding site. This study thus underscores the importance of

using genetic data and mutational analysis of binding sites to identify biologically relevant miRNA:mRNA interactions. Furthermore, identification of the marginal but functional *lin-42A/B* LCSs presents the opportunity to investigate the contribution of factors, such as variation in 3' compensatory biding, 3'UTR context and secondary structure of the target mRNA, which stabilize a weak miRNA:mRNA interaction and can lead to target repression [177, 197, 198].

Low stringency of binding to *lin-42* LCSs may allow regulatory flexibility

A single miRNA target with canonical miRNA binding sites can be regulated by different miRNAs that have the same seed sequence, that is different members of a single miRNA family [172, 179]. However, for genes with atypical miRNA binding sites, such as *lin-41* and *lin-42*, regulation by a miRNA family can be more varied and complex. For example, *let-7* binds and represses *lin-41* during the late L4 stage, but the *let-7-fam* miRNAs do not repress *lin-41* during L2 stage. This difference in regulation likely arises because base-pairing of *lin-41* LCSs with the *let-7* 3' region compensates for imperfect base-pairing in the 5' seed region and leads to *lin-41* repression, while base-pairing with the 3' regions of *mir-48*, *mir-84* and *mir-241* cannot compensate and does not result in repression [195, 196]. Compensatory 3' base-pairing with *let-7*, but not with *let-7-fam* miRNAs, is necessary to restrict *lin-41* down-regulation to the L4 stage and prevent precocious terminal differentiation at L2 stage. Thus, it is not surprising that there has been strong selective pressure against sequence variation of the *lin-41* LCSs, which are 100% conserved between *C. elegans* and *C. briggsae* [176]. Unlike *lin-41*, *lin-42A/B* is repressed by *let-7-fam* miRNAs at L2 stage and non-redundantly by *let-7* at L4 stage. Loss of *lin-42* repression at either stage results in distinct retarded developmental phenotypes, indicating that non-redundant regulation of *lin-42* by *let-7* miRNAs at both developmental time-points is

important for normal seam stem cell development. Therefore, all four members of the *C. elegans let-7* miRNA family must be able to bind to *lin-42* LCSs to mediate repression. We speculate that the marginal nature of the *C. elegans lin-42A/B* LCSs and LCS sequence variability among related nematodes may have arisen from the need to maintain the capability of being regulated by multiple *let-7* family members with redundant and non-redundant functions.

let-7 miRNAs act as rheostats to fine-tune *lin-42* expression during larval development

Based primarily on studies of *lin-4:lin-14* and *let-7:lin-41* interactions in *C. elegans*, miRNAs were originally thought to function as binary on/off switches of gene expression. Subsequent studies in *C. elegans* and other animals have shown that the majority of miRNAs cause more subtle changes in gene expression and function as rheostats to fine-tune gene expression rather than as switches. Another surprising finding has been that for the majority of animal miRNAs, loss of function of a single miRNA results in mutant phenotypes only under environmental or physiological stress [172, 179, 199]. These aspects of regulation by miRNAs have led to the hypothesis that most miRNAs function to buffer developmental and physiological pathways against environmental perturbations and thus ensure the robustness of these pathways [200]. *let-7* has stood as an exception to the rheostat/buffer model of miRNA function and acts as a classic switch to repress expression of its validated targets, such as *C. elegans lin-41* and *hbl-1* and human *Ras* [16, 17, 164, 201]. However, we have found that *let-7* miRNAs also act as rheostats to modulate *lin-42* expression during *C. elegans* larval development. Loss of function of *let-7-fam* miRNAs results in a transient increase in *lin-42* expression during late L2, but does not interrupt the cyclic pattern of *lin-42* expression in subsequent larval stages. Similarly, *let-7* loss of function results in a transient increase in *lin-42* expression during late L4 stage. Thus, *let-*

7 miRs are not the primary negative regulators of *lin-42* expression, as they are of *lin-41* or *hbl-1* expression, but instead provide an additional layer of repression that dampens protein expression to a more optimal level that is still functional in the cell.

let-7 function as a binary switch of *lin-41* expression versus *let-7* function as a rheostat of *lin-42* expression may explain why *lin-41* LCSs are conserved while *lin-42* LCSs are not. *let-7* is the primary negative regulator of *lin-41* expression, and *lin-41* repression is essential for terminal differentiation [16, 17]. Thus, there is likely to be strong evolutionary selective pressure against losing or even weakening *lin-41* repression, resulting in conservation of *lin-41* LCS sequences. On the other hand, *let-7* repression of *lin-42* is a secondary layer of regulation that is added to an independent mechanism that controls the primary pattern of *lin-42* cyclic expression. Thus, one would predict lower selective pressure to maintain strong *let-7*-mediated repression of *lin-42*, allowing for greater sequence variation of *let-7* binding sites in *lin-42* 3'UTRs.

***lin-42* may be a negative regulator of *let-7* microRNAs**

If cyclic expression of *lin-42* is regulated independently of *let-7*, then what is the purpose of *let-7* repression? One possibility is that *let-7* and *lin-42* function in a double-negative feedback loop: *lin-42* directly or indirectly inhibits *let-7*, and thus *let-7* reinforces its own expression and/or function by down-regulating *lin-42*. Double-negative feedback loops appear to be a common feature of miRNA-target interactions, and yeast-one-hybrid assays of interactions between miRNA gene promoters and miRNA target transcription factors have identified at least 23 potential double negative feedback loops in *C. elegans* [202]. The mechanism of a double-negative feedback loop is used to enable transitions in cell fate programs and stabilize a new cellular state [179]. In both vertebrates and invertebrates, *let-7* is necessary for stem cell

differentiation [159, 203]), and is involved in a number of double-negative feedback loops that ensure stable and robust *let-7* expression and function. For example, mouse *lin-41* is expressed in stem cells and encodes an ubiquitin ligase that inhibits the Ago2 component of the miRISC silencing complex, which is necessary for miRNA function [204]. In mice, *lin-28*, another conserved regulatory target of *let-7*, functions to maintain stem cell pluripotency by inhibiting processing of mature *let-7* microRNAs [205-207]. In *C. elegans*, the transition from the early (L1/L2) to the late (L3/L4) seam stem cell developmental program is dependent upon down-regulation of *hbl-1* by *let-7-fam* miRNAs that relieves *hbl-1* repression of *let-7* [205]. Also in *C. elegans*, a complex double-negative feedback loop involving *daf-12* and the *let-7-fam* microRNAs controls the choice between reproductive development, which is favored under normal environmental conditions, and dauer development, which is favored under conditions of environmental stress [208].

In the *C. elegans* heterochronic pathway, *lin-42* functions with *daf-12*, *lin-28* and *hbl-1*to regulate the transition from early to late seam stem cell developmental programs, and with *lin-41* to regulate terminal differentiation at the L4 to adult transition. *lin-42* loss of function phenotypes are similar to the precocious development phenotypes produced by *hbl-1* or *lin-41* mutation. In addition, *lin-42* loss of function almost completely suppresses the retarded heterochronic phenotypes of *let-7* mutation, similar to suppression by *hbl-1* or *lin-41* loss of function. However, genetic analyses indicate that *lin-42* is not epistatic to either *hbl-1* or *lin-41*. These observations suggest that LIN-42 interacts directly with or shares common regulatory targets with HBL-1 and LIN-41. Given that HBL-1 and LIN-41 are negative regulators of *let-7*, we hypothesize that LIN-42 also inhibits *let-7* expression or function.

Period family proteins such as LIN-42 do not have RNA or DNA binding domains but can interact via PAS-domains with other proteins, such as transcription factors, to modify their activities [188]. Thus, LIN-42 could potentially interact with *let-7* transcriptional regulators such as HBL-1, with *let-7*-processing regulators such as LIN-28, with regulators of the miRISC silencing complex such as LIN-41, or directly with components of the miRISC complex. We propose a model in which *lin-42* functions reiteratively during all four larval stages with stage-specific factors, such as LIN-28, HBL-1 and LIN-41, to repress *let-7* microRNAs. This multilayered repression of *let-7* function prevents differentiation and allows for self-renewal of the seam stem cells during larval stages. During early larval development, *let-7* is transcriptionally repressed by HBL-1 [209], and processing of mature *let-7-fam* miRNAs is likely inhibited by *C. elegans* LIN-28, similar to the role of mammalian LIN-28 [205-207]. *C. elegans lin-28* is progressively down-regulated during L2 stage to allow expression of mature *let-7-fam* miRNAs, which in turn repress *hbl-1* and thus allow transcription of *let-7* during L3 and L4 stages. We envision that *let-7-fam* miRNAs are also inhibited by LIN-42, but increasing levels of mature *let-7-fam* miRNAs during late L2 repress *lin-42* and thus reinforce their own activities. Levels of mature *let-7* build progressively during L3 and L4 stages, but *let-7* silencing function is likely inhibited by LIN-41, and we propose also by LIN-42 [204]. As during late L2 stage, *let-7* increases above a threshold level of repression during late L4 stage and down-regulates both *lin-41* and *lin-42* to positively reinforce its own activity, thus allowing terminal differentiation.

let-7 repression of *Period* may be a conserved feature of animal adult stem cell development

The circadian timing pathway, which controls daily patterns of physiology and behavior, has also recently been found to regulate stem cell biology in vertebrates [210, 211]. In particular, Period function is necessary for maintenance of stem cell multipotency in a number of adult stem and progenitor cell lineages. For example, murine Per3 inhibits differentiation of mesenchymal stem cells into adipocytes [157], while Per2 inhibits proliferation and maturation of neural stem/progenitor cells [155] and endothelial progenitor cells [57]. Murine Per1 and Per2 functions are also necessary to maintain epidermal stem cell dormancy and inhibit differentiation [156]. In contrast to the role of Period in stem cell lineages, *let-7* microRNAs promote differentiation and suppress self-renewal in both embryonic stem cells and adult stem/progenitor cells of various tissues including lung, breast, pancreas, as well as neural tissues [156, 159]. The opposing roles of *let-7* and *period* in mammalian adult stem cells are similar to those of *C. elegans let-7* and *lin-42/period* in controlling development of the epidermal seam stem cells. Our work shows that *C. elegans let-7* miRNAs allow terminal differentiation and developmental progression of seam stem cells by direct post-transcriptional repression of *lin-42* transcripts. We further show that *let-7* can post-transcriptionally repress *period* expression in human cells, which suggests that the regulatory relationship between *let-7* and *period* is evolutionarily conserved. We thus hypothesize that differentiation of mammalian mesodermal and ectodermal adult stem/progenitor cells is at least partially dependent upon down-regulation of *period* by *let-7* microRNAs.

MiRNA regulation of *lin-42/period* is not restricted to cell fate pathways in *C. elegans*. MiRNAs post-transcriptionally regulate core genes of the circadian clock, including *clock*, *bmal* and *period*, and influence circadian outputs in both invertebrates and mammals [212, 213]. The homologous miRNAs miR-192, miR-194 and miR-215 directly target transcripts of all three

period genes in human cell culture, and ectopic expression of miR-192/194/215 alters the circadian rhythm of gene expression in these cells [214]. The miR-192/194/215 family also regulates adult stem/progenitor cell development. Similar to the role of *let-7* miRNAs, miR-192/194/215 miRNA expression and function in mice and humans is positively correlated with normally differentiated adult tissues, specifically gastrointestinal and kidney cells, and negatively correlated with cancer stem cells and dedifferentiated tumors of these tissues [215, 216]. These data, together with our work, suggest that both *let-7* and miR-192/194/215 miRNAs promote differentiation by down-regulating *period* genes, which function to maintain multipotency of adult stem/progenitor lineages. miR-192/194/215 are the most highly expressed miRNAs in differentiated intestinal epithelia [217], while *let-7* miRNAs are more broadly expressed [175]. Thus, miR-192/194/215 miRNAs may act with *let-7* to co-regulate *period* in tissues where both miRNAs are expressed, and may be the primary miRNA regulator of *period* in the intestine. There may be additional tissue-specific miRNAs that down-regulate *period* in other tissues. Our work with *lin-42*/*period* and *let-7* in *C. elegans* and human cells reinforces the growing body of evidence that post-transcriptional regulation by miRNAs is a shared feature of *period*-based regulatory modules that control changes in cell state, be they circadian transitions between different physiological states or developmental transitions in stem/progenitor cell fate.

ACKNOWLEDGEMENTS

We thank the CGC for nematode strains, and the *C. elegans* and *C. briggsae* genome sequencing consortia for sequence data. We also thank Kristy Reinert for technical assistance with *C. elegans* microinjections. This work was supported by NIH R01 grant (GM64701) to F.J.S. and Virginia Tech start-up grant to D.B.

FIGURE LEGENDS

Figure 1. *lin-42* genomic structure and putative *let-7* family microRNA binding sites (LCSs) in *lin-42* 3'UTRs. (A) *C. elegans* chromosome II genomic region showing the 12 exons that constitute *lin-42* isoforms, 3' regulatory regions of these isoforms, and positions of primers (horizontal arrows) and probes (horizontal bars) used for qRT-PCR. *lin-42A* qRT-PCR primers anneal exon 8 and to exon 7 (dark grey), which is unique to *lin-42A*, and the qRT-PCR probe spans the junction of these two exons. *lin-42B* qRT-PCR primers anneal to sequences in exons 5 and 8 and the qRT-PCR probe spans the junction of these two exons. *lin-42C* qRT-PCR primers and probe anneal exon 6 (dark grey), which is unique to *lin-42C*. *lin-42A* and *lin-42B* transcripts have the same 0.9 kb of 3'UTR sequence that contains three putative *let-7* miRNA binding sites (A/B-LCSs 1, 2, 3). The *lin-42C* 3'UTR consists of approximately 0.15 kb of sequence that contains one putative *let-7* binding site (C-LCS 1). However, six additional putative *let-7* binding sites (C-LCSs 2, 3, 4, 5, 6, 7) are present within 1.3 kb downstream of the *lin-42C* 3'UTR that may constitute alternative 3'UTRs. (B) Partial sequence and structure of mRNA-miRNA duplexes formed between *let-7* and *lin-42A/B* LCSs 1-3, or mutated/scrambled LCSs (A/B-scLCSs), between *let-7* microRNA and *lin-41* LCSs, and between *let-7* and l*in-42C* LCSs1-7. Each duplex shows 12 bases of the *lin-42* or *lin-41* LCS that are predicted to pair with the *let-7* 5' bases, including the seed sequence (5'-GAGGUAG-3'). Mismatched nucleotides, unmatched nucleotides and wobble base pairing between LCS and *let-7* seed bases are indicated by (*). The seed score refers to the largest number of contiguous base pairs formed between the LCS and the 7-base *let-7* seed sequence.

Figure 2. Post-transcriptional regulation via *lin-42* 3'UTRs. Expression of *lin-42* post-transcriptional *lacZ* reporters in N2/wildtype genetic background during larval development (larval stages L1 to L4). Post-transcriptional reporters consist of *col-10* promoter, *lacZ* reporter gene, and one of the following 3'UTRs: *lin-42A/B* 3' UTR (~1.2 kb) containing 3 LCSs, *lin-42C* 3' regulatory sequences (~1.4 kb) containing 7 LCSs, or *unc-54* 3'UTR (~0.9 kb) containing no LCSs. Reporter expression was quantified as the percentage of worms in a synchronized population showing blue staining in the seam cells. *lacZ* expression was measured in 3 independently transformed lines for each type of reporter (individual lines are shown for *lin-42* 3'UTR reporters and an average of three lines is shown for *unc-54* 3'UTR reporter). Reporter expression values for L2 to L4 populations were normalized to reporter expression in L1 population. (*) indicates statistically significant decrease in reporter expression between two stages, as determined by Students t-test, with $P \leq 0.05$. Error bars represent standard deviation.

Figure 3. *let-7* microRNA dependent regulation of expression via *lin-42A/B* 3'UTR.

(A) Expression of *lin-42A/B* 3'UTR post-transcriptional reporter in wildtype (N2/WT) or *let-7* family microRNA loss-of-function genetic backgrounds: *let-7(lf)* = *let-7(n2853),* VT1066 = *lin-58 mir-241*; *mir-84(n4037),* mir-84(lf) = *mir-84(n4037),* mir-48(lf) = *lin-58(n4097),* mir-241(lf) = *mir-241(n4316)*. *lin-42A/B* 3'UTR reporter lines 1 and 3 were crossed into *let-7* miRNA lf strains, and reporter expression values shown are an average of line 1 and line 3 values for each stage. (B) Expression of *lin-42A/B* 3'UTR post-transcriptional reporters with mutant LCSs in wildtype genetic background. Mutant LCS reporters are modifications of *lin-42A/B* 3'UTR (with three wildtype LCSs, +3LCS) and contain scrambled sequences for all 3 LCSs (sc3LCS) or individual LCSs (scLCS1, scLCS2, scLCS3). (A and B) Reporter expression values for L2 to L4

populations were normalized to reporter expression in L1 population. (*) indicates statistically significant decrease in reporter expression between two stages, as determined by Students t-test. $P \leq 0.05$, except in (A) where $P \leq 0.10$ for mir-48(lf), mir-84(lf) and mir-241(lf) genetic backgrounds. Error bars represent standard deviation.

Figure 4. *lin-42* isoform expression patterns over larval development in wildtype and *let-7* family loss-of-function genetic backgrounds. (A) *lin-42A*, *B* and *C* isoform expression during late L3, late L4 and early adult stages in *let-7(n2853)lf* worms relative to expression in N2/wildtype. Expression of *ama-1* and all three *lin-42* isoforms was determined using *ama-1*-specific and *lin-42* isoform-specific qRT-PCR primers and probes, and *lin-42* expression was expressed relative to *ama-1* expression in the same sample. The graphed *lin-42* levels are a ratio of *lin-42* levels in *let-7(n2853)* to *lin-42* levels in N2/wildtype. Ratios were calculated from averages of 4 independent trials for each time point and worm strain. (B-D) *lin-42A* (B), *lin-42B* (C) and *lin-42C* (D) isoform expression, relative to *ama-1* expression, from L2 to adult stages in N2/wildtype and VT1066 genetic backgrounds. *lin-42* isoform expression was determined by qRT-PCR as in (A). 'Time' on the X axis refers to hours post addition of synchronized, starved L1 larvae to bacterial food and grown at 25°C. Grey vertical bars at approximately 12 hrs, 22 hrs, 32 hrs and 41 hrs indicate the times at which molting worms were observed (molting times, MTs) in the N2 populations. The VT1066 expression profile was shifted by approximately -6 hrs in order to bring the VT1066 MTs in register with the N2 population MTs. Each expression value is the average from 2 independent time course experiments for each worm strain. (A-D) Within a single time course experiment, qRT-PCR reactions were run in triplicate for each *lin-42* isoform or *ama-1* at each time point. Error bars represent standard deviation.

Figure 5. Effect of *lin-42* and/or *let-7* family miRNA loss-of-function on seam cell division and terminal differentiation during larval development. (A) Suppression of adult lethality phenotype of *let-7(n2853)lf* , VT1066 (triple mutant of *mir-48*, *mir-241* and *mir-84*) or nDF51 (double mutant of *mir-48* and *mir-241*) on *lin-42(RNAi)*.

(B) Suppression of retarded adult alae formation in *let-7(n2853)lf*, VT1066 or nDF51 on *lin-42(RNAi)*. (A and B) For each worm strain and RNAi condition, we counted the number of viable worms in a minimum synchronized population of 200 young adult worms (A), or the number of worms showing adult alae for a minimum of 50 worms from a synchronized population (B). Values shown are the averages of at least 3 independent experiments, and error bars represent standard deviation. (C) Seam cell divisions during late larval development (early L3 to early adult) in *wIs79* or *nDF51;wIs79* worm strains on *lin-42(RNAi)*. For each time point and worm strain, we counted the number of GFP-positive seam cell nuclei in a single worm for a minimum of 85 worms from a synchronized population. Each time point count is the average of at least 3 independent time-course experiments, and error bars represent standard deviation.

Figure 6. *let-7a* mediated post-transcriptional repression via *Period* 3'UTRs in HeLa cells.

(A) Putative *let-7* binding sites in the 3'UTRs of human (Hs) *Period1*, *Period2* and *Period3*. Each Hs *Period* 3'UTR contains multiple atypical LCSs. The partial sequence and structure of two *let-7a*:*Per* LCS duplexes, closest to canonical miR:mRNA duplexes, is shown for each Hs *Per* 3'UTR. Each duplex shows 12 bases of the *period* LCS that are predicted to pair with *let-7* 5' bases. Mismatched nucleotides, unmatched nucleotides and wobble base pairing between LCS

and *let-7* seed bases are indicated by (*). The seed score refers to the largest number of contiguous base pairs formed between the LCS and the 7-base *let-7* seed sequence.

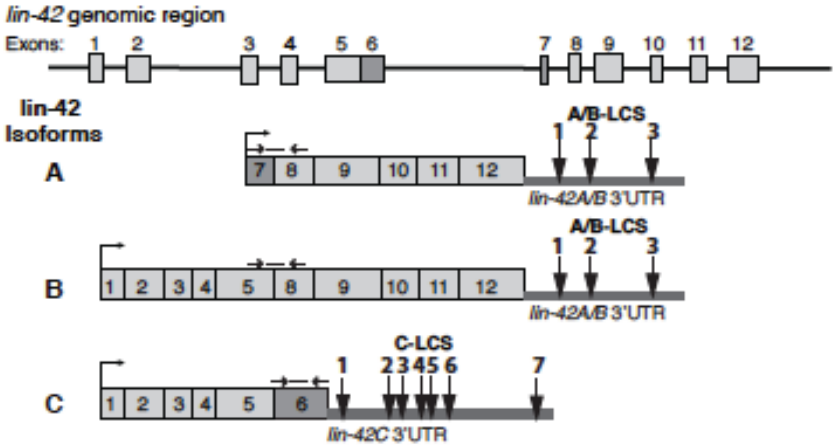
(B) Expression of Hs *Period* 3'UTR reporters in HeLa cells transfected with *let-7* antisense RNA. Post-transcriptional reporters consist of SV40 promoter, *luciferase* reporter gene, and one of the following 3'UTRs: Hs *Period1* 3'UTR (~650 bp) containing 7 putative LCSs, Hs *Period2* 3'UTR (~2400 bp) containing 20 putative LCSs, Hs *Period3* 3'UTR (~2500 bp) containing 16 putative LCSs. Each *Period* post-transcriptional reporter was separately transfected into HeLa cells and co-transfected with *let-7* antisense RNA (anti-*let-7a*) or an antisense RNA of sequence unrelated to *let-7a* (negative control). Reporter expression in cells transfected with anti-*let-7a* is shown normalized to expression from cells transfected with control antisense RNA. Values shown are the average of three independent experiments. (*) indicates statistically significant increase in reporter expression, as determined by Students t-test, with $P \leq 0.05$ for *Per1* and *Per3* reporters, and with $P \leq 0.10$ for *Per2* reporter. Error bars represent standard deviation.

Figure S1. *let-7* microRNA:*lin-42A/B* duplexes. Putative duplexes formed between *C. elegans lin-42A/B* 3'UTR LCS1 (A) or LCS2 (B) and *let-7* family microRNAs: *let-7*, *mir-84*, *mir-48* and *mir-241*. Base-pairs in red indicate binding between the microRNA seed sequence and *lin-42* LCS. Base-pairs in blue indicate three or more contiguous base-pairs formed between 3' region of the microRNA and *lin-42* LCS.

Figure S2. *lin-42A/B* LCSs in *C. briggsae* and *C. remanei*. Putative duplexes formed between (A) *C. briggsae* (Cb) *lin-42A/B* 3'UTR LCSs 1-3, or (B) *C. remanei* (Cr) *lin-42A/B* 3'UTR LCSs 1-3 and *let-7* microRNA. Base-pairs in red indicate binding between the microRNA seed

sequence and *lin-42* LCS. Base-pairs in blue indicate three or more contiguous base-pairs formed between 3' region of the microRNA and *lin-42* LCS. Mismatched nucleotides, unmatched nucleotides and wobble base pairing between LCS and *let-7* seed bases are indicated by (*). The seed score refers to the largest number of contiguous base pairs formed between the LCS and the 7-base *let-7* seed sequence. (C) Seed score and position of *C. elegans* (Ce), *C. briggsae* (Cb) and *C. remain* (Cr) *lin-42A/B* LCSs. LCS position refers to the last 3' base of the LCS that base-pairs with the microRNA 5' seed sequence and is measured starting from the first base of the 3'UTR after the exonic STOP codon.

A

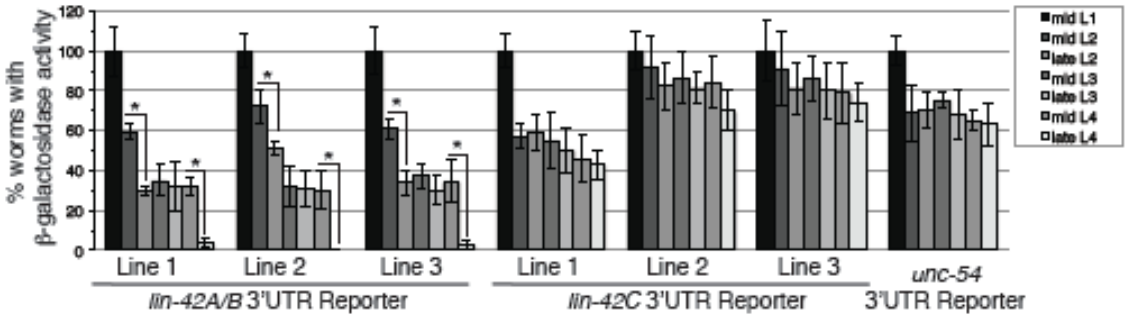


B

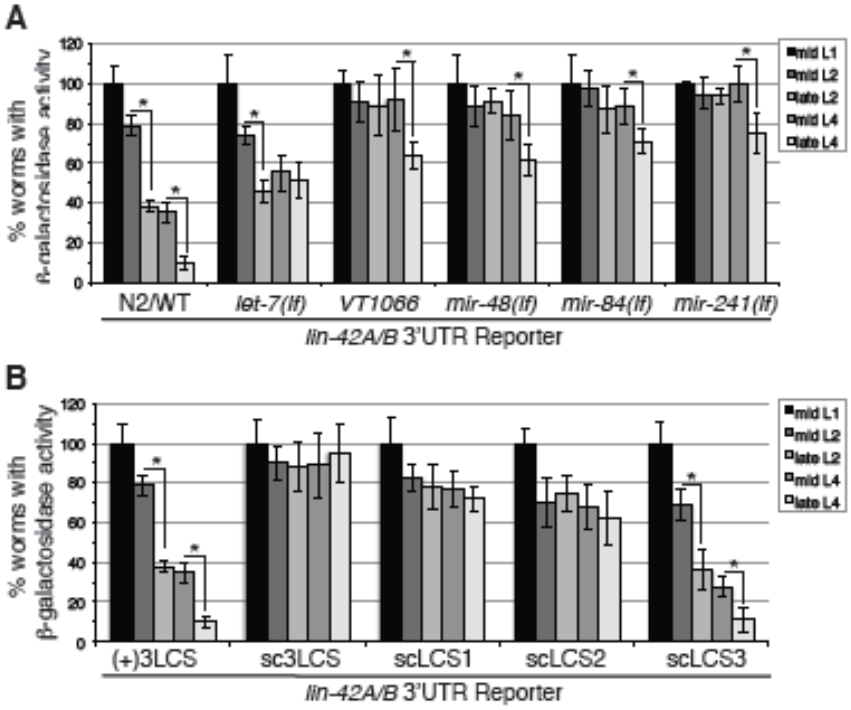
		Seed Score
A/B-LCS1	5'---CAUGACUCCUCG---3'	4/7
let-7	3'---UGGAUGAGGAGU-5'	
A/B-scLCS1	5'---CCCAAGCAAGAC---3'	2/7
let-7	3'---GAUGAUGGAGU-5'	
A/B-LCS2	5'---CAUUUCGCCUCU---3'	5/7
let-7	3'---GAUGA UGGAGU-5'	
A/B-scLCS2	5'---ACCAACCCAAAA---3'	1/7
let-7	3'---GGAUGAUGGAGU-5'	
A/B-LCS3	5'---AUUUGAACCUCU---3'	5/7
let-7	3'---AUGA UGGAGU-5'	
A/B-scLCS3	5'---UUGGGCAAACAC---3'	1/7
let-7	3'---GGAUGAUGGAGU-5'	
lin-41 LCS1	5'---GUUCUACACUCA---3'	4/7
let-7	3'---GAUGAUG GAGU-5'	
lin-41 LCS2	5'---CCAUUCUGCCUC---3'	7/7
let-7	3'---UGGAUGAUGGAGU-5'	

		Seed Score
C-LCS1	5'---UACAGUACCUUA---3'	6/7
let-7	3'---GGAUGAUGGAGU-5'	
C-LCS2	5'---AGGAGUACCUGU---3'	5/7
let-7	3'---GGAUGAUGGAGU-5'	
C-LCS3	5'---UACAGUACCUUU---3'	6/7
let-7	3'---GGAUGAUGGAGU-5'	
C-LCS4	5'---AGGUGUGACCUA---3'	4/7
let-7	3'---UGGAUA UGGAU-5' (G G)	
C-LCS5	5'---ACUUAAUCCUCG---3'	4/7
let-7	3'---AUGAU GGAGU-5'	
C-LCS6	5'---CGAAAAUACCUA---3'	5/7
let-7	3'---UGGAUGAUGGAU-5' (G)	
C-LCS7	5'---CUACCACCCUCC---3'	4/7
let-7	3'---GAUGAU GGAGU-5'	

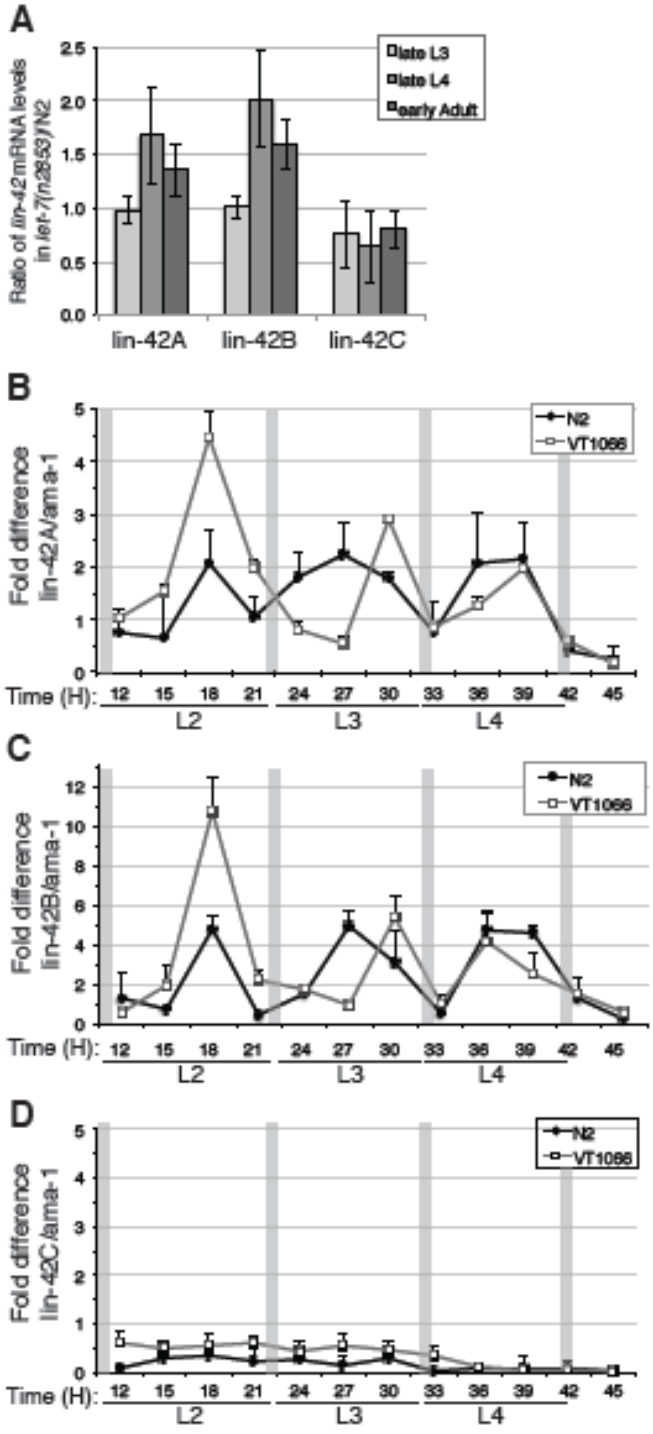
James et al. 2012
Figure 2



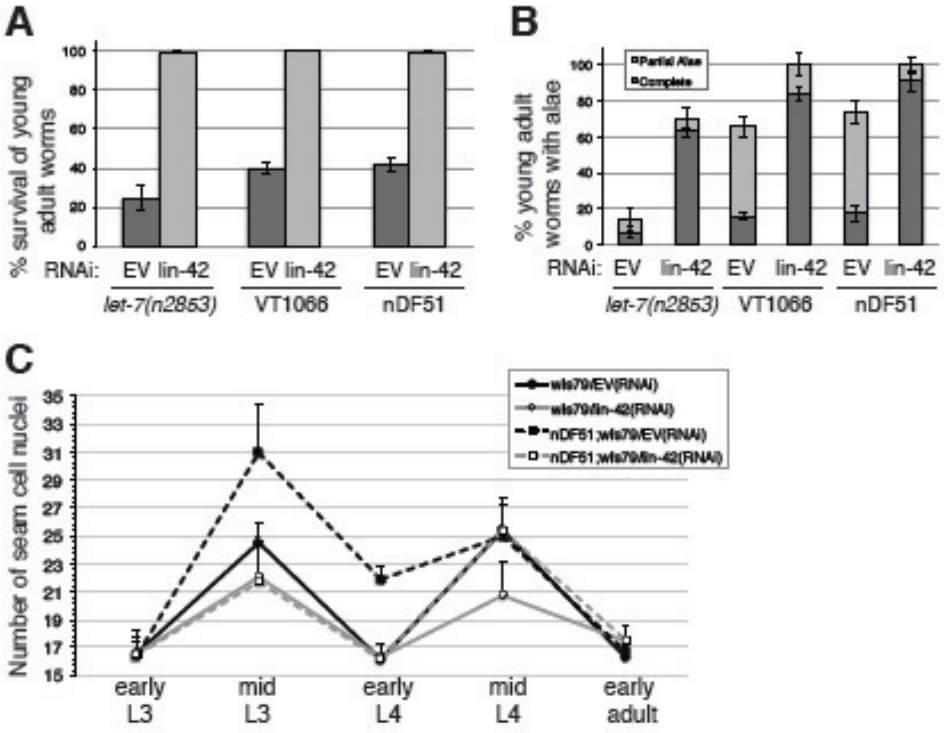
James et al. 2012
Figure 3



James et al. 2012
Figure 4



James et al. 2012
Figure 5

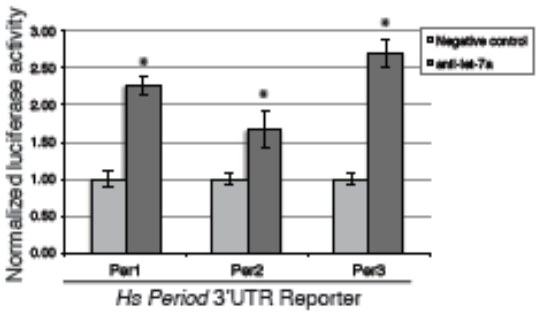


A

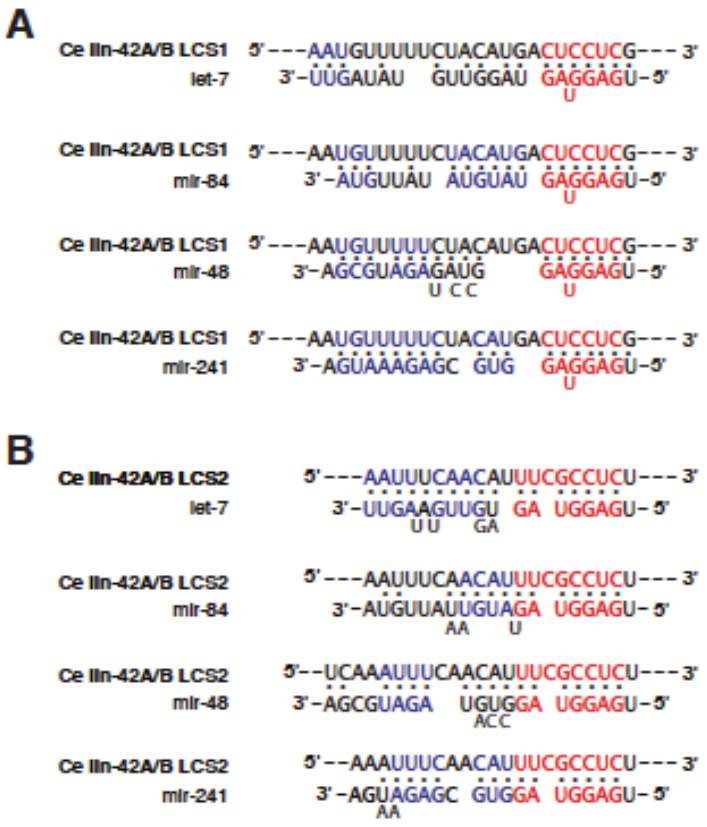
Period gene	3'UTR length	number of LCSs
Hs Per1	596 bp	7
Hs Per2	2329 bp	20
Hs Per3	2422 bp	16

		Seed Score
Hs Per1-LCS1	5'---GGGGCUGCCUCU---3'	7/7
let-7	3'---GGAUGAUGGAGU-5'	
Hs Per1-LCS2	5'---AGUGUCACCUCC---3'	5/7
let-7	3'---UGGAUGUGGAGU-5'	
	A	
Hs Per2-LCS1	5'---ACCUCUGCCUCC---3'	7/7
let-7	3'---GGAUGAUGGAGU-5'	
Hs Per2-LCS2	5'---UCUCCUGCCUCA---3'	7/7
let-7	3'---GGAUGAUGGAGU-5'	
Hs Per3-LCS1	5'---AGCCCUACCACA---3'	5/7
let-7	3'---GGAUGAUGGAGU-5'	
Hs Per3-LCS2	5'---GAGCGCUCCUCC---3'	4/7
let-7	3'---UGGAUGAGGAGU-5'	
	U	

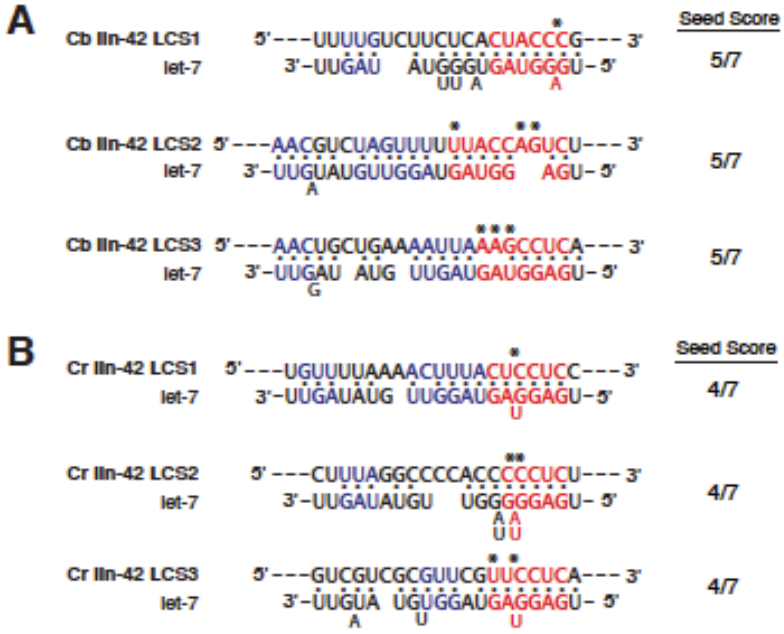
B



James et al. 2012
Figure S1



James et al. 2012
Figure S2



C

lin-42 LCS	3'UTR Position			Seed Score		
	Ce	Cb	Cr	Ce	Cb	Cr
LCS1	204	170	180	4/7	5/7	4/7
LCS2	383	207	321	5/7	5/7	4/7
LCS3	739	730	830	5/7	5/7	4/7

Table S1. Sequences of primers and probes used in study.

Organism	Target Gene	Primer/Probe	Sequence 5' - 3'	Primer/Probe Purpose
C. elegans	lin-42	F47F6.20	TCC CCG CGG TAA GCT ACT GCC CAA TTT TTC	Amplification of lin-42A/B 3'UTR
C. elegans	lin-42	F47F6.21	CCA CAT GTG CTT CCT CGA TGC ACC ATG TTAC	Amplification of lin-42A/B 3'UTR
C. elegans	lin-42	F47F6.22	TCC CCG CGG AAT TCG GGA AAA GGA CCA GAAG	Amplification of lin-42C 3'UTR
C. elegans	lin-42	F47F6.23	CCA CAT GTG GCA CTT GGG ACG GGA GAG AGAG	Amplification of lin-42C 3'UTR
C. elegans	lin-42	F47F6.40	GCC TAG ACA ATG GAA GTT TTG ATA TAA ATCC	Site-directed mutagenesis of lin-42A/B 3'UTR LCS1
C. elegans	lin-42	F47F6.41	ATC AAA ACT TCC ATT GTC TAG GCT TTT CAG GC	Site-directed mutagenesis of lin-42A/B 3'UTR LCS1
C. elegans	lin-42	F47F6.42	AAG CTG ATT TCA CTA TCC TTT ACG TTT CCT GC	Site-directed mutagenesis of lin-42A/B 3'UTR LCS2
C. elegans	lin-42	F47F6.43	ACG TAA AGG ATA GTG AAA TCA GCT TGT CGA GAGC	Site-directed mutagenesis of lin-42A/B 3'UTR LCS2
C. elegans	lin-42	F47F6.44	CAA AAA TCT TCC ATT TCC TGC CTT TTC TCC CGCG	Site-directed mutagenesis of lin-42A/B 3'UTR LCS3
C. elegans	lin-42	F47F6.45	AAG GCA GGA AAT GGA AGA TTT TTG GGT GGG TGG	Site-directed mutagenesis of lin-42A/B 3'UTR LCS3
C. elegans	lin-42	lin-42A F1	ACG ATC TTG CAG AGC CAG TAC	qRT-PCR primer for lin-42A
C. elegans	lin-42	lin-42A R3	GCC CTG GTT CTG ATC CTT G	qRT-PCR primer for lin-42A
C. elegans	lin-42	lin-42B-F6	AAT CGG AGA CTC GGA TGT G	qRT-PCR primer for lin-42B
C. elegans	lin-42	lin-42B-R6	TGA CGG GAG GTG GAT GG	qRT-PCR primer for lin-42B
C. elegans	lin-42	lin42C-F5	TTC AGT AAC CTT CTT ATC ATC CG	qRT-PCR primer for lin-42C
C. elegans	lin-42	lin42C-R4	GCT GCT GAC CAC TCT CG	qRT-PCR primer for lin-42C
C. elegans	arnt-1	arnt1-F3	AAG GTC GCA GGT GGA TGC	qRT-PCR primer for arnt-1
C. elegans	arnt-1	arnt1-R3	GTC CTC ATT CAC GTT CTT CTT CC	qRT-PCR primer for arnt-1
C. elegans	lin-42	lin42-TM-A1	CACAGCCACCACCACCATCACTCAAGC	qRT-PCR Taqman probe for lin-42A
C. elegans	lin-42	lin42-TM-B1	ACCACCACCAGGCATCCAGTCAA	qRT-PCR Taqman probe for lin-42B
C. elegans	lin-42	lin42-TM-C1	TGCCAGTCACCGCTTATCTTCTTCTTCAGT	qRT-PCR Taqman probe for lin-42C
C. elegans	arnt-1	arnt1-TM1	TCGCTATCAACCATCATACCGCCGTGTC	qRT-PCR Taqman probe for arnt-1
H. sapiens	period1	hPer 1.1	GCT CTA GAA CTC CAT TCT GGG ACC ATC TCCA GGAG	Amplification of period1 3'UTR
H. sapiens	period1	hPer 1.2	CGT CTA GAA TGC CAT CGG CAG AGG GTA CAG CTGAG	Amplification of period1 3'UTR
H. sapiens	period2	hPer 2.1	GCT CTA GAC CCC TGC CCC ACC TCA GCC CGGC AGCC	Amplification of period2 3'UTR
H. sapiens	period2	hPer 2.2	CGT CTA GAC TGG GTA AAG AAA GTT TGG TTT GCA CAC	Amplification of period2 3'UTR
H. sapiens	period3	hPer 3.4	GCT CTA GAG AGT GAC TGT GAG GAT GAA CCT TCA TACC	Amplification of period3 3'UTR
H. sapiens	period3	hPer 3.5	CGT CTA GAT GTT GTT TTT ACT TAA AAA GAT ACC	Amplification of period3 3'UTR

APPENDIX B

Protocol for *C. elegans* Nuclear Extracts (adapted from [218])

REQUIREMENTS:

- Solutions I, II and III
- 100 X Halt Protease and Phosphatase inhibitor cocktail from Pierce
- M9 ($MgSO_4$ added)
- Autoclaved 2.0 mL eppendorf tubes
- Table-top centrifuge @ 4°C (13,400g and 16,100g)
- Sonicator
- Vortex
- Liquid nitrogen for flash-freezing
- All tubes must be placed on ice at all times including solutions I, II and III
- Inhibitor cocktail must be supplemented with Pepstatin A (from RT stock) and then added to the homogenate at steps 4, 5 and 8 to a final concentration of 1X.

PROCEDURE:

The following solutions were prepared in advance, filter sterilized and stored at 4°C:

Soln I	Soln II	Soln III
Tris-HCl (pH 7.5) 50mM	Tris-HCl (pH 7.5) 50mM	Tris-HCl (pH 7.5) 50mM
K-acetate 10mM	K-acetate 100mM	K-acetate 400mM
DTT 5mM	DTT 5mM	Mg-acetate 2mM
	Mg-acetate 2mM	
For 20mL of each soln: 50mM Tris 157.6mg, 10mM K-Ac 9.63mg, 5mM DTT 15.4mg		

The worms were washed off the plates with M9 and further washed 3X to get rid of as much media/bacteria as possible. The dry pellets of worms were flash frozen for about 15s and placed in at -80°C if extracting later. For every 100uL of packed worms, 200uL of Soln I, 100uL of

Soln II and 150uL of Soln III was added.. Half the indicated volume of Soln I along with the worm pellet was sonicated in a sonicator- 10 pulses of 0.5 seconds each for a total of 8 times. The pellets were placed on ice in between pulses. The indicated volume of Soln II was then added and incubated on ice for 20-25 min. The homogenate was then centrifuged for 30 min at 4°C @ 13400g. The supernatant was saved in a pre-chilled tube as the Cytoplasmic fraction. The pellet was then re-suspended in the indicated volume of Soln III and incubated on ice for 30 min, vortexing for 15s every 10 min. The homogenate was then centrifuged for 30 min at 4°C @ 16100g. The supernatant was saved in a pre-chilled eppendorf as the nuclear fraction. The protein concentrations were measured by a Bradford assay and the extracts were stored at -80°C.

APPENDIX C

Primers for EMSA Probes

Target Promoter	EMSA Probe	Probe size (bp)	Forward Primer	Forward Primer Sequence (5' - 3')	Reverse Primer	Reverse Primer Sequence (5' - 3')
lin-42B/C	1	200bp	L42_PROM_F1	TTG GCT GAT GGT GCC ACG	L42_PROM_R1	GCA TTT GCA TTA TAT TAC ATT TTA GTA TAC
lin-42B/C	2	200bp	L42_PROM_F6	AGA AAA ATT GAG GTT TGG GGG AAA TTT CG	L42_PROM_R6	CAG TTG TTA ACT TTG TTC TAG CTT TGA TG
lin-42B/C	3	200bp	L42_PROM_F2	TGC AGT TAC AGA GAA TAA ATT TTA GGC	L42_PROM_R2	GAA AGT CCT GTA ACT AGT TTA TAA TTG G
lin-42B/C	4	200bp	L42_PROM_F7	TTT CAT CAA AGC TAG AAC AAA GTT AAC AAC	L42_PROM_R7	TAT TCA AGC ACT TTT TAT TGA TTT TTA TGG G
lin-42B/C	5	200bp	L42_PROM_F3	CTT TCT AGA TGC CAA AAC TTC ACA CAG CTC CAG C	L42_PROM_R3	ATT GCG CGG CGC GGG AG
lin-42B/C	6	200bp	L42_PROM_F8	CCC CAT AAA AAT CAA TAA AAA GTG CTT GAA TAC	L42_PROM_R8	CCC TAA ATT CCC GTT CTA CAC ATC
lin-42B/C	7	200bp	L42_PROM_F4	TCG TGT GCT GGG GCC TAC	L42_PROM_R4	TTC CAG GAA GCT GTC CAA ACA GCA CG
lin-42B/C	8	200bp	L42_PROM_F9	ATG TGT AGA ACG GGA ATT TAG GG	L42_PROM_R9	CTT CTA TGG CTA AAA AAC TCC AGC
lin-42B/C	9	200bp	L42_PROM_F5	CTG GAA GTC TAG CAG TAT CGG ATA AAT GTA G	L42_PROM_R5	TTT GCA TTT TTC CGA CGG CTC
lin-42B/C	10	200bp	L42_PROM_F10	GAA GCA TTT TTC GTG GAG TTC TCT AGG	L42_PROM_R10	GTA CCC TAA AAA AAA ATA CTA AAT TGA GCG TTA GGG G
lin-42B/C	11	200bp	L42_PROM_F11	AAA TCA AAA AAC CAC TAC GAA ACT ACG	L42_PROM_R11	TGT TAT CCC ACC CAA ATC TCA GAG
lin-42B/C	12	200bp	L42_PROM_F12	GTA CTG TAG CTT AAA AAG TAC GCA AAC ACG	L42_PROM_R12	AAA GCT CTA CGG TGT TGG CG
lin-42B/C	13	200bp	L42_Prom_F13	GGG ATA ACA CGC AGA AGA CAA ATT TCT AAC	L42_Prom_R13	ATA TTT CAA TTT TTA TTC ATA AAA ATT GGC TTC AAT GTG TG
lin-42B/C	14	200bp	L42_Prom_F14	TAC GAT AAA ATG CAC AAA TTT TGT AAA TTT GCT AAA AAT TTGGC	L42_Prom_R14	AAT TTT TAA AGA TTT GTG AAA TTT TTC GTA GAG CTC TAC G
lin-42B/C	15	200bp	L42_Prom_F15	TTG AAA TTT TAC AAA AAA TCT GTA CTT TGG GG	L42_Prom_R15	AGA TAT TGA TTT TTA GTT TTA ATT GGT ATT TTT TTA AAT AGA TTT TCG
lin-42B/C	16	200bp	L42_Prom_F16	TGC TAA AAA TTC ATG TGG AAG CTG	L42_Prom_R16	GAA GGT CGT CCA TTT TTA CCC C
lin-42B/C	17	200bp	L42_Prom_F17	TAT GTA GTA AAT CTT GCA AAT TCA AAA AAA AAG TGG C	L42_Prom_R17	TAA AAC ATT AGA CCA TAC CTT ATC AAT TGT CAG TTT C
lin-42B/C	18	200bp	L42_Prom_F18	CAC GGT TTT TTG GAG GGT TTT CTG AC	L42_Prom_R18	ACG CAA CAC AAC GTA GAC CG
lin-42B/C	4-6	65bp	L42_PROM_F3.5	CTCGGTTTAGACCCC	L42_PROM_R3.5	GTGTTGGAACAACGT G
myo-2	F1/R1	200bp	Myo2_PROM_F1	TCC GAT TGC TAT CAT GTG AAT CTG	Myo2_PROM_R1	ATC AGA GTT GTG ATT TCC TGT GC
myo-2	F2/R2	200bp	Myo2_PROM_F2	TTG TGT GGA TAA	Myo2_PROM_R2	AAG AGA GCA GAG

				GAG TAG CAA AAT GGC		AAC CTC AAC G
egl-18	F/R	200bp	egl-18_PROM_F	CGG AGG GAT CAT TAT TAG CAC	egl-18_PROM_R	CGT AGT ATC ACC CCG ACA C

APPENDIX D

C. elegans EMSA Protocol

REQUIREMENTS

- Nuclear extract of known protein concentration
- Biotin-labeled DNA, Unlabeled DNA fragments- specific, non-specific competitors
- Autoclaved de-ionized H_2O
- 5% denaturing polyacrylamide gel
- 0.5X TBE buffer
- Biodyne charged nylon membranes from Pierce
- Transfer apparatus and UV cross-linker
- Chemiluminescent Nucleic acid detection module

Additives:

- 100X inhibitor cocktail supplemented with Pepstatin A
- 80mM $MgCl_2$; 80mM $ZnCl_2$
- 1ug/uL Poly dI:dC
- 1% NP-40
- 1ug/uL BSA 1:100 diluted
- 10X Binding Buffer (adapted from [119]): made and stored at -20°C

Components	10X	1X (final conc. In 20uL reaction)
HEPES	262.25mM	26.25mM
Glycerol	30%	3%
KCL	600mM	60mM
$MgCl_2$	40mM	4mM
Spermidine	40mM	4mM
DTT	5mM	0.5mM

REACTION SET-UP:

Components	DNA only	DNA + extract	DNA + extract + unlabeled excess of SC	DNA+ extract + unlabeled excess of NSC
d.i. H_2O	Up to 20uL	Up to 20uL	Up to 20uL	Up to 20uL
10X Binding buffer	2uL	2uL	2uL	2uL
100X inhibitor cocktail	0.25uL	0.25uL	0.25uL	0.25uL
80mM Zn; 80mM Mg	2.0uL	2.0uL	2.0uL	2.0uL
1ug/uL Poly dI:dC	1.0uL	1.0uL	1.0uL	1.0uL
1% NP-40	1.0uL	1.0uL	1.0uL	1.0uL
1ug/uL BSA 1:100	0.5uL	0.5uL	0.5uL	0.5uL
Nuclear extract	Up to 6ug	Up to 6ug	Up to 6ug	Up to 6ug
INCUBATE ON ICE FOR 10-15 MINS				
Labeled DNA	3.5fmols	3.5fmols	3.5fmols	3.5fmols
Unlabeled DNA Specific (50X)	0	0	175fmols	0
Unlabeled DNA Non-specific (50X)	0	0	0	175fmols
INCUBATE TUBES ON ICE FOR 45-50 MINS				

The reaction was set up in the order shown in table above. The binding reactions were loaded onto a 5% native polyacrylamide gel (wells flushed out), which had been pre-electrophoresed at 110V for 30-40 mins in 0.5X TBE buffer. No loading dye was added into the binding reactions. The gel was run at 145-155V for about an hour or until the dye had run 3/4th the distance. The Biodyne membrane was soaked in 0.5X TBE for at least 15mins prior to transfer. Transfer was carried out at 380mA for 30 mins in cooled (10°C) 0.5X TBE buffer. Once transferred, the membrane was cross-linked at 120mJ/cm2 using an auto cross-linker. Detection and development of signal was done following kit instructions of the Chemiluminescent Nucleic acid detection module.

APPENDIX E

C. elegans Homologs of *Drosophila* and Vertebrate Transcription Factors Predicted to Bind DNA Motifs in *lin-42B/C* Promoter Regions BR1, BR2 and BR3

Table E.1: Candidates tested through the RNAi screen

Candidate #	Binding Region	TF Family	Gene	Gene #	Gene Product Description	Post-Embryonic Expression
1	BR1	HIFF	**aha-1**	C25A1.1 1	bHLH transcription factor	Hypodermal, intestinal, pharyngeal
2	BR1	HIFF	**ahr-1**	C41G7.5	Aryl-hydrocarbon receptor nuclear translocator	Neural
3	BR1	HOMF, NKX6	**alr-1**	R08B4.2	Homeodomain transcription factor	Hypodermal
4	BR1	DHOM, NKX6	**ceh-2**	C27A12. 5	Homeodomain transcription factor	Vulval, muscle, pharyngeal
5	BR1	HOMF	**ceh-22**	F29F11. 5	Hom of Transcription factor tinman/NKX2-3	Intestine, pharynx, neural(VNC)
6	BR1	DHOM	**ceh-24**	F55B12. 1	Homeodomain transcription factor	Vuval, muscle
7	BR1	NKX6	**ceh-6**	K02B12. 1	Hom of OCT1 TF, POU and HOX domains	Hypodermal
8	BR1	HIFF	**cky-1**	C15C8.2	bHLH-PAS TF	Pharyngeal
9	BR1	EBOX	**crh-1**	Y41C4A. 4	cAMP response element binding protein, CREB homolog	Constitutive
10	BR1	PCBE	**fzr-1**	ZK1307. 6	WD40 repeat DNA binding protein	Broad expression
11	BR1	HIFF	**hif-1**	F38A6.3	bHLH TF	Broad expression
12	BR1	DHOM, HOMF	**lin-39**	C07H6.7	hom zerknullt, Dm deformed, sex combs reduced	Ventral epidermal cells, VNC
13	BR1	DHOM, HOMF	**mab-5**	C08C3.3	Homeodomain transcription factor	Hypodermal, Q lineage
14	BR1	EBOX	**mls-1**	H14A12. 4	Tbx1 subfamily of T-box transcription factors	Vulva, uterine, M lineage
15	BR1	NKX6	**pal-1**	C38D4.6	Homeodomain transcription factor, Caudal homolog	Hypodermal, neural, intestine
16	BR1	HOMF	**pax-3**	F27E5.2	Homeodomain transcription factor	Hypodermal, neural
17	BR1	EBOX	**tbx-8**	T07C4.2	TBX2 and related T-box transcription factor	Hypodermal
18	BR1	EBOX	**tbx-9**	T07C4.6	TBX2 and related T-box transcription factor	Hypodermal
19	BR1	HOMF	**vab-15**	R07B1.1	Homeodomain transcription factor	Hypodermal SC, VNC
20	BR1, BR2	E2FF, GATA	**elt-6**	F52C12. 5	GATA transcription factor	Seam cells, neural
21	BR1, BR2	GATA	**zfp-1**	F54F2.2	Leucine Zipper, Zn finger transcription factor	Seam Cells
22	BR1, BR2	GATA	**ztf-16**	R08E3.4	Leucine Zipper, Zn finger transcription factor	Seam Cells
23	BR1, BR2, BR3	HOMF	**pop-1**	W10C8.2	HMG-box transcription factor, TCF/LEF homolog	Seam cells, vulva, somatic gonad
24	BR1, BR2,	NKX6	**skn-1**	T19E7.2	bZIP transcription factor	Hypodermal, neural, intestine

	BR3					
25	BR1, BR3	NKX6	**ces-2**	ZK909.4	bZIP transcription factor	Hypodermal, neural, muscle
26	BR2	PLZF	**blmp-1**	F25D7.3	Zn finger transcription factor	Seam cells, neural
27	BR2	DHOM	**ceh-32**	W05E10.3	Homeodomain transcription factor	Hypodermal, neural
28	BR2	GATA	**egl-18**	F55A8.1	GATA transcription factor	Seam cells, neural
29	BR2	SP1F	**klf-3**	F54H5.4	Zn finger transcription factor	Hypodermal
30	BR2	HOMF	**nhr-23**	C01H6.5	Nuclear hormone receptor transcription factor	Seam cells
31	BR2, BR3	NKX6	**daf-16**	R13H8.1	Forkhead transcription factor, FOXO homolog	Hypodermal, vulva, intestine, neural
32	BR2, BR3	DHOM, NKX6	**irx-1**	C36F7.1	Homeodomain transcription factor	Hypodermal, neural
33	BR2, BR3	OCT1	**lin-11**	ZC247.3	LIM homeodomain transcription factor	Vulval, muscle, neural
34	BR2, BR3	NKX6	**lin-31**	K10G6.1	Winged helix-like transcription factor	Vulval
35	BR2, BR3	OCT1	**mec-3**	F01D4.6	Homeodomain transcription factor	Neural, hypodermal
36	BR2, BR3	HOMF	**nhr-1**	R09G11.2	Nuclear hormone receptor transcription factor	Hypodermal
37	BR2, BR3	HOMF	**nhr-41**	Y104H12A.1	Nuclear hormone receptor transcription factor	Seam cells
38	BR2, BR3	HOMF	**nhr-6**	C48D5.1	Nuclear hormone receptor transcription factor	Neural, spermatheca
39	BR2, BR3	HOMF	**nhr-64**	C45E1.1	Nuclear hormone receptor transcription factor	Hypodermis, neurons, pharynx, gut
40	BR2, BR3	HOMF	**nhr-69**	T23H4.2	Nuclear hormone receptor transcription factor	Hypodermis, gut, uterus
41	BR2, BR3	OCT1	**ttx-3**	C40H5.5	Homeodomain transcription factor	Hypodermal, neural
42	BR2, BR3	OCT1	**unc-86**	C30A5.7	POU-domain transcription factor	Hypodermal, neural
43	BR2, BR3	OCT1	**zag-1**	F28F9.1	Homeodomain transcription factor	Neural, intestine

Table E.2: Candidates of the bioinformatics analysis

Candidate #	Binding Region	TF Family	Gene	Gene #	Gene Product Description	Post Embryonic Expression
44	BR1	EBOX	**mml-1**	T20B12.6	bHLH transcription factor	Hypodermal
45	BR1	EBOX	**atf-7**	C07G2.2	bZIP transcription factor, Homolog of FOSB/c-Fos	ND
46	BR1	EBOX	**C27D6.4**	C27D6.4	CREB/transcription factor family transcription factor	Neural, nerve ring
47	BR1	DHOM	**ceh-12**	F33D11.4	HOX domain protein	Neural
48	BR1	DHOM	**ceh-19**	F20D12.6	Homolog of Bar-like Homolog protein from Dm & vertebrates	ND
49	BR1	HOMF	**ceh-22**	F29F11.5	Homolog of Transcription factor tinman/NKX2-3	Intestine, pharynx, neural(VNC)
50	BR1	DHOM	**ceh-27**	F46F3.1	Homolog of Tinman transcription factor	Embryonic
51	BR1	HOMF	**ceh-31**	C33D12.1	Homeodomain protein	Pharyngeal, neural

52	BR1	DHOM, NKX6	**ceh-5**	C16C2.1	Homeodomain protein	Intestine, neural
53	BR1	DHOM	**ceh-7**	C34C6.8	Homeodomain protein	Male tail, rectum
54	BR1	DHOM	**ceh-9**	Y65B4BR.9	Homeodomain protein	Neural
55	BR1	DHOM, HOMF, NKX6	**cog-1**	R03C1.3	Homolog NKX6 transcription factors	Vulval and uterine
56	BR1	HOMF	**dsc-1**	C18B12.3	ND	Intestine, neural
57	BR1	HOMF	**egl-5**	C08C3.1	Homeodomain Transcription Factor	Neural, rectal
58	BR1	PCBE	**gad-1**	T05H4.14	WD40 repeat protein	Embryonic
59	BR1	EBOX	**hlh-10**	ZK682.4	Transcription factor HAND2/Transcription factor TAL1/TAL2/LYL1	Neural
60	BR1	EBOX	**hlh-11**	F58A4.7	bHLH transcription factor	ND
61	BR1	HIFF	**hlh-30**	W02C12.3	bHLH transcription factor	Intestine
62	BR1	HIFF	**hlh-34**	T01D3.2	Homolog SIM1, bHLH transcription factor	Intestine, neural
63	BR1	EBOX	**hlh-8**	C02B8.4	Homolog of TWIST	M lineage, intestine
64	BR1	PCBE	**K02D7.2**	K02D7.2	Homolog SNAI2/SLUG, Txn Rep?	ND
65	BR1	DHOM, HOMF	**lin-39**	C07H6.7	Homolog zerknullt, Dm deformed, sex combs reduced	Ventral epidermal cells, VNC
66	BR1	EBOX	**mab-9**	T27A1.6	Homolog of TBX1, Brachury	Neural
67	BR1	EBOX	**mdl-1**	R03E9.1	bHLH transcription factor	Hypodermis, pharynx, nerve cord
68	BR1	DHOM, NKX6	**mls-2**	C39E6.4	T-box transcription factor	M lineage, uterine muscle cells
69	BR1	DHSF	**mpz-1**	C52A11.4	multi-PDZ domain containing protein	Vulval, muscle, neural
70	BR1	EBOX	**mxl-1**	T19B10.11	bHLH transcription factor	Intestine, neural
71	BR1	DHOM, HOMF	**nob-1**	Y75B8A.2	Homolog Vertebrate Hox9-13, Dm Abd-B transcription factor,	Neural
72	BR1	DHOM	**pal-1**	C38D4.6	Homolog of Caudal	ND
73	BR1	DHOM, HOMF	**pha-2**	M6.3	HOX domain transcription factor	Pharyngeal, intestinal, rectal
74	BR1	DHOM, HOMF	**php-3**	Y75B8A.1	Homolog Dm Abd-B transcription factor	Pharyngeal
75	BR1	PCBE	**prp-4**	C36B1.5	U4/U6 small nuclear ribonucleoprotein Prp4	ND
76	BR1	PCBE	**T02H6.1 a,b**	T02H6.1	ND	ND
77	BR1	DHOM, HOMF, NKX6	**tab-1**	F31E8.3	HOX domain transcription factor	Neural, intestine
78	BR1	EBOX	**tbx-7**	ZK328.8	T-Box HOX domain transcription factor	ND
79	BR1	HOMF	**ttx-1**	Y113G7A.6	HOX domain transcription factor	Neural, pharyngeal
80	BR1	HOMF	**unc-30**	B0564.10	Homolog of Transcription factor PTX1	Neural
81	BR1	DHOM	**vab-7**	M142.4	Homolog even-skipped, EVX1	Neural
82	BR1, BR2	HEAT	**hsf-1**	Y53C10A.12	Heat shock induced transcription factor	Neural, intestine, muscle
83	BR1,	HEAT	**Y53C10**	Y53C10	paralog of hsf-1	Neural

	BR2		**A.3**	A.3		
84	BR2	PLZF	**ces-1**	F43G9.1	Zn finger protein	Neural
85	BR2	GATA, SP1F	**che-1**	C55B7.1 2	Zn finger protein	Neural
86	BR2	DHSF	**dsh-1**	C34F11. 9	Homolog of Dishevelled, paralog of mig-5	Pharyngeal, neural
87	BR2	DHSF	**dsh-2**	C27A2.6	Homolog of Dishevelled	Intestine, neural
88	BR2	E2FF	**efl-1**	Y102A5 C.18	E2F-like transcription factor	Gonad
89	BR2	E2FF	**efl-2**	Y48C3A. 17	ND	ND
90	BR2	SP1F	**egrh-1**	C27C12. 2	EGR (Early Growth factor Response factor) homolog	ND
91	BR2	SP1F	**egrh-3**	Y94H6A. 11	EGR (Early Growth factor Response factor) homolog	ND
92	BR2	GATA	**ehn-3**	ZK616.1 0	Zn finger transcription factor	ND
93	BR2	GATA	**elt-1**	W09C2.1	GATA transcription factor	Embryonic
94	BR2	E2FF, GATA	**elt-2**	C33D3.1	GATA transcription factor	Intestine
95	BR2	E2FF, GATA	**elt-4**	C39B10. 6	GATA transcription factor	Gut
96	BR2	E2FF, GATA	**elt-7**	C18G1.2	GATA transcription factor	Intestine
97	BR2	GATA	**end-1**	F58E10. 2	GATA-4/5/6 transcription factors	Embryonic
98	BR2	GATA	**end-3**	F58E10. 5	GATA-4/5/6 transcription factors	Intestinal
99	BR2	GATA	**eor-1**	R11E3.6	Zn finger transcription factor	Broad expression
100	BR2	SP1F	**klf-1**	F56F11. 3	Zn finger protein	Intestine
101	BR2	SP1F	**klf-2**	F53F8.1	Kruppel-like transcription factor	ND
102	BR2	GATA	**med-1**	T24D3.1	GATA transcription factor	Embryonic
103	BR2	GATA	**med-2**	K04C2.6	GATA transcription factor	Embryonic
104	BR2	DHSF	**mig-5**	T05C12. 6	Homolog of Dishevelled	Vulval
105	BR2	PLZF	**pag-3**	F45B8.4	Zn finger protein	Neural
106	BR2	GATA	**sdz-12**	F12E12. 5	Zn finger protein	Somatic gonad, uterine
107	BR2	SP1F	**sptf-2**	T22C8.5	Zn finger protein	Intestine
108	BR2	SP1F	**sptf-3**	Y40B1A. 4	specificity protein transcription factor	Intestine, neural
109	BR2	GATA	**tra-1**	Y47D3A. 6	GLI transcription factor family	intestine, neural
110	BR2	E2FF	**unc-89**	C30A5.7	POU Homeodomain transcription factor	Neural
111	BR2	PLZF	**Y38H8A .5**	Y38H8A. 5	ND	ND
112	BR2	PLZF	**Y55F3A M.14**	Y55F3A M.14	no description	Intestine, neural
113	BR2	PLZF	**ZK337.2**	ZK337.2	Zn finger protein	Muscle - body wall, vulval
114	BR2	PLZF, SP1F	**ztf-23**	Y54E10B R.8	Zn finger protein	ND
115	BR2, BR3	OCT1	**lin-11**	ZC247.3	LIM Homeodomain transcription factor	Vulva, neurons, muscle
116	BR2, BR3	OCT1	**mec-3**	F01D4.6	HOX domain transcription factor	Neural

117	BR2, BR3	OCT1	**ttx-3**	C40H5.5	HOX domain transcription factor	Neural
118	BR2, BR3	OCT1	**unc-86**	C30A5.7	POU dom transcription factor	Neural
119	BR2, BR3	OCT1	**zag-1**	F28F9.1	Homolog Drosophila zfh-1 and the vertebrate ZEB genes	Neural, intestine

Table E.3: Candidates of the bioinformatics analysis

Candidate #	Binding Region	TF Family	Gene	Gene #	Gene Product Description	Post Embryonic Expression
120	BR1	PCBE	**lin-23**	K10B2.1	Homolog Beta-TrCP/Slimb proteins	Vulval, muscle, neural
121	BR1	EBOX	**ngn-1**	Y69A2AR.29	Neurogenin family protein	Neural
122	BR2	DHSF	**afd-1**	W03F11.6	Actin filament binding protein	ND
123	BR2	E2FF	**csp-2**	Y73B6BL.7	Caspase homolog	ND
124	BR2	DHSF	**dlg-1**	C25F6.2	Membrane-associated guanylate kinase	ND
125	BR2	PL2F	**F57C9.4b**	F57C9.4b	Not transcription factor, secreted protein	Hypodermis, intestine, somatic gonad
126	BR2	DHSF	**magi-1**	K01A6.2	Membrane associated guanylate kinase	ND
127	BR2	DHSF	**par-3**	F54E7.3	Not transcription factor, PDZ-domain protein	Embryo specific

APPENDIX F

PCR Primers Used in Making RNAi Constructs

TF gene	Gene #	Forward Primer	Forward Primer Sequence (5'-3')	Reverse Primer	Reverse Primer Sequence (5'-3')
zfp-1	F54F2.2	zfp-1_F	AAA CGG AAG TAC GCT TCC ACC	zfp-1_R	CGT CGT TTA CGA AGA CCA GAC
ztf-16	R08E3.4	ztf-16_F	ATG CAA GAG CAT TCA ATG ACA C	ztf-16_R	TTC GGG AGT CGG TGA TTT CTC
klf-3	F54H5.4	kfl-3_F	CAT TGA AAG GGA CTC GAG C	klf-3_R	CTT GTC AAT TCA TCA CTT CGT GC
blmp-1	F25D7.3	blmp-1_F	GCG AGT TTA GCA GAA ACA ATT GTG	blmp-1_R	TGG CAT CAC AAA CAT CAC ACG
alr-1	R08B4.2	alr-1_F	TCT TAT GGA ACA GTT ACA GGC AC	alr-1_R	TCA TGA ACT TTC TTC TTT TGG CTT C
crh-1	Y41C4A.4	crh-1_F	AAG GCA TTA CAA GAA GGT GGG G	crh-1_R	TCA CAT TCC GTC CTT TTC CTT TCG
mml-1	T20B12.6	mml-1_F	GCG TCG GTT ACC ACG AAG	mml-1_R	GAG TTG TTG CAT GAC ATA ATC GGG
mdl-1	R03E9.1	mdl-1_F	ATG GAA CAG CAA CTC AAC CTT GG	mdl-1_R	ACT TGG AGG TTG ATT GGC AAG AAC
fzr-1	ZK1307.6	fzr-1_F	GCT GCC AGT ATC AAC GG	fzr-1_R	TAA ATG AAT AAC CGT GCG TAG AG
eor-1	R11E3.6	eor-1_F	CGA CCA AAA GGA ATG AAA AAA CAT C	eor-1_R	CTT AAG CTT ATC ACG CCG ACT AG
lin-23	K10B2.1	lin-23_F	ATG TCT TCA CCG CAC CG	lin-23_R	GGA TAT GGT TTC GCC ACG C
ceh-1	F16H11.4	ceh-1_F	ATG CGG CGA GCC AGA AC	ceh-1_R	ATT GAA GCC CAC ATT TTG AGG C
tra-1.a	Y47D3A.6a	tra-1a_F	CGA GTT GAT GGT GAC TCT GAC	tra-1a_R	CCA AAC TGC ATA GCT TCA GC
tra-1.b	Y47D3A.6b	tra-1b_F	ATG GCC CCC AGT ACT GAG	tra-1b_R	ACC AAA ACC AAC CTC ACA C
skn-1.2a	T19E7.2a	skn-1.2a_F	ATG GGC GGT TCA TCA CG	skn-1.2a_R	AAA CAA TTC CTT CCG ACA GAG
skn-1.2c	T19E7.2c.4	skn-1.2c_F	TGA CCG AGA TGC AAG AGA TG	skn-1.2c_R	TCA GAT GTA ATG GGA CAT CTT G

APPENDIX G

Genotyping Primers

Gene	Gene #	Strain	Forward primer	Forward primer sequence (5'-3')	Reverse primer	Reverse primer sequence (5'-3')
ztf-16	R08E3.4	ok3028	ZTF-16_IntR	ATG AAA GAC AAA GGT TGC GG	ZTF-16_IntL	CTC TGT TGA CGA TGC TCA CAA