

University of Massachusetts Medical School

eScholarship@UMMS

---

GSBS Dissertations and Theses

Graduate School of Biomedical Sciences

---

2009-04-10

## Delineating the *C. elegans* MicroRNA Regulatory Network: A Dissertation

Natalia Julia Martinez

*University of Massachusetts Medical School*

Let us know how access to this document benefits you.

Follow this and additional works at: [https://escholarship.umassmed.edu/gsbs\\_diss](https://escholarship.umassmed.edu/gsbs_diss)



Part of the [Amino Acids, Peptides, and Proteins Commons](#), [Animal Experimentation and Research Commons](#), [Genetic Phenomena Commons](#), and the [Nucleic Acids, Nucleotides, and Nucleosides Commons](#)

---

### Repository Citation

Martinez NJ. (2009). Delineating the *C. elegans* MicroRNA Regulatory Network: A Dissertation. GSBS Dissertations and Theses. <https://doi.org/10.13028/4rv6-6861>. Retrieved from [https://escholarship.umassmed.edu/gsbs\\_diss/411](https://escholarship.umassmed.edu/gsbs_diss/411)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

**Delineating the *C. elegans* microRNA regulatory network**

A Dissertation Presented

By

**Natalia Julia Martinez**

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

April 10, 2009

Interdisciplinary Graduate Program

# DELINEATING THE *C. ELEGANS* MICRORNA REGULATORY NETWORK

A Dissertation Presented By

Natalia J. Martinez

The signatures of the Dissertation Defense Committee signifies completion and approval as to style and content of the Dissertation

---

A. J. Marian Walhout, Ph.D., Thesis Advisor

---

Kirsten A. Hagstrom, Ph.D., Member of Committee

---

Fabio Piano, Ph.D., Member of Committee

---

Phillip D. Zamore, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written Dissertation meets the requirements of the Dissertation Committee

---

Craig C. Mello, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the School

---

Anthony Carruthers, Ph.D.  
Dean of the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program  
April 10, 2009

## COPYRIGHT INFORMATION

The chapters of this dissertation have appeared in separate publications or as  
part of publications:

Martinez, N. J., Ow, M. C., Barrasa, M. I., Hammell, M., Sequerra, R., Doucette-Stamm, L., Roth, F. P., Ambros, V., and Walhout, A. J. 2008. A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev* **22**: 2535-49.

Martinez, N. J.\*, Ow, M. C.\*, Reece-Hoyes, J. S., Barrasa, M. I., Ambros, V. R., and Walhout, A. J. 2008. Genome-scale spatiotemporal analysis of *Caenorhabditis elegans* microRNA promoter activity. *Genome Res* **18**: 2005-15.  
\*These authors contributed equally to this work

Ow, M. C., Martinez, N. J., Olsen, P. H., Silverman, H. S., Barrasa, M. I., Conradt, B., Walhout, A. J., and Ambros, V. 2008. The FLYWCH transcription factors FLH-1, FLH-2, and FLH-3 repress embryonic expression of microRNA genes in *C. elegans*. *Genes Dev* **22**: 2520-34.

Martinez, N. J., and Walhout, A. J. 2009. The interplay between transcription factors and microRNAs in genome-scale regulatory networks. *Bioessays* **31**: 435-445.

## **DEDICATION**

A mis amigos, los de lejos y los de cerca. A Mike por hacerme reir todos los días.

Y sobretodo a mi querida familia, la mejor del mundo!!

## ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge my graduate advisor, Marian Walhout, for her support, understanding and dedication and for giving me the chance to work in a very exiting project. She has been instrumental in my development as a scientist and her advice has been invaluable throughout this project.

For over four years, we have maintained a very fruitful and exiting collaboration with Dr. Maria C. Ow and Dr. Victor Ambros. The work presented in this dissertation could have not been possible without their help, insights and hard work. It has been a real pleasure to collaborate with them.

I want to express my immense gratitude for the help, friendship and advice from former and present members of the Walhout lab, especially Bartolito, Efsun, Chris, Vanessa, John and Inma. We have shared memorable science and life discussions, including innumerable laughs. I would also like to extend my appreciation to people at the Program in Gene Function and Expression for creating a great atmosphere to work in and in many cases for sharing their knowledge, equipment and reagents. I would especially like to acknowledge Nina Bhabhalia for her immense generosity.

I would also like to thank the members of my committee, Craig Mello, Phil Zamore and Kirsten Hagstrom for their helpful insights, care, time and guidance throughout the course of my project. I would also like to acknowledge Dr. Fabio Piano for being the external member of my committee.

I could not have probably made the decision to come to the United States if it were not for Dr. Lucio Castilla. I truly thank him for giving me the great opportunity to work in his lab and give me a flavor for what “cutting-edge research” was.

And last but not least, I would like to express my gratitude to all my dear friends for their everlasting friendship and to Mike’s family for always making me feel “at home”. Thank you Mike for making me a better person every day. You are my sunshine, love and inspiration. Finally I would like to thank my family for their unconditional love and for being proud of me no matter what. Your support has been pivotal for me in every aspect of my life.

Thanks to all of you!

## ABSTRACT

Metazoan genomes contain thousands of protein-coding and non-coding RNA genes, most of which are differentially expressed, *i.e.*, at different locations, at different times during development, or in response to environmental signals. Differential gene expression is achieved through complex regulatory networks that are controlled in part by two types of *trans*-regulators: transcription factors (TFs) and microRNAs (miRNAs). TFs bind to *cis*-regulatory DNA elements that are often located in or near their target genes, while microRNAs hybridize to *cis*-regulatory RNA elements mostly located in the 3' untranslated region (3'UTR) of their target mRNAs.

My work in the Walhout lab has centered on understanding how these *trans*-regulators interact with each other in the context of gene regulatory networks to coordinate gene expression at the genome-scale level. Our model organism is the free-living nematode *Caenorahbditis elegans*, which possess approximately 950 predicted TFs and more than 100 miRNAs.

Whereas much attention has focused on finding the protein-coding target genes of both miRNAs and TFs, the transcriptional networks that regulate miRNA expression remain largely unexplored. To this end, we have embarked in the task of mapping the first genome-scale miRNA regulatory network. This network contains experimentally mapped transcriptional TF=>miRNA interactions, as well as computationally predicted post-transcriptional miRNA=>TF interactions. The work presented here, along with data reported by other groups, have revealed



the existence of reciprocal regulation between these two types of regulators, as well as extensive coordination in the regulation of shared target genes. Our studies have also identified common mechanisms by which miRNAs and TFs function to control gene expression and have suggested an inherent difference in the network properties of both types of regulators.

Reverse genetic approaches have been extensively used to delineate the biological function of protein-coding genes. For instance, genome-wide RNAi screens have revealed critical roles for TFs in *C. elegans* development and physiology. However, reverse genetic approaches have not been very insightful in the case of non-coding genes: A single null mutation does not result in an easily detectable phenotype for most *C. elegans* miRNA genes. To help delineate the biological function of miRNAs we sought to determine when and where they are expressed. Specifically, we generated a collection of transgenic *C. elegans* strains, each containing a miRNA promoter::*gfp* (*Pmir>::gfp*) fusion construct. The particular pattern of expression of each miRNA gene should help to identify potential genetic interactors that exhibit similar expression patterns, and to design experiments to test the phenotypes of miRNA mutants.

## TABLE OF CONTENTS

<b>TITLE PAGE</b>	i
<b>SIGNATURE PAGE</b>	ii
<b>COPYRIGHT PAGE</b>	iii
<b>DEDICATION</b>	iv
<b>ACKNOWLEDGEMENTS</b>	v
<b>ABSTRACT</b>	vii
<b>TABLE OF CONTENTS</b>	ix
<b>LIST OF FIGURES</b>	xiii
<b>LIST OF TABLES</b>	xv
<b>PREFACE TO CHAPTER I</b>	1
<b>CHAPTER I: The interplay between transcription factors and microRNAs in genome-scale regulatory networks: An introduction.</b>	2
<b>Introduction</b>	3
<b>Overview of the principles of TF and miRNA-mediated gene regulation</b>	5
<i>Transcription factors</i>	5
<i>MicroRNAs</i>	7
<i>Cis-regulatory elements</i>	8
<i>Modulation of TF and miRNA activity</i>	10
<i>Impact on animal development and physiology</i>	12
<b>Integrating TFs and miRNAs into genome-scale gene regulatory networks</b>	13
<i>Gene regulatory networks</i>	14
<i>Mapping regulatory networks involving TFs and miRNAs</i>	15
<i>Network motifs</i>	18
<i>MiRNA and TF-containing circuits in regulatory networks</i>	19
<b>Synopsis</b>	22
<b>PREFACE TO CHAPTER II</b>	27

<b>CHAPTER II: A <i>C. elegans</i> genome-scale microRNA network contains composite feedback motifs with high flux capacity</b>	28
<b>Abstract</b>	29
<b>Introduction</b>	30
<b>Results</b>	34
<i>A genome-scale C. elegans TF=&gt;miRNA transcription regulatory network</i>	34
<i>A C. elegans miRNA=&gt;TF post-transcription regulatory network</i>	39
<i>MiRNAs and TFs function together in composite feedback loops</i>	41
<i>MiRNA&lt;=&gt;TF composite feedback loops form higher order network structures</i>	43
<i>Composite miRNA&lt;=&gt;TF feedback loops constitute a network motif</i>	44
<i>MiRNAs and TFs in composite feedback loops provide a high information flow</i>	45
<b>Discussion</b>	48
<b>Materials and methods</b>	54
<b>Acknowledgments</b>	64
<b>PREFACE TO CHAPTER III</b>	100
<b>CHAPTER III: Genome-scale spatiotemporal analysis of <i>Caenorhabditis elegans</i> microRNA promoter activity</b>	101
<b>Abstract</b>	102
<b>Introduction</b>	103
<b>Results</b>	107
<i>Generation of transgenic PmiRNA::gfp C. elegans strains</i>	110
<i>Characterization of miRNA expression patterns</i>	108
<i>Temporal PmiRNA::gfp activity correlates with Northern blot analysis</i>	110
<i>Post-transcriptional mechanisms contribute to differential miRNA expression</i>	112
<i>The promoters of miRNAs are active later in development</i>	114

<i>Most miRNA promoters drive GFP expression in a tissue-specific manner</i>	115
<i>Members of miRNA families can be expressed in distinct or overlapping patterns</i>	116
<i>Intragenic miRNAs</i>	118
<b>Discussion</b>	120
<i>Note added in proof</i>	122
<b>Methods</b>	124
<b>Acknowledgments</b>	130
<b>PREFACE TO CHAPTER IV</b>	158
<b>CHAPTER IV: The FLYWCH transcription factors FLH-1, FLH-2 and FLH-3 repress embryonic expression of microRNA genes in <i>C. elegans</i></b>	159
<b>Abstract</b>	160
<b>Introduction</b>	161
<b>Results</b>	164
<i>FLH-1 binds to an upstream region of lin-4</i>	164
<i>The FLH-1-binding fragment in the lin-4 promoter is essential for repression of lin-4 in the embryo</i>	164
<i>RNAi of flh genes results in precocious embryonic expression of lin-4</i>	165
<i>Isolation and characterization of deletion mutations in the flh genes</i>	166
<i>Expression pattern of FLH transcription factors</i>	168
<i>Precocious expression of lin-4 in flh mutants reduces LIN-14 levels in embryos</i>	169
<i>Elevated lin-4 levels in flh mutant embryos do not result in post-embryonic heterochronic defects</i>	170
<i>In addition to lin-4, FLH proteins can regulate the levels of other miRNAs</i>	172
<i>Genome-scale Y1H screens reveal additional interactions between</i>	

<i>miRNA promoters and FLH proteins</i>	174
<i>Identification of an FLH-1 consensus binding site</i>	175
<i>FLH-2 can also interact with the FLH-1 consensus binding site</i>	177
<i>FLH-1 represses the expression of flh-2</i>	178
<i>FLYWCH proteins engage in protein-protein interaction</i>	179
<i>FLH transcription factors likely regulate multiple targets</i>	180
<b>Discussion</b>	181
<b>Materials and methods</b>	186
<b>Acknowledgments</b>	195
<b>PREFACE TO CHAPTER V</b>	222
<b>CHAPTER V: Discussion</b>	223
<b>MiRNA networks: Lessons from genome-scale studies</b>	224
<i>Global properties of gene regulatory networks involving miRNAs and TFs</i>	224
<i>Regulatory circuits containing TFs and miRNAs</i>	225
<b>Genome-scale analysis of miRNA promoter activity: Lessons and applications</b>	229
<i>Spatial and temporal analysis of miRNA promoter activity</i>	230
<i>Independent transcription of intragenic miRNAs</i>	231
<i>Conservation of miRNA expression patterns</i>	232
<i>Phenotypic assays</i>	233
<i>Combinatorial action of miRNAs</i>	234
<b>Case study: FLH transcription factors function together to repress miRNA expression in the embryo</b>	235
<b>Future challenges: Complete, dynamic and integrated networks</b>	237
<b>REFERENCES</b>	244

## LIST OF FIGURES

<b>Figure I-1.</b> <i>TF and miRNA-containing regulatory networks.</i>	25
<b>Figure I-2.</b> <i>Coherent and incoherent feedback and feed-forward motifs.</i>	26
<b>Figure II-1.</b> <i>A genome-scale C. elegans TF=&gt;miRNA transcription regulatory network.</i>	81
<b>Figure II-2.</b> <i>A C. elegans miRNA=&gt;TF post-transcriptional regulatory network.</i>	84
<b>Figure II-3.</b> <i>The mir-43&lt;=&gt;LIN-26 composite feedback loop.</i>	90
<b>Figure II-4.</b> <i>Additional LIN-26 data.</i>	92
<b>Figure II-5.</b> <i>DAF-3 regulates the let-7 family of miRNAs.</i>	93
<b>Figure II-6.</b> <i>TFs and miRNAs in composite feedback loops are characterized by a high flux capacity (Fc).</i>	95
<b>Figure II-7.</b> <i>A high flux capacity correlates with composite feedback loops.</i>	97
<b>Figure II-8.</b> <i>Model for the function of feedback loops in gene expression programs.</i>	99
<b>Figure III-1.</b> <i>MiRNA promoters and expression rates.</i>	131
<b>Figure III-2.</b> <i>Temporal PmiRNA::gfp activity correlates with Northern blot analysis and uncovers possible post-transcriptional mechanisms that control miRNA expression.</i>	148
<b>Figure III-3.</b> <i>Temporal miRNA promoter activity.</i>	150
<b>Figure III-4.</b> <i>Spatial miRNA promoter activity.</i>	152
<b>Figure III-5.</b> <i>MiRNAs from a given family can have overlapping as well as different spatiotemporal expression patterns.</i>	155
<b>Figure III-6.</b> <i>Upstream sequences of intragenic miRNAs can drive GFP expression in vivo.</i>	157
<b>Figure IV-1.</b> <i>FLH-1 is required for lin-4 repression in embryos.</i>	196
<b>Figure IV-2.</b> <i>Characterization of FLYWCH mutants.</i>	198
<b>Figure IV-3.</b> <i>Double flh mutants display severe phenotypes compared to single mutants.</i>	200

<b>Figure IV-4.</b> <i>Expression pattern of FLYWCH TFs.</i>	202
<b>Figure IV-5.</b> <i>flh mutants show upregulated levels of LIN-14 protein in embryos but no post-embryonic heterochronic defects.</i>	204
<b>Figure IV-6.</b> <i>MiRNAs that change significantly in flh mutants compared to wild-type embryos.</i>	208
<b>Figure IV-7.</b> <i>Y1H assays identify additional miRNA targets.</i>	211
<b>Figure IV-8.</b> <i>Identification of an FLH-1 consensus binding site.</i>	214
<b>Figure IV-9.</b> <i>Distribution of the FLH-1 consensus binding site.</i>	216
<b>Figure IV-10.</b> <i>FLH-1 and FLH-2 can bind the same consensus binding site.</i>	217
<b>Figure IV-11.</b> <i>FLH-1 represses flh-2 expression.</i>	218
<b>Figure IV-12.</b> <i>FLYWCH TFs engage in protein-protein interactions.</i>	219
<b>Figure IV-13.</b> <i>Working model: FLYWCH TFs regulate miRNA expression.</i>	220
<b>Figure V-1.</b> <i>Summary of presence of hub nodes in transcriptional and post-transcriptional networks.</i>	240
<b>Figure V-2.</b> <i>Network circuits allow the spreading of regulatory effects.</i>	241
<b>Figure V-3.</b> <i>A putative FLH-miRNA regulatory circuit.</i>	242
<b>Figure V-4.</b> <i>Integration of functional data into “meta network models”.</i>	243

## LIST OF TABLES

<b>Table I-1.</b> <i>Overview of TF and miRNA-mediated gene regulation.</i>	24
<b>Table II-1.</b> <i>Summary of miRNA gene location and promoters analyzed by Y1H.</i>	65
<b>Table II-2.</b> <i>Promoter details and primer sequences.</i>	66
<b>Table II-3.</b> <i>Preys used in Y1H matrix experiments.</i>	70
<b>Table II-4.</b> <i>High-confidence TF=&gt;PmiRNA interactions (score <math>\geq 5</math>).</i>	73
<b>Table II-5.</b> <i>TF=&gt;Pmir-788 interactions not included in network analyses.</i>	83
<b>Table II-6.</b> <i>MiRNA target predictions for TF genes found in Y1H assays and common in two or more prediction algorithms.</i>	86
<b>Table II-7.</b> <i>List of composite miRNA&lt;=&gt;TF feedback loops.</i>	89
<b>Table II-8.</b> <i>Network motif analysis.</i>	94
<b>Table II-9.</b> <i>Separation between nodes involved and not involved in composite feedback loops according to different cutoffs of <math>k_{in}</math>, <math>k_{out}</math> and <math>F_c</math>.</i>	96
<b>Table II-10.</b> <i>Introduction of 10% false negative TF=&gt;miRNA interactions does not affect the enrichment of miRNA&lt;=&gt;TF feedback lops.</i>	98
<b>Table III-1.</b> <i>Promoter details and primer sequences.</i>	132
<b>Table III-2.</b> <i>Description of PmiRNA::gfp expression patterns.</i>	139
<b>Table III-3.</b> <i>Expression profile of each PmiRNA::gfp line using a binary code.</i>	144
<b>Table III-4.</b> <i>Temporal expression available of each miRNA.</i>	146
<b>Table III-5.</b> <i>Explanation of the spatiotemporal annotation scheme.</i>	154
<b>Table IV-1.</b> <i>Expression of the col-19 heterochronic marker in wild-type and flh mutants.</i>	206
<b>Table IV-2.</b> <i>Number of seam cells of the V lineage observed in wild-type and flh mutants.</i>	207
<b>Table IV-3.</b> <i>Fold change values of miRNAs that changed significantly in flh mutants compared to wild-type embryos.</i>	210
<b>Table IV-4.</b> <i>Comparison of various analyses used in this study.</i>	213



**Table IV-5.** *List of C. elegans strains used in this study.*

221

## PREFACE TO CHAPTER I

This chapter provides a concise introduction to the main principles of TF- and miRNA-mediated gene regulation. It also provides a brief introduction to the study of gene regulatory networks, specifically those involving TFs and miRNAs. Finally, this chapter presents an outline of the main questions and research described in the following chapters of this thesis.

Part of this chapter has been published separately in:

Martinez, N. J. and Walhout, A. J. 2009. The interplay between transcription factors and microRNAs in genome-scale regulatory networks. *Bioessays* **31**: 435-445.

## CHAPTER I

**The interplay between transcription factors and microRNAs in genome-scale regulatory networks: An introduction.**

## Introduction

Over the last decade, the sequencing of a vast number of genomes revealed that an increase in organismal complexity is not merely explained by a dramatic increase in the number of protein-coding genes. Indeed, highly complex organisms frequently contain roughly the same number of protein-coding genes as organisms with less intricate morphology and behaviors. For example, the nematode *Caenorhabditis elegans* has ~20,000 predicted protein-coding genes with a relatively simple body plan of less than 1000 somatic cells (Consortium 1998). The fruit fly *Drosophila melanogaster* and humans have a much more complex anatomy and physiology than worms, yet their genomes encode only ~14,000 and ~25,000 predicted protein-coding genes, respectively (Adams et al. 2000; Consortium 2002).

It has been proposed that organismal complexity developed from a gradual increase in protein diversity, due mainly to alternative mRNA splicing, combined with a gradual increase in the extent and intricacy of gene regulation (Maniatis and Tasic 2002) (Levine and Tjian 2003; Mattick 2004). For instance, the human genome is 3.2 Gb in length, whereas *C. elegans* has a genome of only 100 Mb. Since exon and ORF length does not increase with animal complexity, this means that the non-coding part of the human genome can be up to 30 times larger than that of *C. elegans*. In addition to an increase in regulatory genomic space, there is also an increase in the number of *trans*-regulators. First, the proportion of proteins that encode TFs increases with organismal complexity;

around 5% of the protein-coding genes code for TFs in flies and nematodes, compared to almost 10% for mouse and human (Kummerfeld and Teichmann 2006; Reece-Hoyes et al. 2005; Wilson et al. 2008) (Table I-1). Second, the number of miRNAs encoded by a genome appears to correlate with organismal complexity as well (Grimson et al. 2008). For example, 154, 337 and 695 miRNAs have been annotated to date in the *C. elegans*, zebrafish *Danio rerio* and human genomes, respectively (miRBase; Griffiths-Jones et al. 2006).

Both TFs and miRNAs can exert a widespread impact on gene expression. Most, if not all, genes in the genome are controlled by TFs, which either up- or downregulate transcription. Overall, miRNAs are predicted to target approximately 10-30% of animal protein-coding genes with each miRNA repressing on average 200 transcripts (Brennecke et al. 2005; Krek et al. 2005; Lall et al. 2006; Lewis et al. 2005).

Hierarchically, miRNAs function downstream of TFs since miRNAs can repress an mRNA only after it has been transcribed. However, recent observations suggest that transcriptional regulation by TFs and post-transcriptional regulation by miRNAs are often highly coordinated. To gain understanding of the coordinated effects of TFs and miRNAs it is critical to delineate and characterize the genome-scale regulatory networks in which these regulators operate. Such networks combine the plethora of regulatory circuits for a tissue, organism, or process of interest, usually into a single model. Analyses of these models provide insights into the mechanisms that control gene

expression at a systems level, rather than at the level of (an) individual gene(s). This chapter briefly describes the main principles of TF- and miRNA-mediated gene regulation, concentrating primarily on animal systems. It also provides a brief introduction to the study of gene regulatory networks involving TFs and miRNAs. Finally, this chapter presents a synopsis of the questions and research described in this thesis.

### **Overview of the principles of TF and miRNA-mediated gene regulation**

A summary of the differences and similarities between TFs and miRNA-mediated gene regulation is provided in Table I-1.

#### *Transcription factors*

TFs are modular proteins that often contain separate domains that participate in DNA binding, protein–protein interactions, and transcriptional activation or repression. TFs work largely by interacting with the basal transcriptional machinery and/or chromatin modifying proteins, thereby altering the rate of gene transcription [for review see: (Latchman 1998; Lee and Young 2000)]. TFs physically interact with genomic *cis*-regulatory DNA elements, referred to as TF binding sites (TFBSs), through a specific DNA binding domain. Numerous types of DNA binding domains have been identified in eukaryotes and known TFs that contain similar DNA binding domains are grouped into families. By compiling the protein sequence information of known members of a TF family,

additional members can be computationally predicted based on protein sequence similarities (Kummerfeld and Teichmann 2006; Reece-Hoyes et al. 2005). Since members of TF families by definition have similar DNA binding domains, their TFBSs can often be highly similar as well. For instance, the 84 homeodomain TFs of *D. melanogaster* can be divided into eleven groups based on their DNA binding specificity, with all the members of a group binding to highly similar DNA sequences (Noyes et al. 2008). However, members of other families, notably C2H2 zinc fingers, can bind a large variety of DNA sequences (Wolfe et al. 2000). An important question in the field of systems biology is how members of TF families attain functional specificity in evolution. It is also important to note that members of a TF family have been reported to bind overlapping sets of target genes and have (partially) redundant functions in biological processes such as development. For instance, a study in human T cell lines showed that three members of the ETS family of TFs frequently bind target genes in a redundant manner (Hollenhorst et al. 2007). Similarly, multiple members of the FLYWCH family of *C. elegans* TFs bind to and repress overlapping sets of miRNAs in the embryo (Ow et al. 2008).

For most organisms, only a handful of TFs have been studied in detail. For instance, even in large efforts such as the ENCODE project, only a limited number of TFs have been analyzed by chromatin immunoprecipitation (ChIP) and other methods, and these assays have been conducted only in a small number of human cell lines (Consortium 2007). Similarly, only ~200 *C. elegans*

promoters have so far been used as DNA baits in yeast one-hybrid (Y1H) assays. Although this already led to the identification of targets for ~25% of all predicted *C. elegans* TFs, these studies are far from complete (Deplancke et al. 2006; Martinez et al. 2008a; Vermeirssen et al. 2007). Extending both types of studies to the genome- and proteome-scale level will continue to uncover more targets for most TFs. There is also accumulating evidence that not all DNA binding domains, and therefore all TFs, have yet been identified. For instance, Snyder and colleagues have found that the yeast enzyme Arg5,6 specifically binds DNA (Hall et al. 2004). Similarly, by using Y1H assays we have retrieved more than 20 *C. elegans* proteins that robustly interact with gene promoters but do not possess a recognizable DNA binding domain (Deplancke et al. 2006; Martinez et al. 2008a; Vermeirssen et al. 2007). Further computational and experimental studies are needed to obtain more comprehensive predictions of TFs.

### *MicroRNAs*

MiRNAs are short non-coding RNAs that are typically transcribed by RNA polymerase II and further processed in a step-wise manner by a common biogenic pathway [reviewed in: (Kim and Nam 2006)]. MiRNAs have been identified in numerous organisms by a combination of experimental and computational strategies (Berezikov et al. 2005; Lagos-Quintana et al. 2001; Lai et al. 2003; Lau et al. 2001; Lee et al. 1993; Lim et al. 2003). For most genomes,



the repertoire of miRNAs is only partially known although efforts such as high-throughput sequencing are currently underway to identify all miRNAs in the organism of choice (Grimson et al. 2008; Landgraf et al. 2007; Ruby et al. 2006; Ruby et al. 2007).

MiRNAs hybridize to complementary *cis*-regulatory elements usually located in the 3'UTR of target mRNAs [for review see: (Filipowicz et al. 2008)]. In animal systems, this interaction leads to translational repression and/or mRNA destabilization. MiRNAs bind mRNAs mainly through an mRNA binding motif, often referred to as the “seed” (see below). MiRNAs share many characteristics with TFs (Table I-1). For instance, they can also be classified into families on the basis of the mRNA binding motif they possess and, to date, hundreds of miRNA families have already been identified (Griffiths-Jones et al. 2006; Hertel et al. 2006). Examples of highly conserved miRNA families include the *let-7* and *mir-1* families that are found in many organisms, including humans. Like TFs, miRNAs from the same family are predicted to share targets and it has been shown that family members can function redundantly to regulate gene expression (Abbott et al. 2005; Miska et al. 2007). For instance, the three related miRNAs *mir-48*, *mir-84* and *mir-241* function redundantly in the control of developmental timing in *C. elegans* (Abbott et al. 2005).

### *Cis-regulatory elements*

TFBSs are short DNA sequences, between 5 and 15 bp long, that can be located proximal to the transcription start site in gene promoters, or can reside in *cis*-regulatory modules, such as enhancers, at more distant genomic locations. TFs do not bind just a single unique DNA sequence; rather they are capable of binding multiple closely related sequences, albeit with different affinities. For these reasons, the identification of functional TFBSs in complex genomes has been extremely challenging. So far the complete spectrum of binding specificities is known only for a handful of TFs. Databases such as TRANSFAC and JASPAR, which collect TFBS information, only contain limited number of TFBSs and only for a fraction of all TFs in any organism (Bryne et al. 2008; Wingender et al. 2001). Two recent efforts experimentally determined the binding specificities of most mouse and *D. melanogaster* homeodomains by protein-binding microarrays and bacterial one-hybrid assays (Berger et al. 2008; Noyes et al. 2008). The extension of these types of efforts to other types of TFs will be important to attain comprehensive TFBS information. Subsequently, it will be crucial to compare how such *in vitro* binding specificities and affinities relate to *in vivo* binding of full-length TFs to their target genes and in the context of chromatinized DNA.

MiRNAs bind to two types of targets. The first contains sequences with perfect complementarity to the miRNA seed, a stretch of 6 to 8 nucleotides located in the 5' end of the miRNA. The second type possesses sequences with imperfect seed complementarity that is compensated by complementarity at the

3' end of the miRNA (Brennecke et al. 2005). So far, it appears that most detectable target genes belong to the first type [for review see: (Rajewski 2006)]. Efforts to study the interaction of miRNAs with their *cis*-regulatory elements have been restricted to the 3'UTR of target mRNAs. The length of 3'UTRs increases with organismal complexity, although even in humans the average 3'UTR length is less than 1kb (Mazumder et al. 2003; Retelska et al. 2006). Interestingly, it has been shown experimentally that miRNAs can associate with artificial sites located in any position on a target mRNA (Lytle et al. 2007). In addition, it has been recently demonstrated that functional *cis*-regulatory elements can be located in the coding sequences of target mRNAs (Duursma et al. 2008; Tay et al. 2008). Whether this is a general mode of miRNA action or an exception still needs to be determined. However, if it is a general function of miRNAs, the spectrum of putative targets will expand greatly.

#### *Modulation of TF and miRNA activity*

TFs and miRNAs are often differentially expressed during the development, differentiation and homeostasis of cells and tissues (Chang et al. 2004; Johnson et al. 2003; Martinez et al. 2008b; Reece-Hoyes et al. 2007). Differential spatiotemporal expression of TFs is in part regulated by other TFs at the transcriptional, and by miRNAs, RNA-binding proteins (RBPs) and alternative splicing at the post-transcriptional level (Chang et al. 2004; Martinez et al. 2008a; Lillycrop and Latchman 1992; Ow et al. 2008; Rodriguez-Gabriel et al. 2003). At

the post-translational level, TF activity can be modulated by selective dimerization with other TFs, interactions with ligands and co-factors, and by modifications such as phosphorylation, acetylation and ubiquitination [for review see (Grove and Walhout 2008)].

The spatiotemporal expression of miRNAs is in part regulated by TFs that bind miRNA promoters (Johnson et al. 2003; Martinez et al. 2008a; O'Donnell et al. 2005). Recent studies have begun to dissect how miRNA activity may be regulated at the post-transcriptional level by miRNA processing and/or (de)stabilization mechanisms (Obernosterer et al. 2006; Thomson et al. 2006; Viswanathan et al. 2008; Wulczyn et al. 2007). For instance, Viswanathan and colleagues have shown that LIN-28, an RBP, can selectively block processing of the primary *let-7* transcript in mouse embryonic stem cells (Viswanathan et al. 2008), while Kawahara *et al* have shown that miRNAs are subject to extensive RNA editing that converts adenosine to inosine (A to I). Although the mechanism and significance of miRNA editing is unknown, it could potentially alter base pairing specificity (changing the set of target genes) or miRNA stability (Kawahara et al. 2007). Further studies will be needed to determine the extent and variety of post-transcriptional miRNA regulation.

Another way of modulating TF and miRNA activity is to regulate the access to their target sequences. Indeed, the mere presence of a TFBS or a miRNA target sequence is a poor predictor of binding and regulation. For TFs it is widely accepted that histone modifications and resulting changes in chromatin

structure can alter the accessibility to TFBSs. Similarly, the accessibility of miRNAs to their binding sequences can be altered by the presence of RBPs. For example, the expression of Dead end (Dnd1) impairs the function of several miRNAs by blocking the accessibility to target mRNAs in zebrafish primordial germ cells (Kedde et al. 2007).

### *Impact on animal development and physiology*

Genome-wide genetic analyses in many organisms have demonstrated a myriad of critical roles that TFs play in controlling gene expression during development, homeostasis and in disease. For instance, more than 30% of *C. elegans* TFs confer a detectable phenotype when knocked down by RNAi [WormBase WS180, (Kamath et al. 2003)]. In addition, genome-wide RNAi analysis of growth and viability in *D. melanogaster* cells found that proteins with a predicted DNA binding domain comprised the largest category of genes that confer the most severe phenotypes (Boutros et al. 2004). These observations are in agreement with important roles that TFs can play as master regulators of development. For instance, *C. elegans* PHA-4 is necessary and sufficient for development of the pharynx and foregut during embryonic development (Mango et al. 1994). Although mutations in specific miRNAs can produce noticeable phenotypes in several organisms, very few miRNAs have been found in genetic screens. With the exceptions of *lin-4* (Lee et al. 1993), *let-7* (Reinhart et al. 2000), *lsey-6* (Johnston and Hobert 2003) and *mir-1* (Simon et al. 2008), a single

null mutation does not result in an easily detectable phenotype for most *C. elegans* miRNAs (Miska et al. 2007). Furthermore, with only one exception (Bernstein et al. 2003), genetic inactivation of Dicer (the enzyme required for mature miRNA processing) only has relatively modest effects on organism patterning (Harfe et al. 2005; Harris et al. 2006; Cobb et al. 2005; Giraldez et al. 2005; Muljo et al. 2005; Wienholds et al. 2003). Together, these observations appear to suggest that miRNAs make auxiliary, rather than critical contributions to organismal morphology, physiology and behavior, likely by fine-tuning rather than establishing gene expression programs (Hornstein and Shomron 2006).

### **Integrating TFs and miRNAs into genome-scale gene regulatory networks**

In order to elucidate the systems-level mechanisms of gene regulation mediated by TFs and miRNAs, one can combine computational and experimental interaction data into functional network models, and then examine such models for their architecture and organization. Compared to regulatory networks involving protein-coding genes, until recently little was known about genome-scale networks that involve miRNAs. Two types of regulatory networks involving miRNAs and TFs can be distinguished: transcription regulatory networks that consist of protein-DNA interactions between TFs and the *cis*-regulatory DNA regions of their target genes (TF=>target) (Fig. I-1A), and post-transcription regulatory networks that consist of RNA-RNA interactions between miRNAs and the 3'UTR of their target mRNAs (miRNA=>target) (Fig. I-1B). Both types of

networks can be combined into gene regulatory networks (Fig. I-1C). While several efforts have identified post-transcriptional miRNA=>target networks, the transcriptional networks that control miRNA expression (TF=>miRNA) had received little attention.

We will discuss below the general properties of regulatory networks, the methods commonly used to map them and our current knowledge of regulatory circuits involving miRNAs and TFs.

### *Gene regulatory networks*

Regulatory network models are composed of nodes (e.g. TFs, miRNAs and target genes) and edges that describe the relationships between nodes (e.g. binding, activation, repression). The visualization of biological interactions as network graphs allows the investigation of network topology and its correlation to network function by using graph-theoretical concepts or network parameters. One example of a network parameter is “node degree”, defined as the number of interacting partners or number of edges per node. The degree distribution of most biological networks characterized to date is scale-free, which means that the majority of the nodes in the network exhibit relatively low connectivity, whereas a small number of nodes, referred to as network hubs, are extremely highly connected (Jeong et al. 2000). The biological significance of this observation became apparent when it was found that hubs in protein-protein interaction networks are more often essential for survival or development of the

organism than other nodes (Jeong et al. 2001). Transcriptional and post-transcriptional networks are bipartite and directed; they each contain two types of nodes, and the edges between these types of nodes are unidirectional (*i.e.* a TF binds its target gene and not the other way around). Because of this, regulatory networks have two types of degree: in- and out-degree, which reflect, for example, the number of TFs that bind a gene and the number of genes bound by a TF, respectively (Fig. I-1). Whereas the out-degree distribution of TF nodes in transcriptional networks is scale-free, the in-degree distribution of gene nodes is not scale-free (Babu et al. 2004; Deplancke et al. 2006; Vermeirssen et al. 2007). Interestingly, TF hubs tend to be essential for viability whereas target genes bound by many TFs do not have a tendency to be essential for viability. Thus, network analysis provides insights into biological systems that cannot be obtained in single gene studies.

### *Mapping regulatory networks involving TFs and miRNAs*

Transcription regulatory networks consist of protein-DNA interactions between TFs and the *cis*-regulatory DNA regions of their target genes (Fig. I-1A). Genome-wide interactions between TFs and their target DNA sequences can be experimentally mapped using TF-centered as well as gene-centered methods (Deplancke et al. 2006). High-throughput TF-centered methods for protein-DNA interaction mapping include chromatin-immunoprecipitation (ChIP), protein-binding microarrays and bacterial one-hybrid systems among others. ChIP is the



most commonly used method. It is based on the precipitation of a TF of choice and its associated DNA fragments using an anti-TF antibody. Multiple readouts of the precipitated DNA can be used, including tiling microarrays (ChIP-chip) and high-throughput sequencing (ChIP-seq) [for review see (Elnitsky et al. 2006) and (Collas et al. 2008)].

The Gateway-compatible yeast one-hybrid (Y1H) system is a gene-centered approach for the large-scale detection of protein–DNA interactions involving “protein preys” (TFs) and “DNA baits” (promoters and *cis*-regulatory elements) (Deplancke et al. 2006). This Y1H system can be used with gene promoters as DNA baits without prior knowledge about the *cis*-regulatory elements (i.e., TFBSs) that reside within the promoter and with cDNA libraries as source of “protein preys” without previous knowledge of the repertoire of TFs in the system of choice. We have used the Gateway-compatible Y1H system to map TF-miRNA promoter interactions in the nematode *C. elegans* (see Chapter II).

Post-transcription regulatory networks consist of RNA-RNA interactions between miRNAs and the 3'UTR of their target mRNAs (Fig. I-1B). Experimental identification of miRNA-target interactions is not an easy task. To map miRNA-target interactions in a genome-wide scale, computational approaches have been the method of choice. There are many algorithms available to predict miRNA-target interactions. Generally, these algorithms take into account the free energy values between the miRNA/target RNA duplex, the level of complementarity

between the miRNA and the target 3'UTR, conservation of the miRNA site in orthologous 3'UTR species, 3'UTR secondary structure as well as other parameters [for review see (Rajewsky 2006)].

Recently, two miRNA-centered experimental approaches have been used, combined with computational miRNA-target predictions, to determine the genome-wide extent of miRNA control. One approach takes into account that miRNAs cannot only repress the translation of its target mRNA but also induce its degradation, presumably through mRNA-decay pathways [reviewed in (Filipowicz et al. 2008)]. Several groups have performed microarray analysis in cell lines transfected with an ectopic miRNA and identified what messages are reduced in abundance. In addition, microarray analyses have been performed in cell lines, before and after an endogenous miRNA has been knockdown (Knutzfeldt et al. 2005; Farh et al. 2005; Lim et al. 2005). The second approach is a proteomics method that utilizes quantitative mass-spectrometry to measure the protein output upon transfection or knockdown of the miRNA of choice (Baek et al. 2008; Selbach et al. 2008).

TF-DNA and miRNA-target interactions mapped using these high-throughput methods will have false positives and false negatives in each of the methods mentioned above. For instance, some of these approaches detect interactions that do occur *in vivo* (e.g., ChIP with endogenous TFs and miRNA knockdown followed by mass spectrometry) and others find interactions that can occur (e.g., Y1H assays and ectopic miRNA addition followed by mass

spectrometry). TF–DNA interactions that occur in a few cells or during a short period of time in development, will likely be missed by ChIP but may be found by Y1H. Similarly, miRNA-target interactions involving targets that are lowly expressed might result in undetectable changes in protein output in miRNA knockdown experiments but could be detected in ectopic miRNA transfection experiments.

### *Network motifs*

Network models can be analyzed globally but can also provide more local information through the characterization of smaller sub-graphs such as network modules and motifs [reviewed in (Barabasi and Oltvai 2004)]. Network motifs were first identified in transcriptional networks of bacteria and yeast and are defined as small gene regulatory circuits that occur more often in real networks than expected by chance, *i.e.* compared to randomized networks in which the edges have been reassigned computationally (Milo et al. 2002; Shen-Orr et al. 2002). Network motifs provide insights into the mechanisms of gene expression at a systems level. Examples of network motifs include feedback, feed-forward, bi-fans, single and multi-input motifs. Gene regulatory networks of different organisms contain the same types of network motifs suggesting that they not only constitute successful mechanisms of gene expression in one particular organism, but throughout evolution as well.

### *MiRNA and TF-containing circuits in regulatory networks*

A handful of feedback and feed-forward regulatory circuits containing TFs and miRNAs have been recently described in a variety of organisms. Feedback loops can be classified into coherent and incoherent loops (Fig. I-2). In coherent loops (also called double-negative loops) the regulatory paths have the same overall effect (either activation or repression of the target), while in incoherent loops (also called single-negative loops) the regulatory paths have opposite effects (Fig. I-2) (Mangan and Alon 2003). It has been proposed that coherent feedback loops can generate mutually exclusive or bi-stable expression of both the miRNA and TF involved (Johnston et al. 2005), and that a transient signal can cause the loop to be locked into an irreversible state, even after the signal is gone. For example, in human hematopoietic cells, *mir-233* and NFI-A function in a coherent feedback loop to control granulocytic differentiation. In undifferentiated cells, *mir-233* levels are low and NFI/A levels are high, however, upon retinoic acid signaling, *mir-233* levels increase (due to activation by the TF C/EBP $\alpha$ ) and NFI/A is repressed, facilitating differentiation to the myeloid lineage (Fazi et al. 2005) (Fig. I-2). This feedback loop ensures mutually exclusive expression of *mir-233* and NFI/A, thereby generating a bi-stable system (undifferentiated *versus* differentiated hematopoietic cells).

Incoherent feedback loops potentially function to fine-tune gene expression and to maintain precise steady state levels of both components of the

loop (Tsang et al. 2007). Additionally, incoherent loops can result in oscillatory expression of both components, which depends on additional input signals (Hirata et al. 2002). Incoherent feedback loops produce overlapping expression patterns of both components of the loop. We have shown that *C. elegans mir-43* and LIN-26 function together in an incoherent feedback loop where both components of the loop are co-expressed (Fig. I-2) (Martinez et al. 2008a).

In feed-forward loops a regulator X regulates the expression of a target Z via a direct as well as an indirect path, through a regulator Y. Note that not necessarily all targets or regulators are TFs and miRNAs (Fig. I-2). Feed-forward loops can also be coherent or incoherent depending on the overall effect of both regulatory paths (Fig. I-2). Coherent feed-forward loops involving TFs and miRNAs can function to suppress leaky transcription or as toggle switches where an initial signal can be converted into a long lasting cellular response (Alon 2007; Tsang et al. 2007). The c-Myc/ E2F/*miR-17-92* circuit is an example of coherent feed-forward loop. The TF c-Myc, which promotes cell cycle progression, activates the expression of many genes, including the E2F family of TFs and the *mir-17-92* cluster. In addition, E2F family members can bind the promoter of the *mir-17-92* cluster, activating its transcription. This coherent feed-forward loop is embedded in a more complex circuit, since some of the miRNAs from the *mir-17-92* cluster negatively regulate E2F family members (Fig. I-2). This feed-forward loop reveals a mechanism through which c-Myc simultaneously activates E2F transcription and limits its translation, through activation of *mir-17-92* cluster,

allowing a tightly controlled proliferative signal (O'Donnell et al. 2005; Sylvestre et al. 2007; Woods et al. 2007). Incoherent feed-forward loops have been proposed to provide response acceleration upon signal detection (Alon 2007). For instance, Marson and colleagues have reported that the core embryonic stem cell (ESC) TFs, Oct-4, Nanog, Sox-2 and Tcf3, promote the transcription of the miRNA *let-7* and the RBP Lin-28. Mature *let-7* is scarce in ESCs but abundant in differentiated cells. Upon signal detection, Lin-28-mediated inhibition is released and mature *let-7* starts to accumulate. This circuit poises ESCs for rapid and efficient cellular differentiation (Fig. I-2) (Marson et al. 2008).

The mapping and analysis of genome-wide gene regulatory networks involving TFs and miRNAs will be necessary to determine if regulatory circuits containing miRNAs and TF, in particular feedback and feed-forward loops, are isolated instances (like the examples mentioned above) or constitute more general mechanism of gene expression at the organismal, or systems level.

## Synopsis

Although feedback mechanisms are prominent throughout biology, Alon and colleagues found that feedback loops are not overrepresented in pure transcriptional networks (networks composed solely of interactions between TFs and target genes) (Shen-Orr et al. 2002). Based on this observation, Margalit and colleagues proposed that feedback loops could be identified when transcriptional interactions are combined with protein-protein interactions (Yeager-Lotem et al. 2004). Several examples of feedback loops involving transcriptional interactions and miRNA-mediated post-transcriptional interactions have been either postulated (Tsang et al. 2007), or demonstrated (see above) (Fazi et al. 2005; Johnston et al. 2005; Kim et al. 2007; Varghese and Cohen 2007). Based on these observations, we hypothesized that miRNAs may, together with TFs, be involved in general mechanisms of feedback regulation in metazoan regulatory networks. In Chapter II of this thesis we describe the mapping of the first genome-scale TF=>miRNA transcription regulatory network in any organism. By combining transcriptional and post-transcriptional interactions, we show that miRNAs are a missing link to form feedback motifs in the *C. elegans* gene regulatory network (Martinez et al. 2008a).

Disruption of gene expression is often the first step in identifying the putative function of a gene during organismal development or in particular pathway or process. Genome-wide RNAi analyses in *C. elegans* and many other organisms have demonstrated a myriad of critical roles that TFs play in

controlling gene expression during development and physiology. However, in the case of miRNAs, reverse genetic analyses have not been very insightful. With the exceptions of a few miRNAs, a single null mutation does not result in an easily detectable phenotype for most *C. elegans* miRNA genes (Miska et al. 2007). Hence, one initial approach that will help to delineate the biological function of *C. elegans* miRNAs is by determining when and where they are expressed. In Chapter III we describe the generation of a resource to study *in vivo* miRNA expression in the nematode *C. elegans*. Specifically, we have generated a collection of transgenic strains expressing GFP under the control of miRNA promoters. Through the global analysis of this resource, we provide insights into miRNA function and we present evidence that post-transcriptional regulation of pri-miRNAs provides an additional layer of differential miRNA expression in nematodes.

The research presented in Chapter IV provides an example of the possible uses and applications of the miRNA transcription regulatory network and the miRNA expression pattern resource described in Chapter II and III, respectively. Specifically, we describe an in depth analysis of a set of TF=>miRNA interactions involving several miRNAs and a poorly characterized family of transcription factors containing a FLYWCH Zn-finger DNA-binding domain. Our data suggests that FLYWCH transcription factors repress the expression of miRNAs in the *C. elegans* embryo.



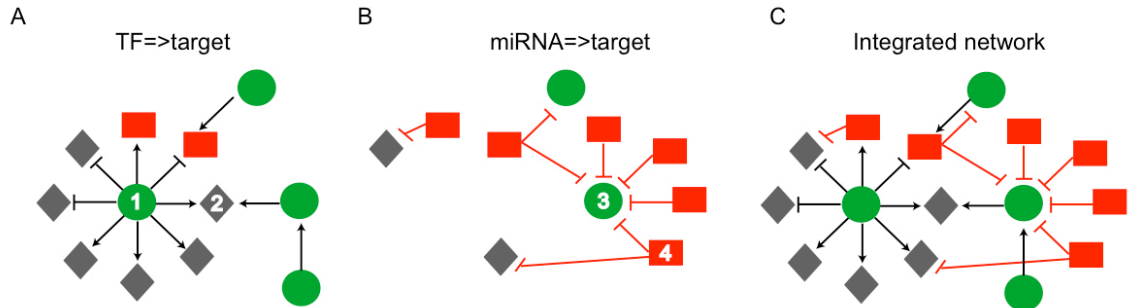
**Table I-1. Overview of TF and miRNA-mediated gene regulation.**

Table 1. Overview of TF and miRNA-mediated gene regulation

	<b>TFs</b>	<b>miRNAs</b>
Abundance in the genome	~5-10%	~1%-5%
<i>De novo</i> identification	Computational	Computational, experimental (high-throughput sequencing)
Classification into families	Yes - based on DNA binding domain	Yes - based on seed sequence
Mode of action	Activation and repression	Mostly repression
Target region	Promoters, <i>cis</i> -regulatory modules	Mostly 3' UTRs, coding regions?
Search space for targets	Tens or hundreds of kilobases	1 or a few kilobases
Mode of binding	Highly plastic - sites can be highly variant	Stringent, dictated by basepairing of seed sequence
Additional variants	Extensive - Created by splicing, PTM, dimerization	Not identified - Possibly by editing?
Biological contributions	Often critical in development and physiology	Mostly not critical in development and physiology (fine-tuning)

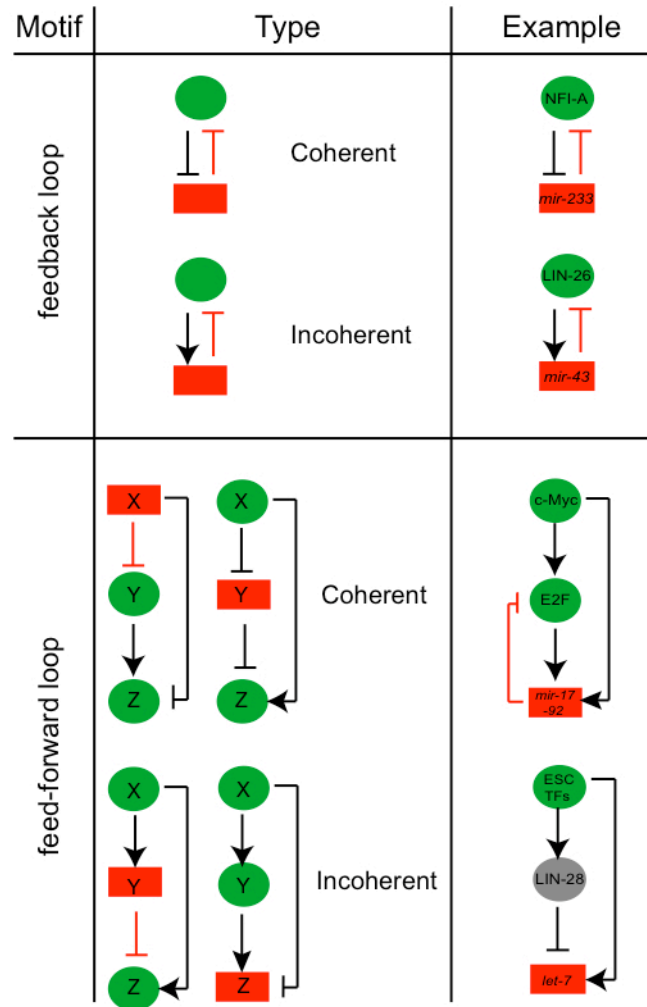
PTM: post-translational modifications

Figure I-1



**Figure I-1.** *TF and miRNA-containing regulatory networks.* (A) Transcriptional network (TF=>target). (B) Post-transcriptional network (miRNA=>target). (C) Integrated gene regulatory network. Nodes: green circles depict TFs; red rectangles depict miRNAs; grey diamonds represent other protein-coding genes. Edges: black arrows depict transcriptional activation; black blunted arrows depict transcriptional repression; red blunted arrows represent post-transcriptional repression. Examples of in- and out-degrees: TF #1 has an out-degree of 8; Target gene #2 has an in-degree of 2; Target gene #3 (in this case a TF) has an in-degree of 5; miRNA #4 has an out-degree of 2.

Figure I-2



**Figure I-2.** *Coherent and incoherent feedback and feed-forward motifs.* Note that for feed-forward loops, other arrangements between regulators and targets are possible. Green circles – TFs; red rectangles – miRNAs; grey diamonds – other protein-coding genes; black arrows – activation; black blunted arrows – repression. Lin-28 is an RNA binding protein – grey circle. ESC TFs: embryonic stem cell TFs Nanog, Oct-4, Tcf3 and Sox-2.

## PREFACE TO CHAPTER II

The work presented in the following chapter describes the first experimentally mapped genome-scale miRNA transcription regulatory network in any organism and provides evidence for the existence of composite miRNA $\leftrightarrow$ TF feedback loops as common mechanism of gene expression. This work is part of a collaboration with the Ambros lab and embodies the joint effort of several people: M. C. Ow contributed to TaqMan PCR data presented in Figures II-1, II-3 and II-4; M. I. Barrasa contributed to network randomizations presented in Figures II-7 and Tables II-8 and II-10; F. P. Roth defined the network parameter “Flux capacity”; M. I. Barrasa and M. Hammell retrieved miRNA-target predictions in Figure II-2 and Table II-6 and myself (Figures II-1 through II-8 and remaining tables). A.J. Walhout and myself wrote the manuscript.

This chapter has been published separately in:

Martinez, N. J., Ow, M. C., Barrasa, M. I., Hammell, M., Sequerra, R., Doucette-Stamm, L., Roth, F. P., Ambros, V., and Walhout, A. J. 2008. A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev* **22**: 2535-49

## CHAPTER II

**A *C. elegans* genome-scale microRNA network contains composite  
feedback motifs with high flux capacity**

**Abstract**

MicroRNAs (miRNAs) and transcription factors (TFs) are primary metazoan gene regulators. Whereas much attention has focused on finding the targets of both miRNAs and TFs, the transcriptional networks that regulate miRNA expression remain largely unexplored. Here, we present the first genome-scale *Caenorhabditis elegans* miRNA regulatory network that contains experimentally mapped transcriptional TF=>miRNA interactions, as well as computationally predicted post-transcriptional miRNA=>TF interactions. We find that this integrated miRNA network contains 23 miRNA<=>TF composite feedback loops in which a TF that controls a miRNA is itself regulated by that same miRNA. By rigorous network randomizations, we show that such loops occur more frequently than expected by chance and, hence, constitute a genuine network motif. Interestingly, miRNAs and TFs in such loops are heavily regulated and regulate many targets. This 'high flux capacity' suggests that loops provide a mechanism of high information flow for the coordinate and adaptable control of miRNA and TF target regulons.

## Introduction

Differential gene expression can be regulated at many levels and by various *trans*-acting factors. MicroRNAs (miRNAs) and transcription factors (TFs) are primary regulators of differential gene expression during organism development, function and in disease. While TFs physically interact with *cis*-regulatory DNA elements to activate or repress transcription of their target genes, miRNAs repress gene expression post-transcriptionally by interacting with complementary sequences located in the 3'UTR of their target mRNAs (Bartel 2004). Following the initial discovery of miRNAs in *C. elegans* (Lee et al. 1993; Wightman et al. 1993), much attention has focused on the identification of their target genes. MiRNA targets are usually predicted computationally, and several algorithms have been developed for this purpose (Sethupathy et al. 2006). Target predictions revealed that miRNAs target TFs more frequently than other types of genes (Shalgi et al. 2007). This suggests that miRNAs and TFs could be intricately connected in the networks that control differential gene expression.

Transcription regulatory networks of protein-coding genes have been mapped and studied in yeast (Harbison et al. 2004), *Caenorhabditis elegans* (Deplancke et al. 2006a; Vermeirssen et al. 2007a), *Drosophila melanogaster* (Sandmann et al. 2007) and mammals (Boyer et al. 2005). While vastly incomplete, these networks have already provided insights into overall network architecture and have also revealed particular network sub-graphs that are overrepresented in real networks compared to randomized networks. Such

enriched sub-graphs are referred to as network motifs (Milo et al. 2002; Shen-Orr et al. 2002). Since network motifs are recurrent regulatory circuits, they provide successful mechanisms of gene expression and, hence, play widespread roles in gene regulation. For instance, feed-forward loops provide a mechanism to ensure a robust transcriptional response to signals (e.g. environmental signals), and to protect against transcriptional noise (Milo et al. 2002; Shen-Orr et al. 2002). Feedback loops are important in homeostasis and cellular differentiation programs (Alon 2007). Surprisingly, whereas feed-forward loops are overrepresented in pure transcription regulatory networks, feedback loops were found to be less abundant. One explanation for the paucity of feedback loops is that they may be generated by a combination of transcriptional and post-transcriptional mechanisms, as opposed to being purely transcriptional (Shen-Orr et al. 2002; Yeger-Lotem et al. 2004). Interestingly, recent bioinformatic studies found that the expression of miRNAs and their targets is often highly correlated or anti-correlated (Farh et al. 2005; Stark et al. 2005; Sood et al. 2006; Tsang et al. 2007). Tsang et al (2007) proposed that such (anti)-correlations can result from various types of feed-forward and feedback loops involving miRNAs, their predicted target genes and upstream regulators (e.g. TFs, kinases). This study was exclusively based on a predicted miRNA=>target network and did not use TF=>miRNA transcriptional interactions because they were not available. Consequently, no *actual* loops were identified. However, several feedback loops involving miRNAs and TFs have recently been found experimentally in a variety



of organisms (Fazi et al. 2005; Johnston et al. 2005; Kim et al. 2007; Varghese and Cohen 2007). Thus, we hypothesized that miRNAs may be a 'missing post-transcriptional link' that, together with TFs generates feedback loops in genome-scale gene regulatory networks.

MiRNAs are transcribed as part of longer primary transcripts (pri-miRNAs), which are then processed in a stepwise manner by protein complexes that include the RNase III enzymes Drosha, to produce pre-miRNAs, and Dicer to produce mature 21-22 nucleotide miRNAs [for review, see (Kim 2005)]. MiRNAs are transcribed by RNA polymerase II (Lee et al. 2004), which suggests that miRNA transcription is subject to similar control mechanisms as protein-coding genes. Although some TFs that regulate miRNA expression have been found (Fazi et al. 2005; Sylvestre et al. 2007), genome-scale transcriptional networks that control miRNA expression remain unexplored. Transcription regulatory networks have predominantly been mapped using protein-DNA interaction mapping methods that are TF-centered such as chromatin immunoprecipitation (ChIP)(Harbison et al. 2004). ChIP is impractical for the comprehensive mapping of TFs that regulate miRNA expression because it would require the testing of all predicted TFs of an organism in multiple tissue types and under different conditions. Moreover, many TFs are not sufficiently broadly or highly expressed to be detected by ChIP, and only few suitable anti-TF antibodies are available (Walhout 2006).

We previously developed and applied a condition-independent yeast one-hybrid (Y1H) method that allows the identification of TFs that can bind a set of promoters of interest (Deplancke et al. 2004; Deplancke et al. 2006a; Vermeirssen et al. 2007a; Vermeirssen et al. 2007b). Here, we use this method to experimentally map a genome-scale TF=>miRNA transcription regulatory network in the nematode *C. elegans*. By integrating this network with a high-confidence miRNA=>TF target network, we identify 23 miRNA<=>TF composite feedback loops. Using rigorous network randomizations, we show that such miRNA<=>TF feedback loops occur more frequently than expected by chance, and, hence, constitute a genuine network motif. We find that most miRNAs and TFs that participate in miRNA<=>TF feedback loops are highly connected: the miRNAs regulate many TFs and are regulated by many TFs, and *vice versa*. We introduce a novel network parameter we named “flux capacity (Fc)” that captures the high flow of information that passes through many miRNAs and TFs involved in composite feedback loops. Finally, we propose a model in which feedback loops provide a mechanism for the highly coordinated and adaptable control of gene batteries, or regulons.

## Results

### *A genome-scale C. elegans TF=>miRNA transcription regulatory network*

The *C. elegans* genome encodes 940 predicted TFs (Reece-Hoyes et al. 2005; Vermeirssen et al. 2007b) and 115 predicted miRNAs (Griffiths-Jones et al. 2006; Ambros et al. 2003; Lim et al. 2003). Of the 115 miRNA genes available in miRBase V4.0, 66 reside in intergenic regions and can be assigned to their own promoter. An additional 16 intergenic miRNAs are transcribed in a total of six intergenic operons, with a single promoter regulating each operon. The remaining 33 miRNAs are embedded within the intron of a protein-coding gene, either in the sense or anti-sense orientation. Thirteen of these intragenic miRNAs are transcribed in the anti-sense orientation, nine of these in two operons. Twenty intragenic miRNAs are located in the sense orientation and are likely co-transcribed with their host gene, and may be controlled by the host gene promoter (Baskerville and Bartel 2005). These latter miRNAs were not included in our analyses.

Although the transcription start site of the majority of *C. elegans* miRNAs has not been mapped, it has been shown that fragments between 1 and 2 kb upstream of the pre-miRNA are sufficient to rescue *lin-4*, *let-7* and *lisy-6* mutant phenotypes (Lee et al. 1993; Johnson et al. 2003; Chang et al. 2004). In addition, reported expression of *mir-84*, *mir-61* and *mir-48* involved 1 to 2.2 kb of genomic sequence upstream of the annotated miRNA (Johnson et al. 2005; Li et al. 2005; Yoo and Greenwald 2005). For these reasons, we have decided to use DNA

fragments that correspond to the intergenic region upstream of the annotated miRNA with a minimum length of 300 bp and a maximum length of 2 kb as miRNA promoters. These fragments may not contain all the regulatory elements necessary for miRNA expression. For instance, Bracht and colleagues have shown that transcription of *let-7* can start either ~200 bp or ~1 kb upstream of pre-*let-7* (Bracht et al. 2004). However, we have recently demonstrated that the vast majority of miRNA promoters as defined here are able to confer GFP expression *in vivo*, and more than 90% of these recapitulate known temporal expression as determined elsewhere by Northern blotting (Martinez et al. 2008). This indicates that the genomic fragment we used indeed encompass miRNA promoters. In total, we cloned 71 miRNA promoters (Table II-1 and Table II-2).

To identify TFs that can interact with each miRNA promoter, we performed three Y1H assays: screens versus a *C. elegans* cDNA library (Walhout et al. 2000b) and a TF mini-library (Deplancke et al. 2004), and matrix assays of all promoters versus all TFs identified in the screens (Vermeirssen et al. 2007a) (Table II-3). Thus, each promoter was directly tested against all TFs in our dataset. We applied a stringent scoring and filtering system to minimize the inclusion of false positives (Vermeirssen et al. 2007a), and obtained 347 high-confidence interactions between 63 miRNA promoters and 116 proteins (Table II-4). These interactions are available in our EDGEDb database (<http://edgedb.umassmed.edu>) (Barrasa et al. 2007). The most highly connected miRNAs belong to the *let-7* and *lin-4* families, implicated in developmental timing,

as well as other miRNAs of unknown function such as *mir-46*, *mir-355* and *mir-243*.

Interestingly, while the majority of the proteins retrieved are predicted TFs (Reece-Hoyes et al. 2005), some do not possess a known DNA binding domain and may constitute novel TFs. We previously demonstrated (by ChIP in yeast) that 9 of 11 tested novel putative TFs (~80%) do interact with their target promoters, suggesting that they may possess an as yet unrecognized DNA binding domain (Deplancke et al. 2006a; Vermeirssen et al. 2007a). The TFs that interact with miRNA promoters are diverse as they represent most of the known TF families in *C. elegans*. The most highly connected TFs include members of the ZF-C2H2 family (DIE-1 and ZTF-1), ZF-NHR (ODR-7), MH-1 (DAF-3), as well as proteins with an unidentified DNA binding domain (Y38C9A.1 and C32D5.1). Together, these observations indicate that there is no DNA binding domain bias in the transcriptional miRNA network (Table II-4 and data not shown).

It is possible that we identified multiple members of a TF family binding to the same promoter in Y1H assays simply because these TFs recognize similar DNA sequences. Alternatively, it may be that members of a TF family function redundantly *in vivo*, as has been shown for mammalian ETS proteins (Hollenhorst et al. 2007). Indeed, we have demonstrated that FLH-1 and FLH-2, members of the FLYWCH family of TFs, redundantly regulate the expression of several miRNAs in the *C. elegans* embryo (Ow et al. 2008).

All high-confidence TF=>miRNA interactions were visualized into a network model (Fig. II-1A). The distribution of both the outgoing connectivity (“out-degree”, or number of miRNA promoters bound by a given TF, Fig. II-1B), and the incoming connectivity (“in-degree”, or number of TFs bound to a given miRNA promoter, Fig. II-1C) of this network are similar to those of *C. elegans* protein-coding gene networks (Deplancke et al. 2006a; Vermeirssen et al. 2007a). This indicates that the overall architecture of miRNA transcription regulatory networks is similar to that of protein-coding gene networks. Hence, at least based on these two network properties, miRNA expression overall is regulated in a similar manner as protein-coding genes.

DAF-3 is a TF that interacted with multiple miRNA promoters (Table II-4). It is involved in dauer formation, a developmentally arrested, alternative third larval stage that occurs under adverse environmental conditions (Patterson et al. 1997). DAF-3 expression increases when worms enter the dauer stage (Wang and Kim 2003). To examine the regulatory effect of DAF-3 on miRNA expression, we compared the levels of 48 miRNAs (see *Materials and methods*) in wild type and *daf-3(mgDf90)* mutant dauer larvae by TaqMan PCR assays, and ranked miRNAs according to their change in expression (Fig. II-1D). Four miRNAs changed significantly in expression levels in *daf-3(mgDf90)* mutants. Three of these were increased in the mutant, and one was slightly decreased (~1.4-fold). One of the miRNAs that increased in the *daf-3(mgDf90)* mutant, *mir-788* (Fig. II-1D, blue bar) was only recently identified (Ruby et al. 2006), and had therefore

not been included in our Y1H experiments. We cloned the *mir-788* promoter and tested it for binding to all 755 available *C. elegans* TFs (Vermeirssen et al. 2007b). *Pmir-788* interacted with eight TFs, including DAF-3 (Fig. II-1E and Table II-5). The promoters of the other two miRNAs that showed a significant increase in expression also bound to DAF-3. Therefore all three miRNAs with significantly increased expression in the *daf-3(mgDf90)* mutant correspond to Y1H positives. This indicates that DAF-3 can function as a transcriptional repressor, which is in agreement with previous observations (Thatcher et al. 1999; Deplancke et al. 2006a). *mir-788* promoter activity was repressed upon dauer formation, which is in agreement with the increase of DAF-3 during this stage (Wang and Kim 2003)(Fig. II-1F). Overall, 25% (3/12) of the miRNAs that interact with DAF-3 in Y1H assays are significantly increased in *daf-3(mgDf90)* dauer animals, while only one out of the 36 miRNAs (3%) that do not interact with DAF-3 showed a small, but significant change in expression (Fig. II-1G). Conversely, the promoters of 75% of the miRNAs that change significantly in *daf-3(mgDf90)* mutants were bound by DAF-3, while only ~20% of the miRNAs that do not change are controlled by a promoter bound by DAF-3 (Fig. II-1H). Together, these results demonstrate that Y1H and TaqMan PCR data correlate (Fisher exact test, P-value=0.04), and provide insights into the transcriptional consequences of physical TF-promoter interactions within the context of an intact animal. It is important to note that those miRNAs that interacted with DAF-3 but that did not change in expression in *daf-3(mgDf90)* animals may be regulated

under different developmental or physiological conditions during the lifetime of the animal. Another possibility is that they may change in expression in one or a few cells within the animal, and thus fall below the detection limit of whole animal TaqMan PCR assays.

#### *A C. elegans miRNA=>TF post-transcription regulatory network*

We generated a post-transcriptional miRNA=>TF network by identifying which of the TFs found in Y1H assays are predicted miRNA targets. Since target prediction algorithms can be noisy (Sethupathy et al. 2006), we only used targets predicted by two or more, from a total of four miRNA target prediction algorithms, including TargetScan, Pictar, miRanda and RNAhybrid (Fig. II-2A and Table II-6, see *Materials and methods*). We identified 252 high-confidence miRNA=>TF interactions involving 67 miRNAs and 73 TFs. The most highly connected miRNAs are members from the *let-7*, *mir-80* and *mir-2* families. The most highly connected TFs are ZAG-1, ZTF-10, and LIN-26, all of which belong to the ZF-C2H2 family, ELT-3 and NHR-14 that belong to the ZF-GATA and ZF-NHR families, respectively.

All miRNA=>TF interactions were visualized into a network model (Fig. II-2B). The out-degree of this network (the number of TFs targeted by a given miRNA) is best fit by an exponential distribution (Fig. II-2C). Most biological networks characterized to date exhibit a different, power-law degree distribution in which a small number of nodes (network hubs) are extremely highly connected



compared to the rest of the nodes in the network (Jeong et al. 2000). The biological significance of this observation became apparent when it was found that hubs in protein-protein interaction networks are often essential for survival or development of the organism (Jeong et al. 2001). For instance, the out-degree distribution of TFs in transcriptional networks follows a power-law, and the TF hubs in these networks tend to be essential for viability (Deplancke et al. 2006a; Vermeirssen et al. 2007a)(Fig. II-1B). The exponential out-degree distribution of the miRNA=>TF post-transcriptional network indicates that no clear miRNA hubs can be identified. Interestingly, *C. elegans* can tolerate removal of most individual miRNAs without obvious developmental defects (Miska et al. 2007). The exponential out-degree distribution of miRNAs and lack of essentiality for most of them both agree with the hypothesis that miRNAs predominantly function to fine tune gene expression instead of establishing crucial developmental gene expression programs (Bartel and Chen 2004; Hornstein and Shomron 2006).

The in-degree distribution of all miRNA target genes (the number of miRNAs that regulate a target) follows a power-law (data not shown). However, the in-degree of the miRNA=>TF post-transcriptional network is best fit by an exponential distribution (Fig. II-2D). When all genes are considered, we find that the target hubs are enriched for TFs (Fisher exact test P-value <0.001), which is in agreement with previous observations in other organisms (Enright et al. 2003; Shalgi et al. 2007). Thus, the exponential in-degree distribution of the

miRNA=>TF network is likely best explained by the exclusive sampling of TF-encoding target genes.

*MiRNAs and TFs function together in composite feedback loops*

We define a 'type I' miRNA<=>TF composite feedback loop as a miRNA and a TF that mutually regulate each other (Fig. II-2E). To systematically identify such loops, we integrated the transcriptional (TF=>miRNA) and post-transcriptional (miRNA=>TF) networks into a directed, bipartite miRNA gene regulatory network and counted the number of composite miRNA<=>TF loops. We found 23 type I composite miRNA<=>TF loops in the integrated network, involving 14 miRNAs and 16 TFs (Fig. II-2F and Table II-7). The 16 TFs represent a variety of families indicating that loops are not biased toward particular types of TFs.

Approaches to confirm the *in vivo* relevance of composite miRNA<=>TF feedback loops require assays to determine the regulatory consequences of Y1H interactions, and to assay TF levels in miRNA mutants. There are two possible type I composite miRNA<=>TF feedback loop subtypes: 'single-negative' feedback loops in which the TF activates the miRNA (also called incoherent loops), and 'double-negative' feedback loops in which the TF represses the miRNA (also called coherent loops) (Fig. II-2E). Here, we focused on the *mir-43<=>LIN-26* composite feedback loop, which we found to belong to the single negative class (Fig. II-3A). To determine whether LIN-26 activates or represses

miRNA expression, we performed TaqMan PCR assays in wild type and *lin-26(ok939)* mutant embryos. We found that *mir-43* levels decrease ~8-fold in *lin-26(ok939)* mutant, compared to wild type embryos. Two other miRNAs, *mir-42* and *mir-44*, are co-expressed with *mir-43* in an operon and, as expected, they also decrease in *lin-26(ok939)* mutant embryos (Fig. II-3B). These data demonstrate that LIN-26 is an activator of *mir-43* expression. One miRNA (*mir-63*) whose promoter was bound by LIN-26 in Y1H assays increases in *lin-26(ok939)* animals, suggesting that LIN-26 may also function as a transcriptional repressor (Fig. II-4A). In total, the expression of four out of six miRNAs targeted by LIN-26 in Y1H assays (67%) changed significantly in *lin-26(ok939)* mutant embryos. After confirming the transcriptional LIN-26=>*mir-43* interaction, we used Western blotting to show that LIN-26 protein levels are increased in *mir-42-44(mgDf49)* mutant embryos (Fig. II-3C) and larvae (Fig. II-4B), which also confirms the post-transcriptional *mir-43*=>LIN-26 interaction. Together, these data demonstrate that *mir-43* and LIN-26 function in a single negative composite feedback loop.

Single negative type I feedback loops can direct stable co-expression of both components (see *Discussion*). Thus, we hypothesized that *mir-43* and LIN-26 are co-expressed at least in some tissues. LIN-26 is expressed throughout the lifetime of *C. elegans*, starting in the early embryo, and is involved in epithelial differentiation (Labouesse et al. 1994). It is expressed in various epithelial tissues, including the hypodermis and seam cells (Landmann et al. 2004). We

created transgenic animals that harbour a *Pmir-42-44::gfp* fusion and found that *Pmir-42-44* drives GFP expression in embryos and throughout development (Fig. II-3D and II-3E). In larval stages, expression was detected in hypodermal seam cells, suggesting that LIN-26 and *mir-43* are indeed co-expressed (Fig. II-3E).

Most single miRNA mutants do not confer a detectable phenotype (Miska et al. 2007) and we were not able to detect a phenotype of a deletion encompassing the *mir-42-44* locus. Thus, *mir-43* could be involved in epithelial differentiation (as suggested by its expression pattern), but may act redundantly with other (miRNA) genes. Comprehensive analysis of gene expression in *C. elegans* will likely help to identify additional genes that may function in this process.

#### *MiRNA<=>TF composite feedback loops form higher order network structures*

Several miRNAs and TFs are involved in higher order network subgraphs that include several loops. For instance, we identified higher order composite feedback loops that contain one miRNA and two TFs ('type II' loops), or one TF and two miRNAs ('type III' loops)(Table II-7). An example of an even more complex sub-graph involving multiple miRNAs, TFs and composite feedback loops is shown in Figure II-5A. The promoters of all members of the *let-7* family of miRNAs (which includes *let-7*, *mir-48*, *mir-84* and *mir-241*) are bound by DAF-3 and these miRNAs are also predicted to target DAF-3. DAF-3 regulates the expression of *mir-241* and *mir-48* in dauer formation and may also regulate *mir-*

84 (Fig. II-2D). Regulation of *let-7* by DAF-3 *in vivo* could not be examined because *let-7* was undetectable in dauer animals using TaqMan PCR assays (data not shown). An additional conserved *let-7* family member, *mir-795*, has only recently been discovered (Ruby et al. 2006). We cloned the promoter of *mir-795* and found that it can also interact with DAF-3 in Y1H assays (Figure II-5B). This highly interconnected sub-graph suggests that the *let-7* family collectively plays a role in dauer formation. We also incorporated available protein-DNA interactions for protein-coding genes (Deplancke et al. 2006a; Vermeirssen et al. 2007a; Vermeirssen et al. 2007b). By doing so, we identified several feed-forward loops, for instance between DAF-3, T27B1.2 and *let-7*. Since we do not yet have comprehensive protein-DNA interaction data for protein-coding genes we cannot examine whether in *C. elegans*, as has been proposed for other systems, feed-forward loops involving miRNAs constitute a network motif (Shalgi et al. 2007; Tsang et al. 2007).

#### *Composite miRNA $\rightleftharpoons$ TF feedback loops constitute a network motif*

To test whether composite miRNA $\rightleftharpoons$ TF feedback loops constitute a genuine network motif, we examined if they are enriched in the integrated miRNA gene regulatory network compared to randomized networks. We used three different methods to generate randomized networks. “Edge switching” (ES) is the most stringent method that maintains the individual degree of each node in the network, and changes only the interaction partners (Milo et al. 2002). “Node

Replacement I” (NR-I) changes the individual degree of the nodes, the identities of the nodes as well as the interaction partners but keeps the overall degree distribution of the network constant. Finally, “Node Replacement II” (NR-II) randomizes everything: the identities of the nodes, the interaction partners and the individual and overall degrees. The use of these three methods not only allows us to determine whether miRNA $\rightleftharpoons$ TF feedback loops constitute a network motif, but also to investigate potential effects of network architecture (see below).

We found that the integrated miRNA gene regulatory network contains approximately twice as many composite miRNA $\rightleftharpoons$ TF feedback loops as the average number of loops found in randomized networks (P-value =0.004 for ES, 0.004 for NR-I and 0.0002 for NR-II, Table II-8). This demonstrates that composite miRNA $\rightleftharpoons$ TF feedback loops constitute a genuine network motif. We have also investigated the presence of composite miRNA $\rightleftharpoons$ TF feedback loops in integrated networks using miRNA predictions retrieved by a single algorithm (as opposed to miRNA predictions common in two or more of the algorithms). In all cases, we observed the same tendency: the number of feedback loops is higher in the real network compared to randomized networks (data not shown).

#### *MiRNAs and TFs in composite feedback loops provide a high information flow*

Interestingly, NR-I and NR-II yielded on average the same number of composite miRNA $\rightleftharpoons$ TF feedback loops (10.6, Table II-7). Randomized

networks generated by NR-II possess a more random degree distribution than randomized networks generated by NR-I. Since both methods produce the same average number of composite feedback loops, this suggests that the overall distribution of in- and out-degrees of either miRNAs or TFs does not contribute to the propensity of forming composite miRNA $\rightleftharpoons$ TF feedback loops. This is in agreement with previously reported mathematical models that examined the expected number of feedback loops in different types of networks with random, scale-free or condensed degree distributions (Itzkovitz et al. 2003).

We investigated the individual degrees of nodes that participate in loops. We found that miRNAs and TFs in composite miRNA $\rightleftharpoons$ TF feedback loops have a higher in- and out-degree than nodes that do not participate in loops (Fig. II-6A and B). In other words, miRNAs in composite feedback loops regulate more TFs and are regulated by more TFs, and *vice versa*. We ranked miRNAs and TFs according to their degree, and annotated whether they participate in a loop or not, and found a significant association between loop participation and a high in- or out-degree (Table II-9). We found the same association when nodes with  $k_{in}=0$  or  $k_{out}=0$  are removed (Table II-8). These observations show that high in- and out-degrees are indicators of loop participation.

To better capture the combined high in- and out-degree properties of a node we introduce a new network parameter, referred to as “flux capacity” ( $F_c = k_{in} \times k_{out}$ , Fig. II-6C). By plotting the out-degree versus the in-degree of each node in the network, we found that a high  $F_c$  better describes the difference between

nodes that participate in loops and those that do not (Fig. II-6D and E, Table II-9). The association between a high  $F_c$  and loop participation suggests that this type of local architecture in a network may predispose loop formation. Indeed, in randomized networks, nodes with a high  $F_c$  participate more frequently in loops than nodes with a low  $F_c$  (Fig. II-7A and B). However, this association is less prominent than the enrichment in the real network (Fig. II-7C). It is important to note that the integrated miRNA network contains twice as many miRNA $\rightleftharpoons$ TF feedback loops than randomized networks, even when the individual and overall degrees remain unaltered (Table II-8, Edge switching). This indicates that, while  $F_c$  is a good indicator for feedback loop participation, there are other determinants involved as well.



## Discussion

In this study, we present the first experimentally mapped genome-scale TF=>miRNA transcription regulatory network in any organism. The integration of this network with a computationally predicted miRNA=>TF post-transcriptional network revealed 23 composite miRNA<=>TF feedback loops in which the TF that binds a miRNA promoter is itself regulated by that same miRNA. This dramatically extends the number of miRNA<=>TF feedback loops identified to date in any organism. The overall number of miRNA<=>TF composite feedback loops is likely even higher because both Y1H assays and computational miRNA target prediction algorithms miss interactions. For instance, miRNA target predictions currently mostly include only those that are conserved in related organisms. However, it is likely that several miRNA=>target interactions may be species-specific.

There are several explanations for missed interactions in the transcriptional TF=>miRNA network. For instance, we did not retrieve any interactions for *lisy-6* or *mir-1*; *lisy-6* is a neuronal miRNA, and *mir-1* is expressed in muscle, so one could expect to retrieve neuronal and muscle TFs, respectively (Chang et al. 2004; Simon et al. 2008). There are several explanations for false negatives in our dataset. First, our library screens are not saturated. For example, when we re-screened *P<sub>lisy-6</sub>*, we retrieved CEH-27 and ODR-7, both of which are neuronal TFs (Vermeirssen et al. 2007a). Interestingly, both TargetScan and RNAhybrid predict putative *lisy-6* binding sites in the 3'UTR of

*ceh-27*, suggesting that they may constitute another composite miRNA $\rightleftharpoons$ TF feedback loop (data not shown, these data have been added to EDGEDb, but are not included in the network). Another composite miRNA $\rightleftharpoons$ TF feedback loop that we did not include in the integrated miRNA network involves *mir-788* and IRX-1. This loop was not included because *mir-788* is one of several miRNAs identified after our initial analyses (Ruby et al. 2006). We found that IRX-1 interacts with the promoter of *mir-788* by Y1H assays (Table II-5); and TargetsScan and RNAhybrid both predict that *mir-788* targets the *irx-1* 3'UTR (Table II-7). Interestingly, *mir-788* and *irx-1* are both expressed in the hypodermis, suggesting that they may function in a single negative composite feedback loop (Reece-Hoyes et al. 2007), this study). Thus, the total number of composite miRNA $\rightleftharpoons$ TF feedback loops identified in this study is actually 25. These loops provide a framework for further functional analysis, both in terms of the underlying biology and the effects they have on gene expression programs.

The second explanation for the presence of false negatives is that some TFs may not function in the context of Y1H assays. For instance, binding that requires heterodimerization or post-translational modification of TFs is missed in Y1H assays (Deplancke et al. 2006a). Finally, transcription regulation of miRNAs may be controlled by *cis*-regulatory elements that reside outside of the promoter fragment used in Y1H assays. In the future, it will be important to map the transcription start site of pri-miRNA transcripts to better delineate their promoters and to further improve Y1H assays to enable the retrieval of heterodimers. To

test the effect of false negatives, we have generated ten networks in which we randomly removed 10% of the TF=>miRNA interactions, and found that feedback loops are still enriched compared to randomized networks (Table II-10). Thus, we conclude that the presence of false negatives does not affect our overall findings.

Both the transcriptional TF=>miRNA and post-transcriptional miRNA=>TF networks may also contain false positive interactions. For instance, many genes do not have an annotated or experimentally determined 3'UTR, and for those genes, the algorithms predict sites in the genomic sequence downstream of the stop codon. Since target prediction algorithms are noisy we did not include any interactions that were identified by only a single miRNA target prediction algorithm.

Y1H assays may also result in false positive TF=>miRNA interactions. For instance, the DNA fragments used may contain regulatory elements that do not regulate the transcription of miRNAs but that of neighboring genes. In addition, although Y1H assays identify reproducible interactions, it may be difficult to detect their regulatory consequence *in vivo* (see below). We aimed to minimize the inclusion of technical false positives in the TF=>miRNA network by applying a stringent Y1H scoring system that takes the quality of the bait, the prey and the interaction into account (see (Vermeirssen et al. 2007a) for a detailed description of the scoring system). After applying this system, we retained 347 interactions, out of 483 that were present in the raw data (data not shown). The quality of the transcriptional miRNA network is demonstrated by the *in vivo* regulatory

confirmation of many physical TF=>miRNA interactions. We found that DAF-3 represses miRNA expression in dauer animals and LIN-26 activates some of its Y1H targets and represses others in embryos. In addition, we found that the FLYWCH TFs FLH-1 and FLH-2 that interact with multiple miRNA promoters, repress miRNA expression in the embryo (Ow et al. 2008). Interactions for which we did not detect a regulatory consequence by TaqMan PCR assays should be regarded as inconclusive because they could fall below the detection limit of TaqMan PCR, or occur in other developmental or environmental conditions. Indeed, we previously observed that some interactions that occur in a particular cell type or tissue can be detected only as a modest effect by quantitative PCR when whole animals are assayed (Deplancke et al. 2006a).

Composite miRNA<=>TF feedback loops likely participate in specific gene regulatory circuits that precisely control gene expression programs in development or homeostasis. For instance, double-negative feedback loops can generate mutually exclusive or bi-stable expression of the miRNA or TF, and, hence their downstream targets (Johnston et al. 2005) (Fig. II-8A). A bi-stable system can switch between two states, depending on which of multiple potential input signals are active (Gardner et al. 2000). Once a state is established, the input signal is no longer necessary. As a result, bi-stable systems provide robust and noise-free gene expression programs. Single-negative feedback loops (Fig. II-8B) can result in stable expression of both components by reducing stochastic fluctuations in gene expression (Tsang et al. 2007). Alternatively, such a loop can

result in oscillatory expression of both components, which depends on additional input signals (Hirata et al. 2002). This could be important in processes such as the cell cycle, molting at different larval stages, or other cyclic processes.

Many miRNAs and TFs that participate in composite feedback loops are characterized by a high Fc. The high out-degree reflects that both the miRNA and the TF have many downstream targets, or regulons, and the loop ensures that the expression of these regulons is tightly correlated. For instance, in bi-stable systems (Fig. II-8A), expression of the miRNA and TF regulons is mutually exclusive. In steady state or oscillatory systems, however, the regulons can be co-expressed, either at steady state levels or in oscillation (Fig. II-8B). The high in-degree of both the miRNAs and the TFs that participate in loops suggests that regulon control is highly adaptable: the systems can be subjected to different stabilizers, switches or modifiers, for instance in different tissues or under different developmental or environmental conditions.

Two recent bioinformatic studies proposed that miRNAs and their targets are involved in feed-forward as well as feedback loops (Shalgi et al. 2007; Tsang et al. 2007). Shalgi and colleagues searched for pairs of miRNAs and TFs that co-regulate target genes by identifying putative miRNA sites and TF binding sites that co-occur in individual genes. They observed that such miRNA-TF pairs are predicted to regulate each other more frequently than randomly picked pairs, suggesting the existence of feedback loops. Tsang *et al.* proposed that correlation or anti-correlation between miRNAs and their targets can result from

various types of feed-forward and feedback loops involving miRNAs, their predicted target genes and upstream regulators (e.g. TFs, kinases), but did not predict any *actual* loops (Tsang et al. 2007). We now provide 23 (25 when *lisy-6*  $\rightleftharpoons$  CEH-27 and *mir-788*  $\rightleftharpoons$  IRX-1 are included) novel miRNA  $\rightleftharpoons$  TF feedback loops and demonstrate that these correspond to a genuine network motif.

Feedback motifs are rare in pure transcriptional networks (Milo et al. 2002; Shen-Orr et al. 2002). We demonstrate that miRNAs are a post-transcriptional missing link to form feedback motifs. It is likely that other post-transcriptional interactions are also involved in feedback regulation. Previously, Margalit and colleagues have shown that protein-protein interactions play a role in generating composite feedback loops in the transcriptional network of *Saccharomyces cerevisiae* (Yeager-Lotem et al. 2004). It is likely that protein-protein interactions also contribute to the formation of loops involving miRNAs (or other regulators) in *C. elegans* networks. In the future, it will be important to integrate miRNA gene regulatory networks with genome-scale protein-protein interaction networks and other functional networks as well.

Taken together, we propose that composite miRNA  $\rightleftharpoons$  TF feedback loops provide a common mechanism of gene regulation at a systems level in *C. elegans*. Similar system level analyses will reveal whether the individual composite miRNA  $\rightleftharpoons$  TF feedback loops found in other organisms are also examples of a network motif, and whether this mechanism is evolutionarily conserved.

## Materials and methods

### *MiRNA promoter definition*

We used the 115 miRNA gene predictions available in WormBase WS130 (<http://www.wormbase.org>) and miRNA registry V4.0 (<http://microna.sanger.ac.uk>). A miRNA promoter is defined here as the intergenic region upstream of the predicted stem-loop sequence annotated in miRBase V4.0 (Table II-1). We used a minimal length of 300 bp and a maximal length of 2 kb. In total, 79 promoters (for a total of 95 miRNAs) were selected as DNA baits for Y1H assays (Table II-2). Seventy-one promoters (controlling 84 miRNAs) were successfully cloned into pMW#2 and pMW#3 by Gateway cloning (Walhout et al. 2000b) and integrated into the genome of *S. cerevisiae* YM4271 (Deplancke et al. 2006b).

### *Gateway-compatible Y1H assays*

Detailed Y1H protocols are described elsewhere (Deplancke et al. 2006b). Y1H screens were performed with each miRNA promoter bait strain versus both AD-wrmcDNA ( $>10^6$  colonies screened)(Walhout et al. 2000b) and AD-TF ( $>3.10^5$  colonies screened)(Deplancke et al. 2004) prey libraries. For *Pmir-61-250*, both reporters were highly self-active and, therefore, this bait could not be used. All interactions were retested by PCR/gap repair (Deplancke et al. 2006b). PCR products corresponding to preys that retested were sequenced by Agencourt Bioscience Corporation. Interactors were identified by BLASTX. In total 669

Interaction Sequence Tags (ISTs) were obtained (Walhout et al. 2000a). Y1H matrix experiments were performed by transforming all interactors obtained in the screens (for which a clone was available), and several TFs found in previous studies (Deplancke et al. 2006a; Vermeirssen et al. 2007a) into each promoter bait strain (130 preys were used in total, Table II-3). In addition, six baits and *Pmir-788* were screened versus our AD-TF yeast array (Vermeirssen et al. 2007b)(Table II-4). Ninety-eight percent of the ~10,500 transformations were successful. All interactions obtained were subjected to a stringent standardized scoring and filtering system (Vermeirssen et al. 2007a). Only interactions with a score  $\geq 5$  were retained (Table II-4). All interactions are available in the EDGEDb database (Barrasa et al. 2007).

### *C. elegans strains*

*C. elegans* N2 wild type, GR1311 [*daf-3(mgDf90)*], VC663 [*lin-26(ok939)*] and MT13372 [*mir-42-44(nDf49)*] strains were cultured on OP50 seeded NGM plates at 20°C unless otherwise noted.

### *TaqMan PCR assays*

Templates for miRNA TaqMan PCR assays were obtained by collecting 50 N2 and *daf-3(mgDf90)* dauer animals, or 100 N2 and homozygous *lin-26(ok939)* segregant mid- to late stage embryos into lysis buffer (50 mM KCl, 10 mM Tris pH 8.3, 2.5 mM MgCl<sub>2</sub>, 0.45% NP-40, 0.45% Tween-20, and 0.01%



gelatin). The samples were subjected to ten cycles of freezing and thawing and incubated at 65°C for 1 hr and 95°C for 20 minutes. After Trizol Reagent (Invitrogen 15596-026) extraction the RNA was co-precipitated with glycogen. MicroRNA TaqMan PCR assays were performed following the recommendations of the manufacturer (Applied Biosystems). A TaqMan PCR assay for the small nuclear RNA sn2343 was used as normalization standard.

#### *Induction of dauer larvae formation*

Dauer pheromone was prepared as described (Vowels and Thomas 1994). N2 and *daf-3(mgDf90)* embryos were hatched on 5 mm pheromone plates (NGM without peptone, supplemented with 100 mg/ml streptomycin and seeded with 6X OP50) and incubated at 25°C for 3 days.

#### *Normalization and analysis of TaqMan PCR data*

*daf-3(mgDf90)* versus N2: A total of five independent biological experiments were performed using TaqMan probes for 107 miRNAs in *daf-3(mgDf90)* and N2 wild type dauer animals, and each experiment was done in triplicate. In each experiment, a  $Ct_{\text{mean}}$  value of the three technical replicates was calculated. The standard deviations (SD) within technical repeats were very low (~0.1-0.4, data not shown). MiRNAs with  $Ct_{\text{mean}}$  values  $\geq 35$  in either wild type or *daf-3(mf90)* animal were discarded (21 miRNAs in total). A  $\Delta\Delta Ct$  value was calculated using the following formula:

$$\Delta\Delta Ct = [Ct_{\text{mean miRNA}} - Ct_{\text{mean control}}]_{\text{daf-3 (mgDf90)}} - [Ct_{\text{mean miRNA}} - Ct_{\text{mean control}}]_{N2}$$

Experiments were normalized by first calculating the average  $\Delta\Delta Ct$  value of all miRNAs within each experiment and then subtracting this value from each individual  $\Delta\Delta Ct$  value. We only used miRNAs that had  $\Delta\Delta Ct$  values in four or five experiments (9 miRNAs were discarded). Normalized  $\Delta\Delta Ct$  values for each miRNA were averaged across all experiments to calculate  $\Delta\Delta Ct_{\text{final}}$  and the standard error of the mean was determined. Only the 48 miRNAs for which both Y1H and TaqMan PCR data were available are visualized in Figure II-1D (the fold difference in expression is defined as  $2^{-\Delta\Delta Ct}$ ). Z-scores were calculated as  $-\Delta\Delta Ct_{\text{final}}/SD$ . Z-scores  $\geq 2$  were considered significant.

*lin-26(ok939) versus N2*: A total of three independent biological experiments were performed using TaqMan probes for the eight miRNAs whose promoters were bound by LIN-26 in Y1H assays. In each experiment, a  $Ct_{\text{mean}}$  value of the three technical replicates was calculated. MiRNAs with  $Ct_{\text{mean}} \geq 35$  in either wild type or *lin-26 (ok939)* animals were discarded (2 miRNAs). A  $\Delta\Delta Ct$  value was calculated using the following formula:

$$\Delta\Delta Ct = [Ct_{\text{mean miRNA}} - Ct_{\text{mean control}}]_{\text{lin-26 (ok939)}} - [Ct_{\text{mean miRNA}} - Ct_{\text{mean control}}]_{N2}$$

The average  $\Delta\Delta Ct$  of the three experiments and the standard error of the mean for all three experiments were calculated and visualized in Figures II-3B and II-4A (the fold difference in expression is defined as  $2^{-\Delta\Delta Ct}$ ).

### *Transgenesis*

Transgenic *promoter::gfp* animals were generated as described (Reece-Hoyes et al. 2007).

### *MiRNA target predictions*

Four programs were used to obtain miRNA target predictions: Pictar (Lall et al. 2006), miRanda Targets Version 4 (Griffiths-Jones et al. 2006), TargetScan Release 3.1 (Lewis et al. 2005) and RNA-hybrid (Rehmsmeier et al. 2004)(by running the algorithm locally). For RNA hybrid predictions 3'UTR sequences were obtained from WormBase WS159. For genes that did not have an annotated 3'UTR, 300 nucleotides downstream of the stop codon of the longest annotated transcript were taken. Only genes with annotated *C. briggsae* orthologs were used. Pairing of the seed region was performed allowing either GU pairs, or one bulge on the mRNA side within the seed region (but no G:U wobbles). Predictions were filtered for minimum free energies (MFE) <-15. RNA-hybrid was run both for *C. elegans* and *C. briggsae* and only predictions common in both were kept. The following modification was made to the RNA-hybrid code: the original program would find the best hybrid (smaller MFE) for a particular miRNA position, block out that entire site in the 3'UTR and no longer consider any of those nucleotides for other possible matches with that miRNA. This resulted in sites with minimal MFE but that may contain seed imperfections, and the program would miss perfect seed match sites with a slightly higher MFE that could be found by just shifting one or two nucleotides. We modified the code to

allow the selection of sites with a better seed match and post-processed the output to eliminate duplicate predictions (Hammell *et al.*, 2008). For subsequent analyses we only used targets predicted by multiple programs.

For miRNA=>TF target predictions, we included all TFs found in Y1H assays, including novel putative TFs. TargetScan targets are defined with GenBank NM identifiers (IDs). Targets predicted by other algorithms are listed with WormBase IDs. We mapped WormBase IDs to NM IDs using a conversion data file kindly provided by G. Bell.

An “all genes” list was assembled by downloading “confirmed” and “partially confirmed” gene IDs from WormBase WS170 (we obtained 14,631 non-TF gene IDs). We were able to match 13,794 WormBase IDs to NM IDs. Therefore, the “all genes” list used to retrieve miRNA predictions had a total of 14,754 genes: 960 TF genes (including novel putative TFs) and 13,794 non-TF genes.

### *Western blotting*

Wild type and *mir-42-44(nDf49)* mutant worms were grown in 60 mm OP50 seeded NGM plates and bleached (25% commercial bleach/0.25 M KOH, 5-10 minutes) when most animals were gravid adults. Eggs were washed in M9 buffer and either collected for extract preparation or incubated at 20°C for 18 hours on S medium to allow hatching. To obtain larval stages, worms were placed on OP50 seeded NGM plates and harvested after 15 (L1) and 46 hrs (L4),

respectively. Worms were washed 4 times in M9 buffer and resuspended in 2 ml M9 buffer. Worms were centrifuged (2,000 rpm, 1 min), transferred to a weighed 1.5 ml Eppendorf tube and centrifuged again. Supernatant was removed and the worm pellet was weighed. To estimate worm pellet volume we assumed that 1 g equals 1 ml. The worm pellet was resuspended in an equal volume of pre-warmed 2X sample buffer (2% SDS, 10% glycerol, 0.1% bromophenol blue, 100mM dithiothreitol, 50mM Tris-HCL pH 6.8) and boiled for 10 minutes. To reduce viscosity, samples were sheared using a 1 cc syringe. Insoluble debris was removed by centrifugation at 14,000 rpm for 1 minute, and supernatant was transferred to a clean Eppendorf tube. Approximately 90,000 embryos, 30,000 L1 and 10,000 L4 animals were used for each genotype. Equal volumes of mutant and wild type extracts were run on NuPAGE 4-12% bis-Tris Gel (Invitrogen NP0323) and electroblotted onto PVDF membranes. PVDF membranes were incubated overnight at 4°C with anti-LIN-26 antibody (a kind gift from J. Polanowska; 1:2,000 dilution in TBS-Tween with 5% dry milk) or for 1 hour with murine anti- $\alpha$ -tubulin antibody (Sigma #T6074) for 1 hr at room temperature. HRP-conjugated secondary antibodies (goat anti-rabbit for LIN-26 and anti-mouse for  $\alpha$ -tubulin, respectively) incubations were done for 1 or 2 hours at room temperature.

### *Network randomizations*

The integrated miRNA-TF gene regulatory network is a bipartite directed network that is composed of two types of nodes and two types of edges: TF=>*PmiRNA* and miRNA=>TF interactions. To avoid randomly generating meaningless interactions such as miRNA=>*PmiRNA* or TF=>TF, the transcriptional and post-transcriptional networks were randomized separately and then combined for motif analysis. Several miRNAs are transcribed from operons that contain two or more miRNAs (Table II-2). Whereas one miRNA within an operon may be part of a network motif, the others do not have to be part of the same motif. To enable correct motif analysis, we added a third type of edge between miRNA promoters and each of the miRNAs they control. For example, LIN-26 binds the promoter of *Pmir-42-44* (that controls *mir-42*, *43* and *44*) but only *mir-43* targets LIN-26. The *Pmir*=>miRNA edges were never randomized.

We used three randomization strategies and only nodes present in the real networks were used (*i.e.* all miRNAs whose promoters were cloned, and TFs retrieved by Y1H assays): 1) *Edge switching (ES)* (Milo et al. 2002). Two edges are randomly picked from the network and the target nodes between them are exchanged (*e.g.* A-B and C-D will become A-D and C-B). The switch is only performed if the new edges are not already present in the newly created network. A random number of switches, between 100 and 200 times the number of edges, are performed to create randomized networks. With ES, the individual in- and out-degrees of nodes are maintained, and therefore the overall distributions of in-degree, out-degree and flux capacity are kept as well. 2) *Node Replacement I*

(*NR-I*). This method maintains the overall in- and out-degree distributions but randomizes the in- and out-degree of individual nodes. The origin nodes (the first component of an edge) are first replaced (e.g. A is replaced with E in the A-B edge). All positions of the origin node are replaced with the same substitute node (e.g. A-B, A-C become E-B, E-C). Then, target nodes are randomized in the same way as the origin nodes (e.g. E-B, D-B become E-F, D-F). 3) *Node Replacement II (NR-II)*. This method randomly replaces the nodes in the networks without preserving the degree distribution. Nodes were randomized one edge at a time and replaced with a randomly picked node of the same type. If a node substitution results in an edge that is already present, we randomly select a different node to replace it. Nodes can be picked multiple times, resulting in a more random degree distribution and a random individual node degree.

To assess the influence of  $F_c$  on loop participation in randomized networks, 400 randomized networks were made using the ES method. We counted the number of times each node, with a specific  $k_{out}$  and  $k_{in}$ , was part of a loop and the number of times it was not part of a loop in the randomized networks. We then calculated and plotted the ratio between the number of times a node was in a loop versus the number of times it was not in a loop.

### *Network motif analysis*

We used Mfinder (Kashtan et al. 2004) to count the number of motifs in the original and randomized networks. Type I composite miRNA $\leftrightarrow$ TF feedback

loops are represented by three nodes in the integrated miRNA gene regulatory network: the TF, the miRNA promoter and the miRNA itself (see above)(Mfinder motif ID98). The higher order motifs in Table II-7 include type II (Mfinder motif ID 4546) and type III (Mfinder motif ID 1090054). P-values were calculated empirically, using the distribution of loop counts in the appropriate set of generated randomized networks. Specifically, the P-value is defined as the proportion of random networks that have the same or larger number of motifs as observed in the original network. A P-value  $\leq 0.01$  was considered significant.



## Acknowledgments

We thank members of the Walhout laboratory and J. Dekker for discussions and critical reading of the manuscript. We thank B. Deplancke for help with the initial cloning of miRNA promoters, and the sequencing staff at Agencourt Bioscience for technical assistance. We thank J. Polanowska for the LIN-26 antibody, M. Green (University of Massachusetts Medical School) for the anti- $\alpha$ TUBULIN antibody and G. Bell (Bioinformatics and Research Computing, Whitehead Institute) for the gene name conversion file. We are grateful to C. Chen from Applied Biosystems for providing reagents for the TaqMan PCR assays to V. A., and to X. Karp for the dauer pheromone. Finally, some nematode strains were provided by the *Caenorhabditis* Genetics Center (CGC), which is funded by the NCCR. This work was supported by NIH grant DK068429 to A.J.M.W. and GM348642 to V.A.; F.P.R. is supported by NIH grants (HG003224, HG0017115, HL81341, and HG004233), and by the Keck Foundation; M. C. O. was funded in part by NIH postdoctoral fellowship GM070118-02.

**Table II-1. Summary of miRNA gene location and promoters analyzed by Y1H.**

Table S1. Summary of miRNA gene location and promoters analyzed by Y1H assays

miRNA location (WS130)	# of miRNAs (miRNA registry V 4.0)	# of promoters attempted	# of promoters cloned into Y1H vectors	# of promoter with Y1H interactions (after scoring)
Intergenic	66	66	61	54
Intergenic/Operon	16 <sup>a</sup>	6 <sup>a</sup>	4 <sup>a</sup>	3 <sup>a</sup>
Intragenic/Antisense orientation	4	4	3	3
Intragenic/Antisense orientation/Operon	9	2	2	2
Intragenic	20	1 <sup>b</sup>	1 <sup>b</sup>	1 <sup>b</sup>
Total	115	79	71	63

<sup>a</sup> *mir-227* is not in the miRNA registry but was included in the analysis.

<sup>b</sup> *mir-2* is annotated in 5th intron of F16A11.3. Sequence upstream of *mir-2* was taken as promoter.

Table II-2. Promoter details and primer sequences.

cel-miRNA (miRNA registry V 4.0)	Promoter	Location WS130	Chr	<sup>a</sup> GW- Forward Primer	<sup>b</sup> GW- Reverse Primer	Cloned	Y1H inx
<i>cel-let-7</i>	<i>Plet-7</i>	Intergenic	X	TTATCTTTTAATG TCTTCAC	CCGGATCCACA GTGTAGACCGTC	yes	yes
<i>cel-lin-4</i>	<i>Plin-4</i>	Intergenic	II	TTCACAGCGCC AGTCGTTTCATG	ACAGGCCGGAA GCATAAACTC	yes	yes
<i>cel-lsy-6</i>	<i>Plsy-6</i>	Intergenic	V	TTTAAACTTACA GTTTCGG	GCTTATTTTCA GAAATTAG	yes	yes
<i>cel-mir-1</i>	<i>Pmir-1</i>	Intergenic	I	ATGACATCTGCA AGAAAC	ATGGGCATGTAA GGAAGTATGC	yes	no
<i>cel-mir-228</i>	<i>Pmir-228</i>	Intergenic	IV	GAGATTCATATT CGATTTTTCCC	TACTTCATTAAC CTCATCACC	yes	yes
<i>cel-mir-230</i>	<i>Pmir-230</i>	Intergenic	X	ACACATTCACAG TTTTACTGG	ACATCAGACAAA AAAATACAAGC	yes	yes
<i>cel-mir-231</i>	<i>Pmir-231</i>	Intergenic	III	GTTTACCATTTTA GAATCCGC	TGTGGTGCTAAC TTTGAATAA	yes	yes
<i>cel-mir-232</i>	<i>Pmir-232</i>	Intergenic	IV	CATACAATTTTAC ATCATAACG	CAAACCTGCCTG GAATCAA	yes	yes
<i>cel-mir-234</i>	<i>Pmir-234</i>	Intergenic	II	TGCATCCAATAC AAATATCAG	TGTGTTTCAACT TACATAG	yes	yes
<i>cel-mir-235</i>	<i>Pmir-235</i>	Intergenic	I	TTTTTTGCAGCT TCGTTTCAGCG	TAATTTTCACGG TGAACAATTTG C	yes	yes
<i>cel-mir-236</i>	<i>Pmir-236</i>	Intergenic	II	CTTTAAATTAATA CGGTGAG	ACCGAAGAATAC GACAGTAGC	yes	yes
<i>cel-mir-237</i>	<i>Pmir-237</i>	Intergenic	X	GACTTCACTCAC GTAGGTTTC	CCACGCAATGTA GAAGTTTTG	yes	yes
<i>cel-mir-238</i>	<i>Pmir-238</i>	Intergenic	III	AAAAGAACGTTG AGTTAC	GGAGAATTGCCA TTTTCTTG	yes	yes
<i>cel-mir-239a</i>	<i>Pmir-239a</i>	Intergenic	X	GAGCTTTTGAG GAAATTT	TGTAGAGTTATT TGACTGAG	yes	yes
<i>cel-mir-239b</i>	<i>Pmir-239b</i>	Intergenic	X	AAACGTCGCATT TTGTAT	CTGTAACCTTTTC AGAATAATAAGG	no	NA
<i>cel-mir-240</i>	<i>Pmir-240</i>	Intergenic	X	TTATCTATCCGA CCATACC	CGAAATAGCAAT AGACAGGA	yes	yes
<i>cel-mir-241</i>	<i>Pmir-241</i>	Intergenic	V	CACTCATTTTAC TTTCTCTC	ACTTTGACACCC CCGCGGTTTG	yes	yes
<i>cel-mir-242</i>	<i>Pmir-242</i>	Intergenic	IV	TCAACGCGTACA CTTTG	CGAGCCACGAG GGAACATTC	yes	yes
<i>cel-mir-243</i>	<i>Pmir-243</i>	Intergenic	IV	ATCGTGTGACAA TGACGTTG	GTGAACAGAATT TCTGTGAGTAAC	yes	yes
<i>cel-mir-244</i>	<i>Pmir-244</i>	Intergenic	I	CCCCCTTTGAC AATCGAATA	ATGGAGGGAAA AAATTAATAATA ΔACTC	yes	no
<i>cel-mir-245</i>	<i>Pmir-245</i>	Intergenic	I	ATTCCAACCGAA TTTAAATC	ATAGCTCTCCAA CATTGAATTC	yes	yes
<i>cel-mir-246</i>	<i>Pmir-246</i>	Intergenic	IV	GAGTGTGAGC GGGAAAC	TGCTGAATATTG TGTAGC	yes	no
<i>cel-mir-247</i>	<i>Pmir-247</i>	Intergenic	X	AATTTGGAAGAA GAAAAATC	ATTTGTAGACAA CTAGTGTGC	yes	yes
<i>cel-mir-249</i>	<i>Pmir-249</i>	Intergenic	X	AGTTTAACTTCA GCTCTTG	GAGTATAGATAC ACAATATGC	yes	yes
<i>cel-mir-251</i>	<i>Pmir-251</i>	Intergenic	X	TCCTATTCAATT GATTGTTCCG	GCAAGGGATCAT TTTCAC	yes	yes

<i>cel-mir-252</i>	<i>Pmir-252</i>	Intergenic	II	AGATGATATTCA AGACATTATCG	TGGAATAGTAAG ATCAGAATGAC	yes	yes
<i>cel-mir-255</i>	<i>Pmir-255</i>	Intergenic	V	AACGAGTTGTGT GAGTCTTG	CACTGAAGAATT AACAG	yes	yes
<i>cel-mir-256</i>	<i>Pmir-256</i>	Intergenic	V	TCTCGTGGAGA AAATGGC	TTCTCAATATTCA TCTAGACAGC	no	NA
<i>cel-mir-257</i>	<i>Pmir-257</i>	Intergenic	V	AATCTCCCTGAC GACTCC	TTACTTCGCAAA CAGCTTC	yes	yes
<i>cel-mir-258.1</i>	<i>Pmir-258.1</i>	Intergenic	X	CCTAAATTTCCC AAAAAATGG	CCTGCGGATCG ACTATTG	yes	yes
<i>cel-mir-258.2</i>	<i>Pmir-258.2</i>	Intergenic	X	NA	NA	NA	NA
<i>cel-mir-259</i>	<i>Pmir-259</i>	Intergenic	V	ACTTTAAAAGTC TTCTGGAAAAAG	AACCGAACTTAC CACTATG	yes	yes
<i>cel-mir-261</i>	<i>Pmir-261</i>	Intergenic	II	TGGTAAACATCA AAGCAAAC	ATCCATATTCAC CGTGAAA	yes	yes
<i>cel-mir-262</i>	<i>Pmir-262</i>	Intergenic	V	CGGTGGAATATT GTTAAAATTG	CTAATTGTAGAG AAACCATAC	yes	yes
<i>cel-mir-264</i>	<i>Pmir-264</i>	Intergenic	X	GATTTTTGTAC GGGAATG	AGTGCTGCTTCT CGAATG	no	NA
<i>cel-mir-265</i>	<i>Pmir-265</i>	Intergenic	IV	AGTAGAAAATAG AAAGTCAGATCC	GTCAGAATAACC ACACAAAATTAG	yes	no
<i>cel-mir-266</i>	<i>Pmir-266</i>	Intergenic	X	GACACGGGAGA TGTAACC	CAGCTTCTGATC CAAGTTA	yes	no
<i>cel-mir-268</i>	<i>Pmir-268</i>	Intergenic	V	ATATTCCATATTC ACATGC	CTCTCCAAAATG TTTCG	yes	yes
<i>cel-mir-269</i>	<i>Pmir-269</i>	Intergenic	IV	AGAACATGTCTG TTTTGAGATAG	CAAGTTTAGCCA ACTTCTG	yes	yes
<i>cel-mir-270</i>	<i>Pmir-270</i>	Intergenic	IV	TCCGCTGCCAA GAAGCTTC	CTTTGTGCCTAC ATAAAAACC	yes	yes
<i>cel-mir-271</i>	<i>Pmir-271</i>	Intergenic	X	ATGGTCTTATCA GCAGAAGC	TGACCTCCTCTT TCTCAC	yes	yes
<i>cel-mir-34</i>	<i>Pmir-34</i>	Intergenic	X	TAAAGCCCCTG GTGACT	GCATTGTCCGTT ATAAGAAT	yes	yes
<i>cel-mir-355</i>	<i>Pmir-355</i>	Intergenic	II	CATTCTGGAAAA ACTAGTCGG	TATTAATCTCAAA CAGCGAC	yes	yes
<i>cel-mir-359</i>	<i>Pmir-359</i>	Intergenic	X	TGGTCTTAGTGG TCTTAGTTC	GAGCATTAATAT CAGTACTGTAG	yes	no
<i>cel-mir-360</i>	<i>Pmir-360</i>	Intergenic	X	GGTTATGGGGTT TTATGA	GCCGAATAATCA ATAAGAA	yes	no
<i>cel-mir-392</i>	<i>Pmir-392</i>	Intergenic	X	CTTCGTGTAAC TACGTTTG	TATCCAGAATAAA TTTTGAAG	yes	yes
<i>cel-mir-45</i>	<i>Pmir-45</i>	Intergenic	II	GCCGATTGTGAT CAAAAAAATC	GGTGCATGAGC AGCTGAATAC	yes	yes
<i>cel-mir-46</i>	<i>Pmir-46</i>	Intergenic	III	GCCACCCAATTC ACGAGCC	TTTGACCCTTA CCTTCAACTTTT CG	yes	yes
<i>cel-mir-47</i>	<i>Pmir-47</i>	Intergenic	X	ATCGTGAAACA ATAGGTCT	GATCAACTAACC TTGTCCG	yes	yes
<i>cel-mir-48</i>	<i>Pmir-48</i>	Intergenic	V	ATGGCAACATTA ACATCAAAG	AGTTCCCGGGA GTTTCAATTG	yes	yes
<i>cel-mir-49</i>	<i>Pmir-49</i>	Intergenic	X	ATGCGTTGTATT GTTATC	GTACACTAGCAT CTGAATAAA	no	NA
<i>cel-mir-51</i>	<i>Pmir-51</i>	Intergenic	IV	CACAATATATGG GTCGCCAGTC	CAAGTGAGCAG GCAAATGTTGG	yes	yes
<i>cel-mir-52</i>	<i>Pmir-52</i>	Intergenic	IV	GGCCACATACA AACTGC	GTTGGAAGAACA TGAATATG	yes	yes
<i>cel-mir-53</i>	<i>Pmir-53</i>	Intergenic	IV	CGGATCCTTAAG TCAGAAATTG	GAATCGGGAGA AATTTAT	yes	yes

<i>cel-mir-57</i>	<i>Pmir-57</i>	Intergenic	II	ATAATTCAAACG GAACGATCG	CGATGAGTTCCAC ATACCTTTTTG	yes	yes
<i>cel-mir-59</i>	<i>Pmir-59</i>	Intergenic	IV	CTTCTTTCTACC CATTCTG	ATGTCATATCTAG ATCAACTC	yes	yes
<i>cel-mir-60</i>	<i>Pmir-60</i>	Intergenic	II	AGAAGATGATAC ACTTTTT	CGAGTAGAATAT GAATACGAA	yes	yes
<i>cel-mir-63</i>	<i>Pmir-63</i>	Intergenic	X	GTGTAAACTTTC GATATCAC	TGAATGAATAAA GTTGCTCC	yes	yes
<i>cel-mir-72</i>	<i>Pmir-72</i>	Intergenic	II	ACTCAATTCTCA CCTCCATT	TCAGCAGAATAG TAGTAGTAG	yes	yes
<i>cel-mir-75</i>	<i>Pmir-75</i>	Intergenic	X	GCCAAATAAAAG TTACATT	AAGTAGAAAAA AAACCGC	yes	yes
<i>cel-mir-76</i>	<i>Pmir-76</i>	Intergenic	III	TCAAAAACCAGG AAAG	TCTTACCTGAGG ATTGTAG	yes	yes
<i>cel-mir-77</i>	<i>Pmir-77</i>	Intergenic	II	ACTTCGAGTTTT TGAAATCCG	CAAACCTCTATAC CAATTTGG	yes	yes
<i>cel-mir-78</i>	<i>Pmir-78</i>	Intergenic	IV	TAGCGCCTGATC TTGTCATTG	GTTATTTTACGG ACACTATG	no	NA
<i>cel-mir-79</i>	<i>Pmir-79</i>	Intergenic	I	GATTAATTGACAT GATATC	TCTACTAATTTCA ACTTACC	yes	yes
<i>cel-mir-83</i>	<i>Pmir-83</i>	Intergenic	IV	AATACGAATTC CCACC	CCGAGTGGTGC TTTTAAATTG	yes	yes
<i>cel-mir-84</i>	<i>Pmir-84</i>	Intergenic	X	ACTAAACAAAAGT CAGATCAC	GATGCCACAGG CAGACGTATG	yes	yes
<i>cel-mir-90</i>	<i>Pmir-90</i>	Intergenic	III	CATTGTGTTTAA CAATGCCATGC	CGCCCGTTTTGT TAGAAATGG	yes	yes
<i>cel-mir-227</i>	<i>Pmir-227-80</i>	Intergenic/Operon	III	TTGATTGGAAAA TTGAAAGC	GAGTGTCCATTG AAAAAGG	yes	yes
<i>cel-mir-229</i>	<i>Pmir-229-64-65-66</i>	Intergenic/Operon	III	GCAAATTGCCG GAATTG	GAGTCTTCCCG CGTTG	no	NA
<i>cel-mir-250</i>	<i>Pmir-61-250</i>	Intergenic/Operon	V	NA	NA	NA	NA
<i>cel-mir-42</i>	<i>Pmir-42-43-44</i>	Intergenic/Operon	II	GATTAATTATGGA TGTGCG	GTTTATGGGGAT TTTAGAGG	yes	yes
<i>cel-mir-43</i>	<i>Pmir-42-43-44</i>	Intergenic/Operon	II	NA	NA	NA	NA
<i>cel-mir-44</i>	<i>Pmir-42-43-44</i>	Intergenic/Operon	II	NA	NA	NA	NA
<i>cel-mir-54</i>	<i>Pmir-54-55-56</i>	Intergenic/Operon	X	TGATACCAAAAC AATTGAGTG	ACATTACCAATT CCAGAG	yes	yes
<i>cel-mir-55</i>	<i>Pmir-54-55-56</i>	Intergenic/Operon	X	NA	NA	NA	NA
<i>cel-mir-56</i>	<i>Pmir-54-55-56</i>	Intergenic/Operon	X	NA	NA	NA	NA
<i>cel-mir-61</i>	<i>Pmir-61-250</i>	Intergenic/Operon	V	TAGAGACAAAAT TGATGTTT	ATAGGCAGGGA GTAGCAG	yes	no
<i>cel-mir-64</i>	<i>Pmir-229-64-65-66</i>	Intergenic/Operon	III	AATTTTCTGGCA AATTGCCGG	GAGTCTTCCCG CGTTGGG	NA	NA
<i>cel-mir-65</i>	<i>Pmir-229-64-65-66</i>	Intergenic/Operon	III	NA	NA	NA	NA
<i>cel-mir-66</i>	<i>Pmir-229-64-65-66</i>	Intergenic/Operon	III	NA	NA	NA	NA
<i>cel-mir-73</i>	<i>Pmir-73-74</i>	Intergenic/Operon	X	TTGGCCAATGG GGACGAA	GAATAAATGAAT GAAGAGTG	no	NA
<i>cel-mir-74</i>	<i>Pmir-73-74</i>	Intergenic/Operon	X	NA	NA	NA	NA
<i>cel-mir-80</i>	<i>Pmir-227-80</i>	Intergenic/Operon	III	NA	NA	NA	NA
<i>cel-mir-124</i>	<i>Pmir-124</i>	Intragenic	IV	NA	NA	NA	NA
<i>cel-mir-2</i>	<i>Pmir-2</i>	Intragenic	I	TGTTGAATGGGA AAGTGC	AGATTCGTCGAA ATACGATTG	yes	yes
<i>cel-mir-233</i>	<i>Pmir-233</i>	Intragenic	X	NA	NA	NA	NA
<i>cel-mir-248</i>	<i>Pmir-248</i>	Intragenic	X	NA	NA	NA	NA
<i>cel-mir-253</i>	<i>Pmir-253</i>	Intragenic	V	NA	NA	NA	NA
<i>cel-mir-254</i>	<i>Pmir-254</i>	Intragenic	X	NA	NA	NA	NA

<i>cel-mir-272</i>	<i>Pmir-272</i>	Intragenic	III	NA	NA	NA	NA
<i>cel-mir-273</i>	<i>Pmir-273</i>	Intragenic	II	NA	NA	NA	NA
<i>cel-mir-353</i>	<i>Pmir-353</i>	Intragenic	I	NA	NA	NA	NA
<i>cel-mir-354</i>	<i>Pmir-354</i>	Intragenic	I	NA	NA	NA	NA
<i>cel-mir-356</i>	<i>Pmir-356</i>	Intragenic	III	NA	NA	NA	NA
<i>cel-mir-50</i>	<i>Pmir-50</i>	Intragenic	I	NA	NA	NA	NA
<i>cel-mir-58</i>	<i>Pmir-58</i>	Intragenic	IV	NA	NA	NA	NA
<i>cel-mir-62</i>	<i>Pmir-62</i>	Intragenic	X	NA	NA	NA	NA
<i>cel-mir-67</i>	<i>Pmir-67</i>	Intragenic	III	NA	NA	NA	NA
<i>cel-mir-70</i>	<i>Pmir-70</i>	Intragenic	V	NA	NA	NA	NA
<i>cel-mir-71</i>	<i>Pmir-71</i>	Intragenic	I	NA	NA	NA	NA
<i>cel-mir-82</i>	<i>Pmir-82</i>	Intragenic	X	NA	NA	NA	NA
<i>cel-mir-86</i>	<i>Pmir-86</i>	Intragenic	III	NA	NA	NA	NA
<i>cel-mir-87</i>	<i>Pmir-87</i>	Intragenic	V	NA	NA	NA	NA
<i>cel-mir-260</i>	<i>Pmir-260</i>	Intragenic/antisense	II	CAGGACCGACG AAGGCAT	TTACATGGAGGT CCCGC	no	NA
<i>cel-mir-267</i>	<i>Pmir-267</i>	Intragenic/antisense	II	GCAGCCGACAC ATTACGGGATC	GGGTTTTGAAAC TTTTATTTTA	yes	yes
<i>cel-mir-81</i>	<i>Pmir-81</i>	Intragenic/antisense	X	GTGTAGCAACC GTAGGAG	TTTTATGCATTGA ATAGAA	yes	yes
<i>cel-mir-85</i>	<i>Pmir-85</i>	Intragenic/antisense	II	TAAACAAGTGTA TTCCT	AGAATACGAACG ACGAGAAAA	yes	yes
<i>cel-mir-35</i>	<i>Pmir-35-36-37-38- 39-40-41</i>	Intragenic/antisense/ Operon	II	CCGTCATATCTG ACCCAC	AATAGTTGGGAA TGGAAATTAGG	yes	yes
<i>cel-mir-357</i>	<i>Pmir-358-357</i>	Intragenic/antisense/ Operon	V	AAACTACAGTAA TTCCTTAAATG	GAATGATTGGGA GAAAGGTGAAG	yes	yes
<i>cel-mir-358</i>	<i>Pmir-358-357</i>	Intragenic/antisense/ Operon	V	NA	NA	NA	NA
<i>cel-mir-36</i>	<i>Pmir-35-36-37-38- 39-40-41</i>	Intragenic/antisense/ Operon	II	NA	NA	NA	NA
<i>cel-mir-37</i>	<i>Pmir-35-36-37-38- 39-40-41</i>	Intragenic/antisense/ Operon	II	NA	NA	NA	NA
<i>cel-mir-38</i>	<i>Pmir-35-36-37-38- 39-40-41</i>	Intragenic/antisense/ Operon	II	NA	NA	NA	NA
<i>cel-mir-39</i>	<i>Pmir-35-36-37-38- 39-40-41</i>	Intragenic/antisense/ Operon	II	NA	NA	NA	NA
<i>cel-mir-40</i>	<i>Pmir-35-36-37-38- 39-40-41</i>	Intragenic/antisense/ Operon	II	NA	NA	NA	NA
<i>cel-mir-41</i>	<i>Pmir-35-36-37-38- 39-40-41</i>	Intragenic/antisense/ Operon	II	NA	NA	NA	NA
<i>cel-mir-788</i>	<i>Pmir-788</i>	Intergenic	X	TTCATTCACACT TAAATGTTTATTC	CTTCCAGCTGAA AATTTGTAATAC	yes	yes

NA: Not Applicable

<sup>a</sup>attB4 tail: aaaaacaactttatataaaaaatta

<sup>b</sup>attB1R tail: aaaaactactttttatatacaaaacttatcat

Chr: chromosome

Y1H inx: Y1H interactions after scoring

**Table II-3. Preys used in Y1H matrix experiments.**

Sequence name	Protein name	DNA-binding domain (wTF2.1)	Source	Sequence name	Protein name	DNA-binding domain (wTF2.1)	Source
T23D8.8	CFI-1	ARID/BRIG HT	screen	D1081.8	PHI-7	MYB - 2 domains	screen
C05D10.1	C05D10.1	AT Hook	screen	C32D5.1	C32D5.1	Novel	other studies
T05A7.4	HMG-11	AT Hook x3	other studies	C44F1.1	C44F1.1	Novel	screen
Y116A8C.22	Y116A8C.22	AT Hook x5	screen	C46H3.2	C46H3.2	Novel	screen
C44F1.2	C44F1.2	AT Hook, SAND	screen	C52G5.2	C52G5.2	Novel	screen
B0304.1	HLH-1	bHLH	screen	C54G4.7	C54G4.7	Novel	other studies
ZK682.4	HLH-10	bHLH	screen	C56C10.10	C56C10.10	Novel	screen
W02C12.3	HLH-30	bHLH	screen	T05G5.6	ECH-6	Novel	screen
T14F9.5	LIN-32	bHLH	screen	F08G12.3	F08G12.3	Novel	screen
F46G10.6	MXL-3	bHLH	screen	F43C11.7	F43C11.7	Novel	screen
F45E6.2	ATF-6	bZIP	screen	C34E10.1	GOP-3	Novel	screen
C48E7.11	C48E7.11	bZIP	screen	H02I12.5	H02I12.5	Novel	screen
T19E7.2	SKN-1	bZIP	other studies	Y51H1A.4	ING-3	Novel	screen
K02F3.4	ZIP-2	bZIP	screen	K02F3.10	K02F3.10	Novel	screen
F46F11.2	CEY-2	COLD BOX	screen	JC8.6	LIN-54	Novel	other studies
F45C12.2	F45C12.2	HD	screen	C34C6.6	PRX-5	Novel	screen
C17H12.9	CEH-48	HD - CUT	screen	R13.4	R13.4	Novel	screen
C08C3.3	MAB-5	HD - HOX	screen	R52.3	R52.3	Novel	screen
C38D4.6	PAL-1	HD - HOX	screen	T22F3.11	T22F3.11	Novel	screen
F46C8.5	CEH-14	HD - LIM	screen	Y17G7B.20	Y17G7B.20	Novel	screen
ZC247.3	LIN-11	HD - LIM	screen	Y38C9A.1	Y38C9A.1	Novel	screen
F29F11.5	CEH-22	HD - NK	mating	Y52E8A.2	Y52E8A.2	Novel	screen
F46F3.1	CEH-27	HD - NK	mating	Y53G8AR.2	Y53G8AR.2	Novel	screen
C39E6.4	MLS-2	HD - NK	screen	Y62E10A.14	Y62E10A.14	Novel	screen
F31E8.3	TAB-1	HD - NK	screen	Y66H1A.4	Y66H1A.4	Novel	mating
R08B4.2	ALR-1	HD - PRD	screen	ZC204.12	ZC204.12	Novel	screen
W03A3.1	CEH-10	HD - PRD	other studies	ZK287.6	ZK287.6	Novel	screen
D1007.1	CEH-17	HD - PRD	screen	F42G4.3	ZYX-1	Novel	screen
C37E2.5	CEH-37	HD - PRD	screen	C01H6.3	C01H6.3	Novel	screen
T13C5.4	T13C5.4	HD - PRD	screen	F26H9.2	F26H9.2	RPEL - 2 domains	screen
F26C11.2	UNC-4	HD - PRD	screen	F40H6.4	TBX-11	T-box	other studies
F58E6.10	UNC-42	HD - PRD	screen	F21H11.3	TBX-2	T-box	screen

ZK265.4	CEH-8	HD - PRD - 2 domains	screen	ZK829.5	TBX-36	T-box	mating
Y53C12C.1	Y53C12C.1	HD - PRD, Paired Domain - CPAX	screen	Y73F8A.17	TBX-40	T-box	screen
Y75B8A.1	PHP-3	HD-HOX	screen	T07C4.2	TBX-8	T-box	screen
H21P03.1	MBF-1	HTH	screen	F28B12.2	EGL-44	TEA/ATTS	screen
F25E2.5	DAF-3	MH1	screen	F53B2.6	HAM-1	WH	mating
F26B1.7	LET-381	WH - Fork Head	screen	C26E6.2	FLH-2	ZF - FLYWCH	screen
F25H8.6	F25H8.6	ZF - BED	screen	T14F9.4	PEB-1	ZF - FLYWCH	other studies
Y47D3B.9	Y47D3B.9	ZF - BED	other studies	Y11D7A.12	FLH-1	ZF - FLYWCH	screen
D2030.7	D2030.7	ZF - C2H2 - 1 finger	screen	C33D3.1	ELT-2	ZF - GATA	screen
F18A1.2	LIN-26	ZF - C2H2 - 1 finger	screen	K02B9.4	ELT-3	ZF - GATA	screen
F18A1.3	LIR-1	ZF - C2H2 - 1 finger	screen	C39B10.6	ELT-4	ZF - GATA	screen
T24C4.7	T24C4.7	ZF - C2H2 - 1 finger	screen	F58E10.5	END-3	ZF - GATA	screen
Y48A6C.3	Y48A6C.3	ZF - C2H2 - 1 finger	screen	R09G11.2	NHR-1	ZF - NHR	screen
Y48C3A.4	Y48C3A.4	ZF - C2H2 - 1 finger	screen	B0280.8	NHR-10	ZF - NHR	screen
C18D1.1	DIE-1	ZF - C2H2 - 3 fingers	screen	F44G3.9	NHR-111	ZF - NHR	screen
T27B1.2	T27B1.2	ZF - C2H2 - 3 fingers	screen	M02H5.7	NHR-123	ZF - NHR	screen
Y39B6A.46	Y39B6A.46	ZF - C2H2 - 3 fingers	screen	C56E10.4	NHR-137	ZF - NHR	screen
ZC395.8	ZTF-8	ZF - C2H2 - 3 fingers	screen	T01B10.4	NHR-14	ZF - NHR	screen
F13G3.1	ZTF-2	ZF - C2H2 - 3 fingers	screen	F25E5.6	NHR-141	ZF - NHR	screen
F43G9.11	CES-1	ZF - C2H2 - 4 fingers	screen	F56E3.4	NHR-29	ZF - NHR	screen
T07D10.3	T07D10.3	ZF - C2H2 - 4 fingers	screen	F58G6.5	NHR-34	ZF - NHR	screen
ZK867.1	TAG-239	ZF - C2H2 - 4 fingers	screen	C07A12.3	NHR-35	ZF - NHR	screen
F54F2.5	ZTF-1	ZF - C2H2 - 4 fingers	screen	Y104H12A.1	NHR-41	ZF - NHR	screen
F28F9.1	ZAG-1	ZF - C2H2 - 4 fingers, HD - ZFHD	screen	C29E6.5	NHR-43	ZF - NHR	screen
F49H12.1	LSY-2	ZF - C2H2 - 5 fingers	screen	C45E5.6	NHR-46	ZF - NHR	screen
F52F12.6	EKL-2	ZF - C2HC 2 fingers	screen	C48D5.1	NHR-6	ZF - NHR	screen



W02A2.7	MEX-5	ZF - CCCH - 2 domains	screen	W01D2.2	NHR-61	ZF - NHR	screen
Y49E10.14	PIE-1	ZF - CCCH - 2 domains	screen	T23H4.2	NHR-69	ZF - NHR	screen
F52E1.1	POS-1	ZF - CCCH - 2 domains	screen	T06C12.7	NHR-84	ZF - NHR	screen
C43H6.7	C43H6.7	ZF - DHHC	screen	Y40B10A.8	NHR-86	ZF - NHR	screen
T22E7.2	T22E7.2	ZF - DHHC	screen	Y15E3A.1	NHR-91	ZF - NHR	other studies
C27C12.6	DMD-4	ZF - DM	screen	M02H5.6	NHR-98	ZF - NHR	screen
F10C1.5	DMD-5	ZF - DM	screen	T18D3.2	ODR-7	ZF - NHR	screen

Table II-4. High-confidence TF=>PmiRNA interactions (score  $\geq 5$ ).

Promoter	Interactor	AD-cDNA	AD-TF	gp/dt	Rt H	Rt L	wTF2.1	SA H	SA L	Sticky H/L	M	Found multiple times	Found once form AD-TF library	Found with both libraries	Sum of weights	Score
<i>Pmir-48</i>	FLH-2	0	0	DT	2	2	1	1	1	0	0	0	0	0	14	5
<i>Plin-4</i>	F09B12.2	0	1	YES	NA	NA	1	0	1	3	0	0	0	0	10	5
<i>Pmir-48</i>	NHR-111	0	0	DT	2	2	1	1	1	0	0	0	0	0	14	5
<i>Plet-7</i>	T27B1.2	1	0	YES	0	NA	1	1	0	3	0	0	1	0	12	5
<i>Pmir-48</i>	TAB-1	0	1	YES	2	2	1	1	1	0	0	0	0	0	14	5
<i>Pmir-48</i>	UNC-42	0	1	YES	2	2	1	1	1	0	0	0	0	0	14	5
<i>Plin-4</i>	W05B10.2	0	1	YES	NA	NA	1	0	1	3	0	0	0	0	10	5
<i>Pmir-241</i>	ATF-6	1	0	YES	0	0	1	1	1	3	NA	0	1	0	13	5.38
<i>Pmir-241</i>	GOP-3	0	0	DT	2	0	0	1	1	3	NA	0	0	0	13	5.38
<i>Pmir-84</i>	C46H3.2	1	0	YES	2	0	0	1	0	3	NA	0	1	0	13	5.38
<i>Pmir-243</i>	CEH-17	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-77</i>	CEH-17	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-243</i>	PHI-7	1	0	YES	0	0	1	1	1	3	NA	0	1	0	13	5.38
<i>Pmir-228</i>	EGL-44	1	1	YES	2	0	1	1	1	0	NA	1	1	0	13	5.38
<i>Pmir-230</i>	END-3	0	1	YES	2	0	1	0	1	3	NA	0	0	0	13	5.38
<i>Pmir-231</i>	F26H9.2	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-238</i>	F26H9.2	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-245</i>	F26H9.2	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-252</i>	F26H9.2	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-271</i>	F26H9.2	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-77</i>	F26H9.2	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-81</i>	F26H9.2	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-72</i>	LIN-26	1	0	YES	NA	NA	1	1	1	3	NA	0	1	0	9	5.38
<i>Pmir-355</i>	MAB-5	1	0	YES	0	0	1	1	1	3	NA	0	1	0	13	5.38
<i>Pmir-268</i>	MXL-3	1	0	YES	0	0	1	1	1	3	NA	0	1	0	13	5.38
<i>Pmir-243</i>	NHR-111	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-355</i>	NHR-111	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-90</i>	NHR-29	1	0	YES	0	0	1	1	1	3	NA	0	1	0	13	5.38
<i>Pmir-234</i>	NHR-41	1	0	YES	0	0	1	1	1	3	NA	0	1	0	13	5.38
<i>Pmir-2</i>	TAB-1	10	0	YES	2	0	1	1	1	0	NA	1	1	0	13	5.38
<i>Pmir-271</i>	TAB-1	0	2	YES	2	0	1	1	1	0	NA	1	1	0	13	5.38
<i>Pmir-83</i>	TAB-1	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-269</i>	UNC-42	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.38
<i>Pmir-2</i>	Y17G7B.20	2	0	YES	0	0	0	1	1	3	NA	1	1	0	13	5.38
<i>Pmir-57</i>	Y38C9A.1	0	0	NA	2	2	0	0	0	3	NA	0	0	0	13	5.38
<i>Pmir-72</i>	ZC204.12	2	0	YES	0	0	0	1	1	3	NA	1	1	0	13	5.38
<i>Pmir-249</i>	ZTF-8	0	6	YES	0	2	1	1	1	0	NA	1	1	0	13	5.38
<i>Pmir-358-357</i>	ZTF-8	0	3	YES	0	2	1	1	1	0	NA	1	1	0	13	5.38
<i>Pmir-47</i>	ZTF-8	0	5	YES	0	2	1	1	1	0	NA	1	1	0	13	5.38
<i>Pmir-243</i>	ZIP-2	1	0	YES	0	0	1	1	1	3	NA	0	1	0	13	5.38
<i>Plet-7</i>	C32D5.1	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-259</i>	C32D5.1	0	0	NA	NA	2	0	0	1	3	NA	0	0	0	11	5.45

<i>Pmir-35-36</i>																
37-38-39-40-41	C32D5.1	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-76</i>	C32D5.1	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-79</i>	C32D5.1	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Plet-7</i>	GOP-3	0	0	DT	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-227-80</i>	CEH-48	1	0	YES	0	NA	1	1	0	3	NA	0	1	0	11	5.45
<i>Pmir-259</i>	EGL-44	2	0	YES	NA	2	1	0	1	0	NA	1	1	0	11	5.45
<i>Pmir-34</i>	ELT-3	1	0	NA	NA	0	1	0	1	3	NA	0	1	0	11	5.45
<i>Plet-7</i>	ZYX-1	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Plin-4</i>	ZYX-1	0	0	NA	NA	2	0	0	1	3	NA	0	0	0	11	5.45
<i>Pmir-259</i>	ZYX-1	0	0	NA	NA	2	0	0	1	3	NA	0	0	0	11	5.45
<i>Pmir-34</i>	ZYX-1	0	0	NA	NA	2	0	0	1	3	NA	0	0	0	11	5.45
<i>Pmir-247</i>	MLS-2	0	0	NA	2	NA	1	0	0	3	NA	0	0	0	11	5.45
<i>Pmir-247</i>	NHR-41	0	0	NA	2	NA	1	0	0	3	NA	0	0	0	11	5.45
<i>Pmir-85</i>	NHR-41	0	0	NA	2	NA	1	0	0	3	NA	0	0	0	11	5.45
<i>Pmir-85</i>	NHR-61	0	2	NO	0	NA	1	0	0	3	NA	1	1	0	11	5.45
<i>Pmir-57</i>	PAL-1	0	0	NA	2	NA	1	0	0	3	NA	0	0	0	11	5.45
<i>Pmir-257</i>	PEB-1	0	0	NA	NA	2	1	0	0	3	NA	0	0	0	11	5.45
<i>Pmir-57</i>	PHP-3	0	0	NA	2	NA	1	0	0	3	NA	0	0	0	11	5.45
<i>Pmir-84</i>	POS-1	1	0	YES	NA	NA	1	1	0	3	0	0	1	0	11	5.45
<i>Pmir-240</i>	Y38C9A.1	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-258</i>	Y38C9A.1	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-259</i>	Y38C9A.1	0	0	NA	NA	2	0	0	1	3	NA	0	0	0	11	5.45
<i>Pmir-34</i>	Y38C9A.1	0	0	NA	NA	2	0	0	1	3	NA	0	0	0	11	5.45
<i>Pmir-35-36</i>																
37-38-39-40-41	Y38C9A.1	0	0	YES	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-46</i>	Y38C9A.1	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-60</i>	Y38C9A.1	0	0	NA	NA	2	0	0	1	3	NA	0	0	0	11	5.45
<i>Pmir-75</i>	Y38C9A.1	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-46</i>	ING-3	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-240</i>	Y52E8A.2	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-227-80</i>	Y62E10A.14	2	0	YES	0	NA	0	1	0	3	NA	1	1	0	11	5.45
<i>Pmir-79</i>	Y62E10A.14	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Plet-7</i>	Y66H1A.4	0	0	DT	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Pmir-57</i>	ZAG-1	0	0	NA	2	NA	1	0	0	3	NA	0	0	0	11	5.45
<i>Pmir-76</i>	ZC204.12	0	0	NA	2	NA	0	1	0	3	NA	0	0	0	11	5.45
<i>Plin-4</i>	DNJ-11	0	1	YES	NA	NA	1	0	1	3	NA	0	0	0	9	5.56
<i>Pmir-63</i>	ECH-6	0	1	YES	NA	NA	0	1	1	3	NA	0	0	0	9	5.56
<i>Pmir-241</i>	Y94H6A.10	0	1	YES	NA	NA	0	1	1	3	NA	0	0	0	9	5.56
<i>Pmir-241</i>	FLH-2	0	0	M	2	2	1	1	1	0	1	0	0	0	14	5.71
<i>Pmir-48</i>	CEH-17	0	0	M	2	2	1	1	1	0	1	0	0	0	14	5.71
<i>Pmir-84</i>	CEH-22	0	0	M	0	2	1	1	0	3	1	0	0	0	14	5.71
<i>Pmir-48</i>	CEH-27	0	0	DT	0	2	1	1	1	3	0	0	0	0	14	5.71
<i>Pmir-241</i>	NHR-35	12	0	YES	0	0	1	1	1	3	0	1	1	0	14	5.71
<i>Pmir-84</i>	R13.4	0	0	M	0	2	1	1	0	3	1	0	0	0	14	5.71

<i>Plin-4</i>	FLH-2	0	0	M	2	2	1	0	1	0	1	0	0	0	12	5.83
<i>Plet-7</i>	PHI-7	0	0	DT	2	NA	1	1	0	3	0	0	0	0	12	5.83
<i>Plin-4</i>	DIE-1	0	0	NA	NA	2	1	0	1	3	0	0	0	0	12	5.83
<i>Plin-4</i>	DMD-5	0	0	DT	NA	2	1	0	1	3	NA	0	0	0	12	5.83
<i>Plet-7</i>	R13.4	0	0	DT	2	NA	1	1	0	3	0	0	0	0	12	5.83
<i>Plet-7</i>	W05B10.2	0	0	DT	2	NA	1	1	0	3	0	0	0	0	12	5.83
<i>Pmir-241</i>	LIN-26	0	1	YES	NA	NA	1	1	1	3	0	0	0	0	10	6
<i>Pmir-84</i>	MLS-2	1	0	YES	NA	NA	1	1	0	3	0	0	1	0	10	6
<i>Pmir-237</i>	T24C4.7	0	1	YES	NA	NA	1	1	1	3	0	0	0	0	10	6
<i>Pmir-267</i>	C32D5.1	0	0	NA	2	2	0	0	1	3	NA	0	0	0	13	6.15
<i>Pmir-46</i>	C32D5.1	0	0	NA	2	2	0	1	0	3	NA	0	0	0	13	6.15
<i>Pmir-63</i>	CFI-1	2	0	YES	0	0	1	1	1	3	NA	1	1	0	13	6.15
<i>Pmir-237</i>	DAF-3	0	0	DT	2	0	1	1	1	3	0	0	0	0	13	6.15
<i>Pmir-42-43</i> <i>44</i>	DIE-1	6	0	YES	0	0	1	1	1	3	NA	1	1	0	13	6.15
<i>Pmir-90</i>	DIE-1	0	1	YES	2	0	1	1	1	3	NA	0	0	0	13	6.15
<i>Pmir-84</i>	ZYX-1	0	0	NA	2	2	0	1	0	3	NA	0	0	0	13	6.15
<i>Pmir-48</i>	MIG-5	0	0	DT	0	2	1	1	1	3	NA	0	0	0	13	6.15
<i>Pmir-252</i>	TAB-1	7	5	YES	2	0	1	1	1	0	NA	1	1	1	13	6.15
<i>Pmir-53</i>	FLH-1	0	0	NA	2	0	1	1	1	3	NA	0	0	0	13	6.15
<i>Plin-4</i>	Y38C9A.1	0	0	NA	2	2	0	0	1	3	NA	0	0	0	13	6.15
<i>Pmir-76</i>	Y38C9A.1	0	0	NA	2	2	0	1	0	3	NA	0	0	0	13	6.15
<i>Pmir-392</i>	Y53C12C.1	0	1	YES	2	0	1	1	1	3	NA	0	0	0	13	6.15
<i>Pmir-47</i>	ZC204.12	7	3	YES	0	0	0	1	1	3	NA	1	1	1	13	6.15
<i>Pmir-243</i>	ZTF-10	0	6	YES	0	0	1	1	1	3	NA	1	1	0	13	6.15
<i>Pmir-259</i>	CEH-27	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	CEH-31	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-259</i>	CEH-37	0	2	YES	NA	0	1	0	1	3	NA	1	1	0	11	6.36
<i>Pmir-76</i>	CEH-48	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-259</i>	CEH-8	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-60</i>	CEH-8	0	1	YES	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	CEH-9	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-240</i>	DAF-3	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-258</i>	DAF-3	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-35-36</i> <i>37-38-39-</i> <i>40-41</i>	DAF-3	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-258</i>	DIE-1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-259</i>	DIE-1	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Plet-7</i>	HAM-1	0	0	DT	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-258</i>	LIR-1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-76</i>	LSY-2	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-75</i>	MIG-5	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	NHR-273	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-76</i>	NHR-43	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-59</i>	NHR-6	4	0	YES	0	NA	1	1	0	3	NA	1	1	0	11	6.36
<i>Pmir-76</i>	NHR-86	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-240</i>	ODR-7	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-258</i>	ODR-7	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	ODR-7	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36

<i>Pmir-76</i>	ODR-7	0	1	YES	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-76</i>	PEB-1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-259</i>	PHP-3	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-59</i>	PHP-3	0	9	YES	0	NA	1	1	0	3	NA	1	1	0	11	6.36
<i>Pmir-76</i>	PHP-3	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-79</i>	PHP-3	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-227-80</i>	PIE-1	0	2	YES	0	NA	1	1	0	3	NA	1	1	0	11	6.36
<i>Pmir-34</i>	R13.4	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-76</i>	T27B1.2	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	TBX-11	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-79</i>	TBX-39	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-34</i>	TBX-8	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	TBX-8	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-227-80</i>	HLH-30	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-34</i>	HLH-30	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	Y48A6C.3	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	Y53C12C.1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-76</i>	Y53C12C.1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	ZIP-1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	ZIP-4	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Plet-7</i>	ZTF-1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Plin-4</i>	ZTF-1	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-227-80</i>	ZTF-1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-258</i>	ZTF-1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-259</i>	ZTF-1	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-34</i>	ZTF-1	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	ZTF-1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-76</i>	ZTF-1	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-46</i>	ZTF-2	0	0	NA	2	NA	1	1	0	3	NA	0	0	0	11	6.36
<i>Pmir-241</i>	CEH-27	0	0	M	0	2	1	1	1	3	1	0	0	0	14	6.43
<i>Pmir-84</i>	DAF-3	0	0	DT	2	2	1	1	0	3	0	0	0	0	14	6.43
<i>Pmir-241</i>	NHR-111	0	4	YES	2	2	1	1	1	0	0	1	1	0	14	6.43
<i>Plin-4</i>	ODR-7	0	0	NA	2	2	1	1	1	3	NA	0	0	0	10	6.57
<i>Plin-4</i>	PQM-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	10	6.57
<i>Pmir-54-5556</i>	D2030.7	2	0	YES	NA	NA	1	0	0	3	NA	1	1	0	9	6.67
<i>Plin-4</i>	DAF-3	0	0	M	NA	2	1	0	1	3	1	0	0	0	12	6.67
<i>Pmir-237</i>	DMD-5	0	0	DT	0	2	1	1	1	3	NA	0	0	0	12	6.67
<i>Pmir-255</i>	ECH-6	1	0	NO	NA	NA	0	1	1	3	NA	0	1	0	9	6.67
<i>Pmir-252</i>	F09B12.2	0	1	NO	NA	NA	1	1	1	3	NA	0	0	0	9	6.67
<i>Pmir-230</i>	ZYX-1	0	0	NA	NA	2	0	0	1	3	NA	0	0	0	9	6.67
<i>Plsy-6</i>	HMG-12	0	1	NO	NA	NA	1	1	1	3	NA	0	0	0	9	6.67
<i>Plin-4</i>	MIG-5	0	0	DT	NA	2	1	0	1	3	1	0	0	0	12	6.67
<i>Pmir-230</i>	PRX-5	2	0	YES	NA	0	0	0	1	3	NA	1	1	0	9	6.67
<i>Pmir-59</i>	R06C7.9	1	0	YES	NA	NA	1	1	0	3	NA	0	1	0	9	6.67
<i>Pmir-230</i>	Y38C9A.1	0	0	NA	NA	2	0	0	1	3	NA	0	0	0	9	6.67
<i>Pmir-83</i>	Y39B6A.46	0	1	YES	NA	NA	1	1	1	3	NA	0	0	0	9	6.67

<i>Pmir-63</i>	Y54E10BR.8	0	1	YES	NA	NA	1	1	1	3	NA	0	0	0	9	6.67
<i>Pmir-234</i>	C32D5.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-355</i>	C32D5.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-392</i>	C32D5.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-63</i>	C32D5.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-83</i>	C32D5.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-243</i>	C44F1.2	1	0	YES	0	2	1	1	1	3	NA	0	1	0	13	6.92
<i>Pmir-77</i>	C46H3.2	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-237</i>	CEH-8	0	0	M	0	2	1	1	1	3	1	0	0	0	13	6.92
<i>Pmir-227-80</i>	DAF-3	0	0	NA	2	2	1	1	0	3	NA	0	0	0	13	6.92
<i>Pmir-243</i>	EGL-44	0	5	YES	2	2	1	1	1	0	NA	1	1	0	13	6.92
<i>Pmir-48</i>	F08G12.3	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-262</i>	ZYX-1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-48</i>	ZYX-1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-63</i>	ZYX-1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-48</i>	F43C11.7	0	0	DT	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-46</i>	MIG-5	0	0	NA	2	2	1	1	0	3	NA	0	0	0	13	6.92
<i>Pmir-52</i>	NHR-86	0	0	NA	2	2	1	1	0	3	NA	0	0	0	13	6.92
<i>Pmir-52</i>	ODR-7	0	0	NA	2	2	1	1	0	3	NA	0	0	0	13	6.92
<i>Pmir-46</i>	PAX-3	0	0	NA	2	2	1	1	0	3	NA	0	0	0	13	6.92
<i>Pmir-232</i>	PRX-5	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-261</i>	PRX-5	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-75</i>	TBX-39	0	0	NA	2	2	1	1	0	3	NA	0	0	0	13	6.92
<i>Pmir-46</i>	TBX-9	0	0	NA	2	2	1	1	0	3	NA	0	0	0	13	6.92
<i>Plet-7</i>	Y38C9A.1	1	0	YES	2	2	0	1	0	3	NA	0	1	0	13	6.92
<i>Pmir-2</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-228</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-232</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-234</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-241</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-242</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-243</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-261</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-355</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-42-4344</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-63</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-81</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-83</i>	Y38C9A.1	0	0	NA	2	2	0	1	1	3	NA	0	0	0	13	6.92
<i>Pmir-84</i>	Y38C9A.1	1	0	YES	2	2	0	1	0	3	NA	0	1	0	13	6.92
<i>Pmir-84</i>	ZTF-1	0	0	NA	2	2	1	1	0	3	NA	0	0	0	13	6.92
<i>Pmir-48</i>	POS-1	1	0	YES	NA	NA	1	1	1	3	0	0	1	0	10	7
<i>Pmir-48</i>	DAF-3	0	0	DT	2	2	1	1	1	3	0	0	0	0	14	7.14
<i>Pmir-48</i>	LSY-2	0	0	NA	2	2	1	1	1	3	0	0	0	0	14	7.14
<i>Pmir-237</i>	NHR-43	0	0	DT	2	2	1	1	1	3	0	0	0	0	14	7.14
<i>Pmir-84</i>	NHR-43	2	0	YES	2	0	1	1	0	3	1	1	1	0	14	7.14
<i>Pmir-241</i>	ODR-7	0	0	NA	2	2	1	1	1	3	0	0	0	0	14	7.14
<i>Pmir-48</i>	T13C5.4	0	1	YES	2	2	1	1	1	3	0	0	0	0	14	7.14

<i>Pmir-241</i>	TBX-11	0	0	DT	2	2	1	1	1	3	0	0	0	0	14	7.14
<i>Pmir-51</i>	DMD-4	3	4	YES	0	NA	1	1	0	3	NA	1	1	1	11	7.27
<i>Pmir-258</i>	NHR-86	1	0	YES	2	NA	1	1	0	3	NA	0	1	0	11	7.27
<i>Pmir-227-80</i>	ODR-7	1	0	YES	2	NA	1	1	0	3	NA	0	1	0	11	7.27
<i>Pmir-259</i>	ODR-7	0	12	YES	NA	2	1	0	1	3	NA	0	1	0	11	7.27
<i>Plin-4</i>	FLH-1	1	0	YES	NA	2	1	0	1	3	NA	0	1	0	11	7.27
<i>Pmir-227-80</i>	Y38C9A.1	0	0	NA	2	2	0	1	0	3	NA	0	0	0	11	7.27
<i>Pmir-77</i>	ELT-7	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-237</i>	CEH-22	0	0	DT	2	2	1	1	1	3	0	0	0	0	13	7.69
<i>Pmir-238</i>	CEH-27	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-245</i>	CEH-27	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-251</i>	CEH-27	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-45</i>	CEH-27	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-392</i>	CEH-8	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-228</i>	DAF-3	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-261</i>	DAF-3	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-355</i>	DAF-3	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-261</i>	DIE-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-81</i>	DIE-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-77</i>	ELT-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-355</i>	ELT-2	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-77</i>	ELT-2	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-77</i>	ELT-3	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-77</i>	ELT-6	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-77</i>	END-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-77</i>	END-3	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-77</i>	F38C2.5	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-392</i>	ZYX-1	1	0	YES	2	2	0	1	1	3	NA	0	1	0	13	7.69
<i>Pmir-242</i>	LSY-2	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-392</i>	MAB-5	5	0	YES	0	2	1	1	1	3	NA	1	1	0	13	7.69
<i>Pmir-63</i>	MBF-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-245</i>	MTCE.26	1	0	YES	2	2	0	1	1	3	NA	0	1	0	13	7.69
<i>Pmir-355</i>	NHR-10	0	1	YES	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-90</i>	NHR-34	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-261</i>	NHR-43	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-239.a</i>	NHR-86	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-72</i>	NHR-86	0	1	YES	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-232</i>	ODR-7	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-243</i>	ODR-7	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-262</i>	ODR-7	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-241</i>	PEB-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-63</i>	PHP-3	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-48</i>	SDZ-12	0	0	DT	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-77</i>	SOX-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-77</i>	T09A5.12	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-228</i>	TBX-11	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-231</i>	TBX-11	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69

<i>Pmir-269</i>	TBX-39	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-228</i>	TBX-8	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-231</i>	TBX-8	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-243</i>	HLH-30	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-228</i>	Y17G7B.20	1	0	YES	2	2	0	1	1	3	NA	0	1	0	13	7.69
<i>Pmir-392</i>	Y38C9A.1	1	0	YES	2	2	0	1	1	3	NA	0	1	0	13	7.69
<i>Pmir-48</i>	Y38C9A.1	1	0	YES	2	2	0	1	1	3	NA	0	1	0	13	7.69
<i>Pmir-228</i>	ZTF-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-234</i>	ZTF-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-261</i>	ZTF-1	0	1	YES	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-355</i>	ZTF-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-392</i>	ZTF-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-47</i>	ZTF-1	0	3	YES	2	0	1	1	1	3	NA	1	1	0	13	7.69
<i>Pmir-48</i>	ZTF-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-63</i>	ZTF-1	0	1	YES	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-83</i>	ZTF-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-42-43</i> <i>44</i>	ZTF-2	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-48</i>	ZTF-2	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.69
<i>Pmir-358-</i> <i>357</i>	FLH-1	11	0	YES	0	2	1	1	1	3	NA	1	1	0	13	7.69
<i>Pmir-2</i>	C52G5.2	3	0	YES	NA	NA	0	1	1	3	NA	1	1	0	9	7.78
<i>Pmir-230</i>	CEH-27	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	9	7.78
<i>Pmir-230</i>	DAF-3	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	9	7.78
<i>Pmir-230</i>	DIE-1	0	0	NA	NA	2	1	0	1	3	NA	0	0	0	9	7.78
<i>Pmir-243</i>	ELT-4	1	0	YES	NA	NA	1	1	1	3	NA	0	1	0	9	7.78
<i>Pmir-257</i>	PAL-1	5	5	YES	NA	NA	1	0	0	3	NA	1	1	1	9	7.78
<i>Pmir-231</i>	TBX-9	1	0	YES	NA	NA	1	1	1	3	NA	0	1	0	9	7.78
<i>Pmir-243</i>	Y116A8C.22	1	0	YES	NA	NA	1	1	1	3	NA	0	1	0	9	7.78
<i>Pmir-48</i>	MLS-2	1	0	YES	2	2	1	1	1	3	0	0	1	0	14	7.86
<i>Pmir-231</i>	CES-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	5	7.86
<i>Plet-7</i>	DMD-5	3	0	YES	2	NA	1	1	0	3	NA	1	1	0	11	8.18
<i>Plet-7</i>	DAF-3	2	2	YES	2	NA	1	1	0	3	0	1	1	1	12	8.33
<i>Pmir-90</i>	CEH-48	1	0	YES	2	2	1	1	1	3	NA	0	1	0	13	8.46
<i>Pmir-231</i>	DIE-1	1	0	YES	2	2	1	1	1	3	NA	0	1	0	13	8.46
<i>Pmir-243</i>	DIE-1	1	0	YES	2	2	1	1	1	3	NA	0	1	0	13	8.46
<i>Pmir-228</i>	LSY-2	4	11	YES	0	2	1	1	1	3	NA	1	1	1	13	8.46
<i>Pmir-355</i>	NHR-1	1	0	YES	2	2	1	1	1	3	NA	0	1	0	13	8.46
<i>Pmir-355</i>	NHR-43	1	0	YES	2	2	1	1	1	3	NA	0	1	0	13	8.46
<i>Pmir-243</i>	PIE-1	1	0	YES	2	2	1	1	1	3	NA	0	1	0	13	8.46
<i>Pmir-355</i>	PRX-5	24	0	YES	2	2	0	1	1	3	NA	1	1	0	13	8.46
<i>Pmir-231</i>	TBX-40	1	0	YES	2	2	1	1	1	3	NA	0	1	0	13	8.46
<i>Pmir-243</i>	ZTF-1	1	0	YES	2	2	1	1	1	3	NA	0	1	0	13	8.46
<i>Plin-4</i>	NHR-43	2	0	YES	2	2	1	0	1	3	1	1	1	0	14	8.57
<i>Pmir-241</i>	NHR-69	4	0	YES	2	2	1	1	1	3	0	1	1	0	14	8.57
<i>Pmir-48</i>	ZAG-1	2	0	YES	2	2	1	1	1	3	0	1	1	0	14	8.57
<i>Plet-7</i>	C44F1.1	1	2	YES	NA	NA	1	1	0	3	NA	1	1	1	9	8.89
<i>Pmir-236</i>	LIN-26	9	1	YES	NA	NA	1	1	0	3	NA	1	1	1	9	8.89
<i>Pmir-63</i>	LIN-26	2	0	YES	NA	NA	1	1	1	3	NA	1	1	0	9	8.89



<i>Pmir-355</i>	NHR-14	2	0	YES	NA	NA	1	1	1	3	NA	1	1	0	9	8.89
<i>Pmir-392</i>	PAL-1	13	0	YES	NA	NA	1	1	1	3	NA	1	1	0	9	8.89
<i>Pmir-59</i>	FLH-1	2	5	YES	2	NA	1	1	0	3	NA	1	1	1	11	9.09
<i>Pmir-267</i>	DIE-1	22	4	YES	2	2	1	0	1	3	NA	1	1	1	13	9.23
<i>Pmir-243</i>	LSY-2	0	21	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-355</i>	NHR-34	0	2	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-358-357</i>	NHR-43	0	2	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-355</i>	NHR-46	0	6	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-261</i>	ODR-7	0	9	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-358-357</i>	ODR-7	0	10	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-83</i>	ODR-7	0	13	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-392</i>	PHP-3	0	9	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-241</i>	TBX-8	2	0	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-355</i>	TBX-8	2	0	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-48</i>	FLH-1	3	0	YES	2	2	1	1	1	3	NA	1	1	0	13	9.23
<i>Pmir-48</i>	Y17G7B.20	4	0	YES	2	2	0	1	1	3	NA	1	1	1	13	9.23
<i>Pmir-48</i>	NHR-34	0	2	YES	2	2	1	1	1	3	1	1	1	0	14	9.29
<i>Pmir-237</i>	NHR-46	5	6	YES	2	2	1	1	1	3	0	1	1	1	14	9.29
<i>Pmir-48</i>	NHR-46	0	2	YES	2	2	1	1	1	3	1	1	1	0	14	9.29
<i>Pmir-241</i>	FLH-1	6	0	YES	2	2	1	1	1	3	1	1	1	0	14	9.29
<i>Pmir-235</i>	CES-1	7	20	YES	2	2	1	1	1	3	NA	1	1	1	13	10
<i>Pmir-63</i>	DIE-1	5	15	YES	2	2	1	1	1	3	NA	1	1	1	13	10
<i>Pmir-355</i>	DMD-5	1	5	YES	2	2	1	1	1	3	NA	1	1	1	13	10
<i>Pmir-247</i>	IRX-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	10	10
<i>Pmir-42-43-44</i>	LIN-26	6	28	YES	NA	NA	1	1	1	3	NA	1	1	1	9	10
<i>Pmir-270</i>	ODR-7	0	0	NA	2	2	1	1	1	3	NA	0	0	0	10	10
<i>Pmir-241</i>	DAF-3	0	0	DT	0	2	1	1	1	3	0	0	0	0	14	5.71
<i>Pmir-270</i>	TBX-8	0	0	NA	2	2	1	1	1	3	NA	0	0	0	10	10

NA: Not applicable

gp/dt: gap repair/ direct transformation

SA: Self-activation

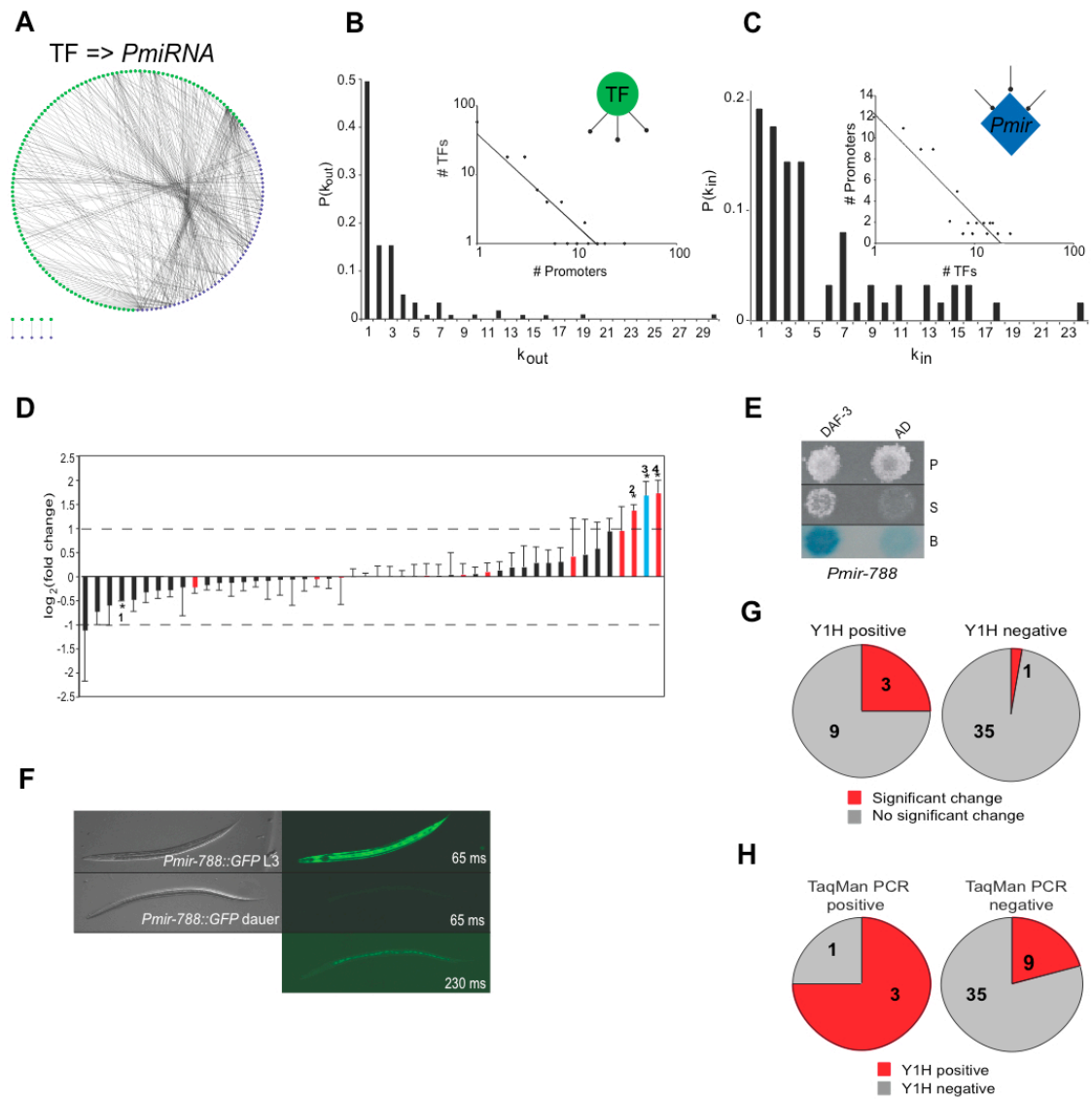
M: Mating

Rt: Re-test

H: *HIS3*

L: *LacZ*

Figure II-1



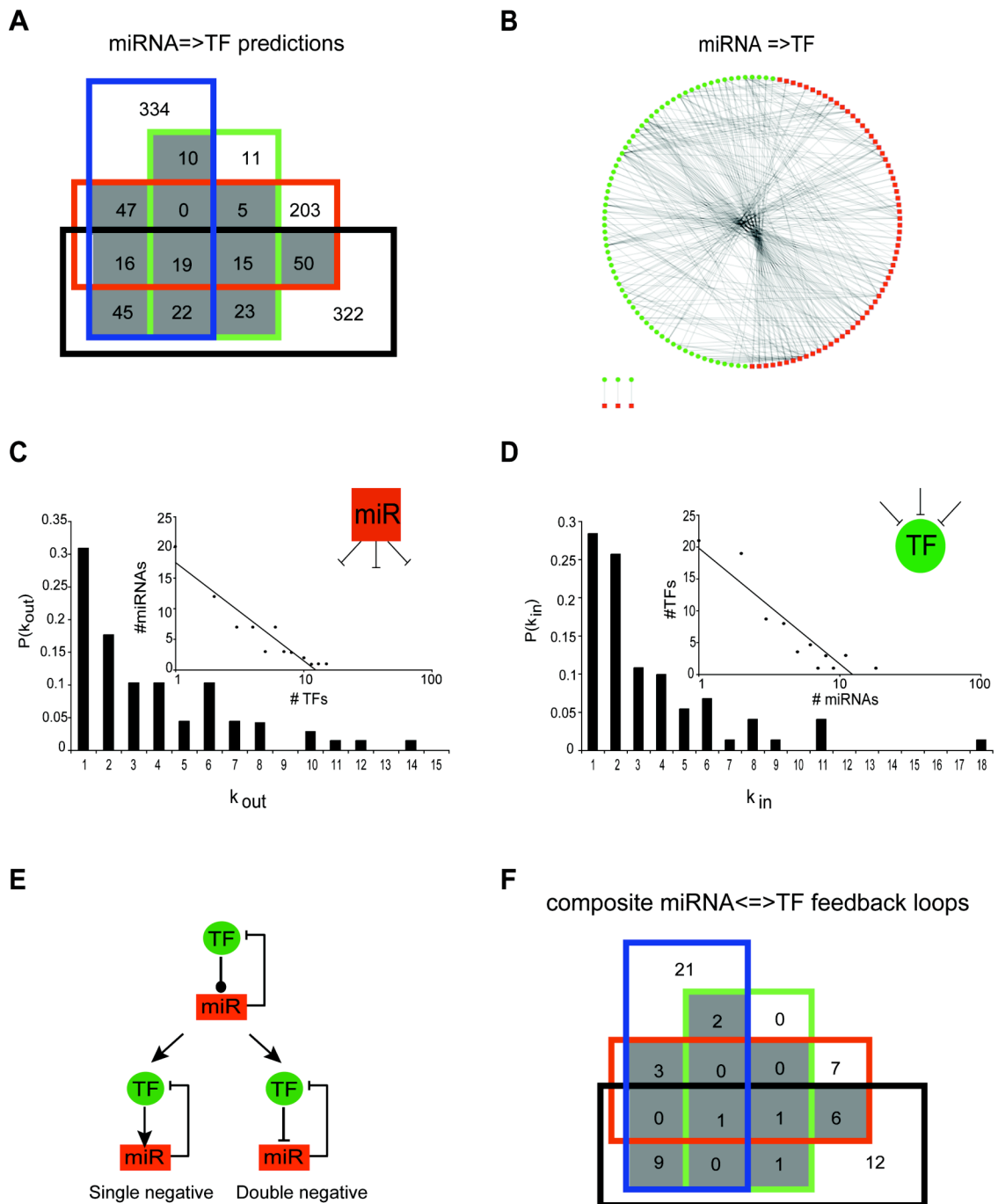
**Figure II-1. A Genome-Scale *C. elegans* TF=>miRNA transcription regulatory network.** (A) TF=>miRNA interactions identified by high-throughput Y1H assays were visualized into a transcription regulatory network using Cytoscape (Shannon et al. 2003). Blue diamonds – miRNA promoters; green circles – TFs. (B) Out-degree;  $P(k_{out})$  is the proportion of miRNA promoters per TF. The out-degree distribution best fits a power law ( $R^2 = 0.82$ , inset). (C) In-degree;  $P(k_{in})$  is the proportion of TFs per miRNA promoter. The in-degree best fits an exponential distribution ( $R^2 = 0.84$ , inset). (D) TaqMan PCR assays of 48 miRNAs in N2 and *daf-3(mgDf90)* mutant dauer larvae. The average  $\log_2$ (fold change) of five experiments is shown. Error bars indicate standard error of the mean. Asterisks indicate significant changes. 1 – *mir-85*, 2 – *mir-48*; 3 – *mir-788*; 4 – *mir-241*. Red bars – miRNAs bound by DAF-3 in Y1H assays. Blue bar – *mir-788*. The dashed line indicates a 2-fold difference. (E) Y1H assay confirming the interaction between DAF-3 and *Pmir-788*. P – permissive media; S – selective media; B –  $\beta$ -Galactosidase assay; AD – empty vector. (F) *mir-788* is repressed in dauer larvae. Left – Nomarski image; Right – GFP fluorescence. The top right and right middle panels are 65 ms exposures, whereas the bottom right panel is a 230 ms exposure of the same field as in the right middle panel to visualize the presence of the animal. (G, H) Correlation between Y1H and TaqMan PCR data. Both the proportion and the actual numbers are depicted.

**Table II-5.** *TF=>Pmir-788 interactions that were not included in network analyses.*

Table S5. TF=>Pmir-788 interactions that were not included in the network analysis

Promoter	Interactor	AD-cDNA	AD-TF	gp/dt	Retest HIS3	Retest LacZ	wTF2.1	SA HIS3	SA LacZ	Sticky HIS3/lacZ	Mating	Found multiple times	Found once form AD-TF library	Found with both libraries	Sum of all weights	Score
<i>Pmir-788</i>	IRX-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.692
<i>Pmir-788</i>	ZTF-1	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.692
<i>Pmir-788</i>	NHR-2	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.692
<i>Pmir-788</i>	NHR-111	0	0	NA	2	2	1	1	1	0	NA	0	0	0	13	5.385
<i>Pmir-788</i>	DAF-3	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.692
<i>Pmir-788</i>	ODR-7	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.692
<i>Pmir-788</i>	TBX-8	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.692
<i>Pmir-788</i>	T09A5.12	0	0	NA	2	2	1	1	1	3	NA	0	0	0	13	7.692

Figure II-2



**Figure II-2.** A *C. elegans* miRNA=>TF post-transcription regulatory network. (A) Four-way Venn diagram showing the number of miRNA=>TF predictions for TFs found in the transcriptional network. Blue – RNA-hybrid; green – Pictar; red – miRanda and black – TargetScanS. Grey – predictions common in two or more algorithms. (B) The predicted miRNA=>TF post-transcription regulatory network. Red squares – miRNAs; green circles – TFs. (C) Out-degree;  $P(k_{out})$  is the proportion of TF targets per miRNA. The out-degree best fits an exponential distribution ( $R^2 = 0.90$ , inset). (D) In-degree;  $P(k_{in})$  is the proportion of miRNA targeting a TF. The in-degree best fits an exponential distribution ( $R^2 = 0.84$ , inset). (E, top) Cartoon of the two types of composite miRNA<=>TF feedback loops: single negative (or incoherent) (bottom left) and double negative (or coherent) (bottom right). Line with dot – physical interaction; blunt arrow – repression. (F) Four-way Venn diagram showing the number of composite miRNA<=>TF feedback loops obtained after network integration. Grey – loops common in two or more algorithms.

**Table II-6.** MiRNA target predictions for TF genes found in Y1H assays and common in two or more prediction algorithms.

miRNA	Target ORF name	Target gene name	miRNA	Target ORF name	Target gene name
<i>mir-359</i>	B0280.8	<i>nhr-10</i>	<i>mir-55</i>	C45E5.6	<i>nhr-46</i>
<i>mir-72</i>	B0280.8	<i>nhr-10</i>	<i>mir-79</i>	C45E5.6	<i>nhr-46</i>
<i>mir-250</i>	C07A12.3	<i>nhr-35</i>	<i>let-7</i>	C48D5.1	<i>nhr-6</i>
<i>mir-2</i>	C07A12.3	<i>nhr-35</i>	<i>mir-268</i>	C48D5.1	<i>nhr-6</i>
<i>mir-43</i>	C07A12.3	<i>nhr-35</i>	<i>mir-85</i>	C48D5.1	<i>nhr-6</i>
<i>mir-52</i>	C07A12.3	<i>nhr-35</i>	<i>mir-90</i>	C48D5.1	<i>nhr-6</i>
<i>mir-228</i>	C07A12.3	<i>nhr-35</i>	<i>mir-228</i>	C52G5.2	
<i>mir-80</i>	C08C3.3	<i>mab-5</i>	<i>mir-80</i>	D1007.1	<i>ceh-17</i>
<i>mir-81</i>	C08C3.3	<i>mab-5</i>	<i>mir-81</i>	D1007.1	<i>ceh-17</i>
<i>mir-80</i>	C17H12.9	<i>ceh-48</i>	<i>mir-230</i>	D2030.7	
<i>mir-81</i>	C17H12.9	<i>ceh-48</i>	<i>mir-2</i>	F08G12.3	
<i>let-7</i>	C18D1.1	<i>die-1</i>	<i>mir-43</i>	F08G12.3	
<i>mir-244</i>	C18D1.1	<i>die-1</i>	<i>mir-60</i>	F08G12.3	
<i>mir-241</i>	C18D1.1	<i>die-1</i>	<i>mir-85</i>	F10C1.5	<i>dmd-5</i>
<i>mir-48</i>	C18D1.1	<i>die-1</i>	<i>mir-357</i>	F13G3.1	<i>ztf-2</i>
<i>mir-80</i>	C18D1.1	<i>die-1</i>	<i>mir-48</i>	F13G3.1	<i>ztf-2</i>
<i>mir-81</i>	C18D1.1	<i>die-1</i>	<i>mir-236</i>	F18A1.2	<i>lin-26</i>
<i>mir-80</i>	C26E6.2	<i>flh-2</i>	<i>mir-357</i>	F18A1.2	<i>lin-26</i>
<i>mir-81</i>	C26E6.2	<i>flh-2</i>	<i>mir-2</i>	F18A1.2	<i>lin-26</i>
<i>let-7</i>	C29E6.5	<i>nhr-43</i>	<i>mir-43</i>	F18A1.2	<i>lin-26</i>
<i>mir-241</i>	C29E6.5	<i>nhr-43</i>	<i>mir-51</i>	F18A1.2	<i>lin-26</i>
<i>mir-48</i>	C29E6.5	<i>nhr-43</i>	<i>mir-52</i>	F18A1.2	<i>lin-26</i>
<i>mir-76</i>	C29E6.5	<i>nhr-43</i>	<i>mir-54</i>	F18A1.2	<i>lin-26</i>
<i>mir-80</i>	C29E6.5	<i>nhr-43</i>	<i>mir-55</i>	F18A1.2	<i>lin-26</i>
<i>mir-84</i>	C29E6.5	<i>nhr-43</i>	<i>mir-56</i>	F18A1.2	<i>lin-26</i>
<i>mir-236</i>	C32E12.5	<i>sox-1</i>	<i>mir-60</i>	F18A1.2	<i>lin-26</i>
<i>mir-75</i>	C32E12.5	<i>sox-1</i>	<i>mir-72</i>	F18A1.2	<i>lin-26</i>
<i>mir-79</i>	C32E12.5	<i>sox-1</i>	<i>mir-239</i>	F18A1.3	<i>lir-1</i>
<i>mir-235</i>	C33D12.1	<i>ceh-31</i>	<i>mir-34</i>	F18A1.3	<i>lir-1</i>
<i>lsy-6</i>	C33D3.1	<i>elt-2</i>	<i>let-7</i>	F25E2.5	<i>daf-3</i>
<i>mir-246</i>	C36F7.1	<i>irx-1</i>	<i>mir-241</i>	F25E2.5	<i>daf-3</i>
<i>mir-46</i>	C36F7.1	<i>irx-1</i>	<i>lsy-6</i>	F25E2.5	<i>daf-3</i>
<i>mir-47</i>	C36F7.1	<i>irx-1</i>	<i>mir-48</i>	F25E2.5	<i>daf-3</i>
<i>mir-34</i>	C38D4.6	<i>pal-1</i>	<i>mir-63</i>	F25E2.5	<i>daf-3</i>
<i>mir-85</i>	C38D4.6	<i>pal-1</i>	<i>mir-75</i>	F25E2.5	<i>daf-3</i>
<i>mir-235</i>	C39E6.4	<i>mls-2</i>	<i>mir-84</i>	F25E2.5	<i>daf-3</i>
<i>mir-63</i>	C39E6.4	<i>mls-2</i>	<i>mir-231</i>	F25E2.5	<i>daf-3</i>
<i>mir-241</i>	C44F1.1		<i>mir-241</i>	F27E5.2	<i>pax-3</i>
<i>mir-243</i>	C44F1.2		<i>mir-48</i>	F27E5.2	<i>pax-3</i>
<i>mir-51</i>	C45E5.6	<i>nhr-46</i>	<i>mir-34</i>	F28B12.2	<i>egl-44</i>
<i>mir-52</i>	C45E5.6	<i>nhr-46</i>	<i>mir-228</i>	F28B12.2	<i>egl-44</i>

<i>mir-235</i>	F28F9.1	<i>zag-1</i>	<i>mir-359</i>	F53B2.6	<i>ham-1</i>
<i>mir-266</i>	F28F9.1	<i>zag-1</i>	<i>mir-268</i>	F53B2.6	<i>ham-1</i>
<i>mir-61</i>	F28F9.1	<i>zag-1</i>	<i>lsy-6</i>	F58E10.2	<i>end-1</i>
<i>mir-250</i>	F28F9.1	<i>zag-1</i>	<i>mir-80</i>	F58E6.10	<i>unc-42</i>
<i>mir-240</i>	F28F9.1	<i>zag-1</i>	<i>mir-81</i>	F58E6.10	<i>unc-42</i>
<i>mir-241</i>	F28F9.1	<i>zag-1</i>	<i>mir-236</i>	F58G6.5	<i>nhr-34</i>
<i>mir-245</i>	F28F9.1	<i>zag-1</i>	<i>mir-237</i>	F58G6.5	<i>nhr-34</i>
<i>mir-247</i>	F28F9.1	<i>zag-1</i>	<i>lin-4</i>	F58G6.5	<i>nhr-34</i>
<i>mir-268</i>	F28F9.1	<i>zag-1</i>	<i>mir-252</i>	F58G6.5	<i>nhr-34</i>
<i>mir-269</i>	F28F9.1	<i>zag-1</i>	<i>mir-43</i>	F58G6.5	<i>nhr-34</i>
<i>mir-355</i>	F28F9.1	<i>zag-1</i>	<i>mir-35</i>	H21P03.1	<i>mbf-1</i>
<i>mir-44</i>	F28F9.1	<i>zag-1</i>	<i>mir-36</i>	H21P03.1	<i>mbf-1</i>
<i>mir-45</i>	F28F9.1	<i>zag-1</i>	<i>mir-38</i>	H21P03.1	<i>mbf-1</i>
<i>mir-48</i>	F28F9.1	<i>zag-1</i>	<i>mir-39</i>	H21P03.1	<i>mbf-1</i>
<i>mir-72</i>	F28F9.1	<i>zag-1</i>	<i>mir-40</i>	H21P03.1	<i>mbf-1</i>
<i>mir-80</i>	F28F9.1	<i>zag-1</i>	<i>mir-41</i>	H21P03.1	<i>mbf-1</i>
<i>mir-81</i>	F28F9.1	<i>zag-1</i>	<i>mir-230</i>	H21P03.1	<i>mbf-1</i>
<i>mir-83</i>	F28F9.1	<i>zag-1</i>	<i>mir-266</i>	K02B9.4	<i>elt-3</i>
<i>mir-43</i>	F29F11.5	<i>ceh-22</i>	<i>mir-250</i>	K02B9.4	<i>elt-3</i>
<i>mir-356</i>	F38A5.13	<i>dnj-11</i>	<i>mir-268</i>	K02B9.4	<i>elt-3</i>
<i>mir-90</i>	F40F8.7	<i>pqm-1</i>	<i>mir-2</i>	K02B9.4	<i>elt-3</i>
<i>mir-34</i>	F42G4.3	<i>zyx-1</i>	<i>mir-43</i>	K02B9.4	<i>elt-3</i>
<i>mir-355</i>	F42G4.3	<i>zyx-1</i>	<i>mir-46</i>	K02B9.4	<i>elt-3</i>
<i>mir-61</i>	F43G9.11	<i>ces-1</i>	<i>mir-47</i>	K02B9.4	<i>elt-3</i>
<i>mir-44</i>	F43G9.11	<i>ces-1</i>	<i>mir-228</i>	K02B9.4	<i>elt-3</i>
<i>mir-45</i>	F43G9.11	<i>ces-1</i>	<i>mir-72</i>	K02B9.4	<i>elt-3</i>
<i>mir-79</i>	F43G9.11	<i>ces-1</i>	<i>mir-230</i>	K02B9.4	<i>elt-3</i>
<i>let-7</i>	F45E6.2	<i>atf-6</i>	<i>mir-85</i>	K02B9.4	<i>elt-3</i>
<i>mir-392</i>	F45E6.2	<i>atf-6</i>	<i>mir-228</i>	R09G11.2	<i>nhr-1</i>
<i>let-7</i>	F46G10.6	<i>mxl-3</i>	<i>mir-79</i>	R09G11.2	<i>nhr-1</i>
<i>mir-80</i>	F46G10.6	<i>mxl-3</i>	<i>mir-80</i>	R09G11.2	<i>nhr-1</i>
<i>mir-81</i>	F46G10.6	<i>mxl-3</i>	<i>mir-81</i>	R09G11.2	<i>nhr-1</i>
<i>mir-63</i>	F49H12.1	<i>lsy-2</i>	<i>mir-235</i>	R13.4	
<i>mir-230</i>	F49H12.1	<i>lsy-2</i>	<i>mir-61</i>	R13.4	
<i>mir-85</i>	F52C12.5	<i>elt-6</i>	<i>mir-241</i>	R13.4	
<i>mir-246</i>	F52E1.1	<i>pos-1</i>	<i>mir-247</i>	R13.4	
<i>mir-61</i>	F52E1.1	<i>pos-1</i>	<i>mir-44</i>	R13.4	
<i>mir-247</i>	F52E1.1	<i>pos-1</i>	<i>mir-45</i>	R13.4	
<i>mir-42</i>	F52E1.1	<i>pos-1</i>	<i>mir-61</i>	T01B10.4	<i>nhr-14</i>
<i>mir-44</i>	F52E1.1	<i>pos-1</i>	<i>mir-247</i>	T01B10.4	<i>nhr-14</i>
<i>mir-45</i>	F52E1.1	<i>pos-1</i>	<i>mir-355</i>	T01B10.4	<i>nhr-14</i>
<i>mir-392</i>	T01B10.4	<i>nhr-14</i>	<i>mir-250</i>	W02C12.3	<i>hlh-30</i>
<i>mir-44</i>	T01B10.4	<i>nhr-14</i>	<i>mir-237</i>	W02C12.3	<i>hlh-30</i>
<i>mir-45</i>	T01B10.4	<i>nhr-14</i>	<i>lin-4</i>	W02C12.3	<i>hlh-30</i>
<i>mir-228</i>	T01B10.4	<i>nhr-14</i>	<i>mir-80</i>	W02C12.3	<i>hlh-30</i>



<i>mir-60</i>	T01B10.4	<i>nhr-14</i>	<i>let-7</i>	W09C2.1	<i>elt-1</i>
<i>mir-63</i>	T01B10.4	<i>nhr-14</i>	<i>mir-241</i>	W09C2.1	<i>elt-1</i>
<i>mir-246</i>	T05C12.6	<i>mig-5</i>	<i>mir-249</i>	W09C2.1	<i>elt-1</i>
<i>mir-75</i>	T07C4.6	<i>tbx-9</i>	<i>mir-271</i>	W09C2.1	<i>elt-1</i>
<i>mir-79</i>	T07C4.6	<i>tbx-9</i>	<i>mir-2</i>	W09C2.1	<i>elt-1</i>
<i>mir-44</i>	T09A5.12		<i>mir-43</i>	W09C2.1	<i>elt-1</i>
<i>mir-63</i>	T09A5.12		<i>mir-48</i>	W09C2.1	<i>elt-1</i>
<i>mir-75</i>	T09A5.12		<i>mir-84</i>	W09C2.1	<i>elt-1</i>
<i>mir-90</i>	T09A5.12		<i>mir-38</i>	Y104H12A.1	<i>nhr-41</i>
<i>mir-358</i>	T13C5.4		<i>let-7</i>	Y11D7A.12	<i>flh-1</i>
<i>mir-235</i>	T14F9.4	<i>peb-1</i>	<i>mir-241</i>	Y11D7A.12	<i>flh-1</i>
<i>mir-265</i>	T14F9.4	<i>peb-1</i>	<i>mir-48</i>	Y11D7A.12	<i>flh-1</i>
<i>mir-61</i>	T14F9.4	<i>peb-1</i>	<i>mir-230</i>	Y11D7A.12	<i>flh-1</i>
<i>mir-247</i>	T14F9.4	<i>peb-1</i>	<i>mir-84</i>	Y11D7A.12	<i>flh-1</i>
<i>mir-2</i>	T14F9.4	<i>peb-1</i>	<i>mir-355</i>	Y38C9A.1	
<i>mir-44</i>	T14F9.4	<i>peb-1</i>	<i>mir-239</i>	Y39B6A.46	
<i>mir-45</i>	T14F9.4	<i>peb-1</i>	<i>mir-48</i>	Y39B6A.46	
<i>mir-63</i>	T14F9.4	<i>peb-1</i>	<i>mir-84</i>	Y39B6A.46	
<i>mir-2</i>	T18D3.2	<i>odr-7</i>	<i>mir-241</i>	Y40B10A.8	<i>nhr-86</i>
<i>mir-80</i>	T18D3.2	<i>odr-7</i>	<i>mir-249</i>	Y40B10A.8	<i>nhr-86</i>
<i>mir-81</i>	T18D3.2	<i>odr-7</i>	<i>mir-90</i>	Y40B10A.8	<i>nhr-86</i>
<i>mir-231</i>	T23D8.8	<i>cfi-1</i>	<i>mir-249</i>	Y49E10.14	<i>pie-1</i>
<i>mir-232</i>	T23D8.8	<i>cfi-1</i>	<i>mir-84</i>	Y51H1A.4	<i>ing-3</i>
<i>mir-268</i>	T23H4.2	<i>nhr-69</i>	<i>mir-230</i>	Y53C12C.1	
<i>mir-72</i>	T23H4.2	<i>nhr-69</i>	<i>mir-80</i>	Y53C12C.1	
<i>mir-259</i>	T24C4.7		<i>mir-81</i>	Y53C12C.1	
<i>mir-51</i>	T24C4.7		<i>mir-246</i>	Y62E10A.14	
<i>mir-52</i>	T24C4.7		<i>mir-357</i>	Y66H1A.4	
<i>mir-53</i>	T24C4.7		<i>mir-232</i>	Y75B8A.1	<i>php-3</i>
<i>mir-54</i>	T24C4.7		<i>let-7</i>	ZK867.1	<i>ztf-10</i>
<i>mir-56</i>	T24C4.7		<i>mir-61</i>	ZK867.1	<i>ztf-10</i>
<i>let-7</i>	T27B1.2		<i>mir-241</i>	ZK867.1	<i>ztf-10</i>
<i>mir-236</i>	T27B1.2		<i>mir-247</i>	ZK867.1	<i>ztf-10</i>
<i>mir-72</i>	T27B1.2		<i>mir-392</i>	ZK867.1	<i>ztf-10</i>
<i>mir-80</i>	T27B1.2		<i>mir-44</i>	ZK867.1	<i>ztf-10</i>
<i>mir-81</i>	T27B1.2		<i>mir-45</i>	ZK867.1	<i>ztf-10</i>
<i>mir-2</i>	W01D2.2	<i>nhr-61</i>	<i>mir-48</i>	ZK867.1	<i>ztf-10</i>
<i>mir-43</i>	W01D2.2	<i>nhr-61</i>	<i>mir-84</i>	ZK867.1	<i>ztf-10</i>
<i>mir-75</i>	W01D2.2	<i>nhr-61</i>	<i>mir-85</i>	ZK867.1	<i>ztf-10</i>
<i>mir-79</i>	W01D2.2	<i>nhr-61</i>	<i>mir-232</i>	ZK867.1	<i>ztf-10</i>

Table II-7. List of composite miRNA $\rightleftharpoons$ TF feedback loops.

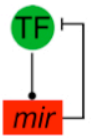
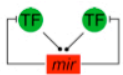
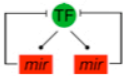
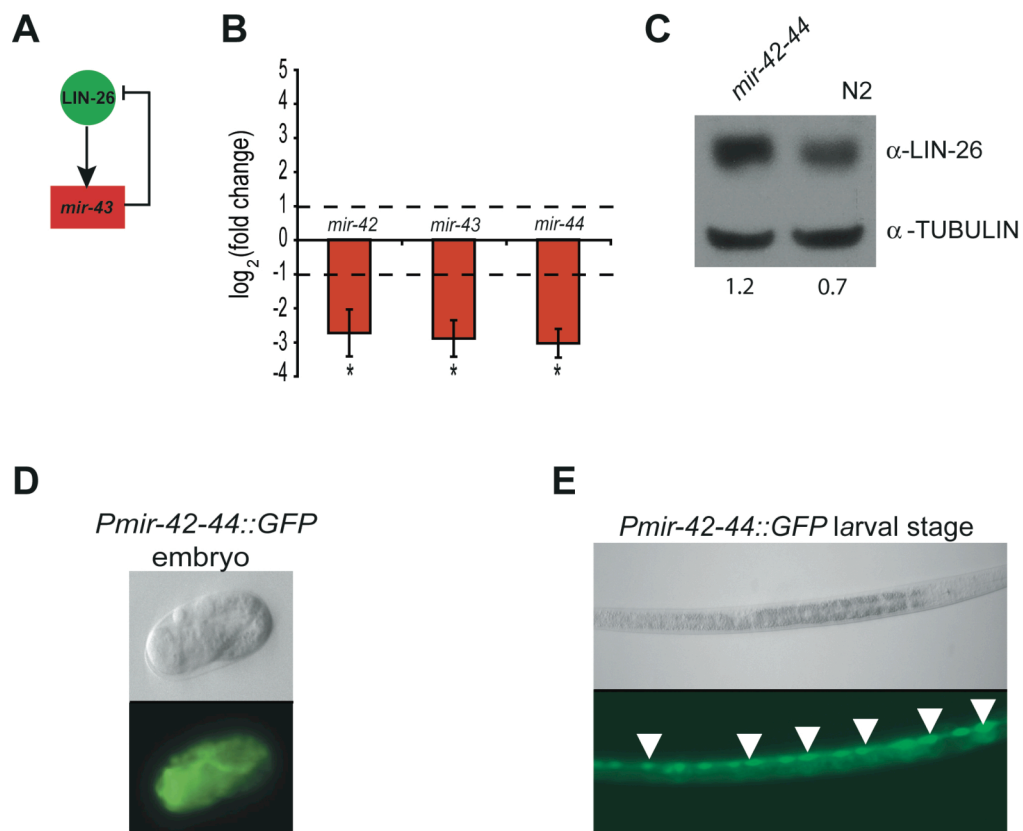
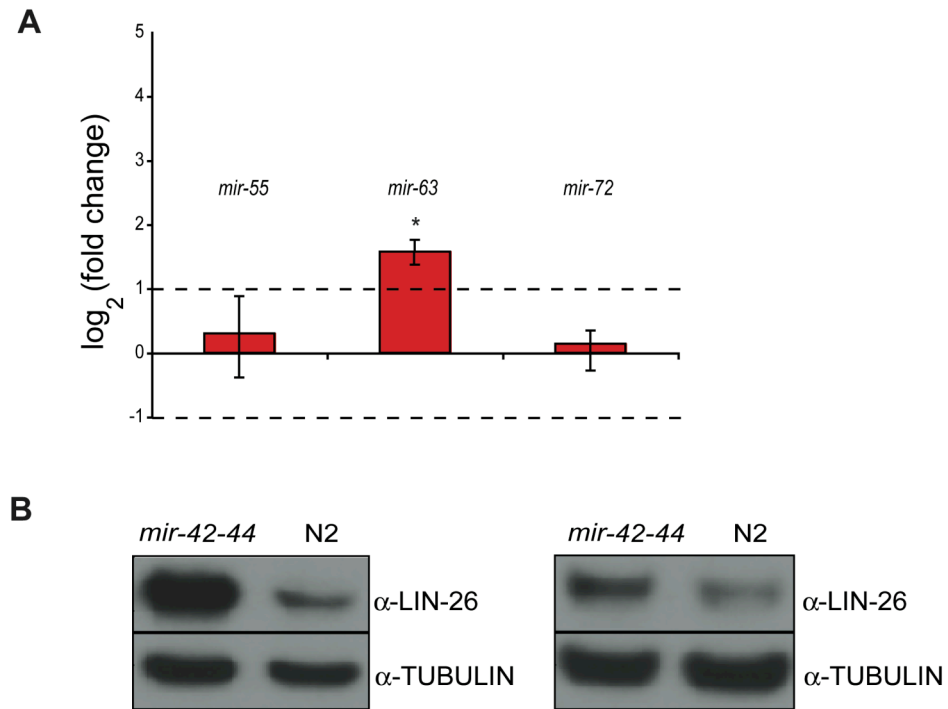
Motif	miRNA-1	miRNA-2	TF-1	TF-2	Promoter 1	Promoter 2	
	Type I	<i>let-7</i>	NA	DAF-3	NA	<i>Plet-7</i>	NA
	<i>let-7</i>	NA	T27B1.2	NA	<i>Plet-7</i>	NA	
	<i>mir-228</i>	NA	EGL-44	NA	<i>Pmir-228</i>	NA	
	<i>mir-236</i>	NA	LIN-26	NA	<i>Pmir-236</i>	NA	
	<i>mir-241</i>	NA	FLH-1	NA	<i>Pmir-241</i>	NA	
	<i>mir-241</i>	NA	DAF-3	NA	<i>Pmir-241</i>	NA	
	<i>mir-243</i>	NA	C44F1.2	NA	<i>Pmir-243</i>	NA	
	<i>mir-34</i>	NA	ZYX-1	NA	<i>Pmir-34</i>	NA	
	<i>mir-355</i>	NA	NHR-14	NA	<i>Pmir-355</i>	NA	
	<i>mir-355</i>	NA	Y38C9A.1	NA	<i>Pmir-355</i>	NA	
	<i>mir-43</i>	NA	LIN-26	NA	<i>Pmir-42-43-44</i>	NA	
	<i>mir-48</i>	NA	DAF-3	NA	<i>Pmir-48</i>	NA	
	<i>mir-48</i>	NA	ZAG-1	NA	<i>Pmir-48</i>	NA	
	<i>mir-48</i>	NA	FLH-1	NA	<i>Pmir-48</i>	NA	
	<i>mir-48</i>	NA	ZTF-2	NA	<i>Pmir-48</i>	NA	
	<i>mir-72</i>	NA	LIN-26	NA	<i>Pmir-72</i>	NA	
	<i>mir-76</i>	NA	NHR-43	NA	<i>Pmir-76</i>	NA	
	<i>mir-80</i>	NA	HLH-30	NA	<i>Pmir-227-80</i>	NA	
	<i>mir-80</i>	NA	ODR-7	NA	<i>Pmir-227-80</i>	NA	
	<i>mir-80</i>	NA	CEH-48	NA	<i>Pmir-227-80</i>	NA	
	<i>mir-81</i>	NA	DIE-1	NA	<i>Pmir-81</i>	NA	
	<i>mir-84</i>	NA	DAF-3	NA	<i>Pmir-84</i>	NA	
	<i>mir-84</i>	NA	NHR-43	NA	<i>Pmir-84</i>	NA	
	<i>mir-788</i>	NA	IRX-1	NA	<i>Pmir-788</i>	NA	
	Type II	<i>let-7</i>	NA	DAF-3	T27B1.2	<i>Plet-7</i>	NA
	<i>mir-241</i>	NA	DAF-3	FLH-1	<i>Pmir-241</i>	NA	
	<i>mir-355</i>	NA	NHR-14	Y38C9A.1	<i>Pmir-355</i>	NA	
	<i>mir-48</i>	NA	ZTF-2	FLH-1	<i>Pmir-48</i>	NA	
	<i>mir-48</i>	NA	DAF-3	FLH-1	<i>Pmir-48</i>	NA	
	<i>mir-48</i>	NA	ZAG-1	FLH-1	<i>Pmir-48</i>	NA	
	<i>mir-48</i>	NA	ZTF-2	ZAG-1	<i>Pmir-48</i>	NA	
	<i>mir-48</i>	NA	DAF-3	ZAG-1	<i>Pmir-48</i>	NA	
	<i>mir-48</i>	NA	ZTF-2	DAF-3	<i>Pmir-48</i>	NA	
	<i>mir-80</i>	NA	CEH-48	ODR-7	<i>Pmir-227-80</i>	NA	
	<i>mir-80</i>	NA	CEH-48	HLH-30	<i>Pmir-227-80</i>	NA	
	<i>mir-80</i>	NA	ODR-7	HLH-30	<i>Pmir-227-80</i>	NA	
	<i>mir-84</i>	NA	NHR-43	DAF-3	<i>Pmir-84</i>	NA	
		Type III	<i>let-7</i>	<i>mir-48</i>	DAF-3	NA	<i>Plet-7</i>
<i>mir-236</i>		<i>mir-43</i>	LIN-26	NA	<i>Pmir-236</i>	<i>Pmir-42-43-44</i>	
<i>mir-236</i>		<i>mir-72</i>	LIN-26	NA	<i>Pmir-236</i>	<i>Pmir-72</i>	
<i>mir-241</i>		<i>mir-48</i>	FLH-1	NA	<i>Pmir-241</i>	<i>Pmir-48</i>	
<i>mir-241</i>		<i>let-7</i>	DAF-3	NA	<i>Pmir-241</i>	<i>Plet-7</i>	
<i>mir-241</i>		<i>mir-48</i>	DAF-3	NA	<i>Pmir-241</i>	<i>Pmir-48</i>	
<i>mir-241</i>		<i>mir-84</i>	DAF-3	NA	<i>Pmir-241</i>	<i>Pmir-84</i>	
<i>mir-43</i>		<i>mir-72</i>	LIN-26	NA	<i>Pmir-42-43-44</i>	<i>Pmir-72</i>	
<i>mir-48</i>		<i>mir-84</i>	DAF-3	NA	<i>Pmir-48</i>	<i>Pmir-84</i>	
<i>mir-76</i>		<i>mir-84</i>	NHR-43	NA	<i>Pmir-76</i>	<i>Pmir-84</i>	
<i>mir-84</i>		<i>let-7</i>	DAF-3	NA	<i>Pmir-84</i>	<i>Plet-7</i>	

Figure II-3



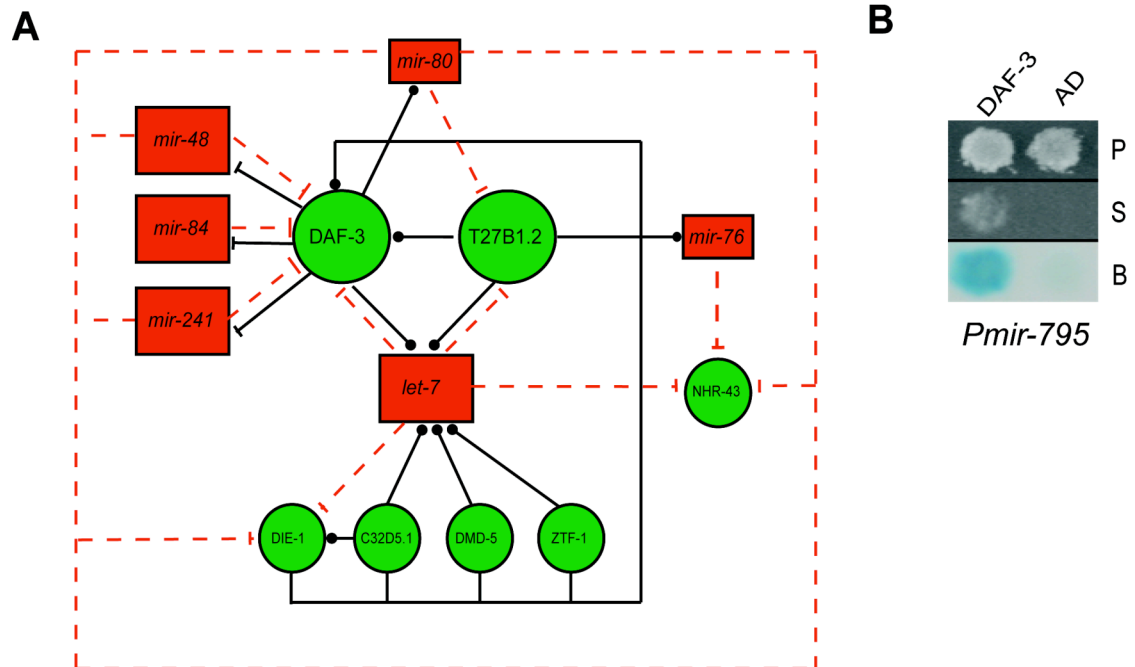
**Figure II-3.** *The mir-43<=>LIN-26 composite feedback loop.* (A) LIN-26 and *mir-43* function in a single negative composite feedback loop. (B) TaqMan PCR analysis shows that *mir-43* and the two miRNAs with which it is transcribed from an operon (*mir-42* and *mir-44*) are downregulated in *lin-26(ok939)* mutants compared to wild type animals. The average  $\log_2$ (fold change) of three experiments with standard error of the mean is shown. The dashed lines indicate a two-fold difference. Asterisks indicate significant changes. (C) Western blotting shows that LIN-26 is upregulated in *mir-42-44(nDf49)* mutant embryos compared to N2 wild type embryos.  $\alpha$ -tubulin antibody was used as a loading control. Numerical values represent LIN-26 levels after normalization to tubulin. (D) *Pmir-42-44* drives expression in the developing embryo. (E) *Pmir-42-44* drives expression in seam cells (a subset of seam cells is indicated by white arrows) in *C. elegans* larvae.

Figure II-4



**Figure II-4. Additional LIN-26 data.** (A) TaqMan PCR of remaining miRNAs that interact with LIN-26 in Y1H assays. The average log<sub>2</sub>(fold difference) of three experiments with standard error of the mean is shown. The dashed lines indicate a two-fold difference. Asterisks indicate significant changes. (B) Western blotting shows that LIN-26 is upregulated in *mir-42-44(nDf49)* mutant larvae compared to wild type worms. Left – L1 stage, right – L4 stage. α-tubulin antibody was used as a loading control.

Figure II-5

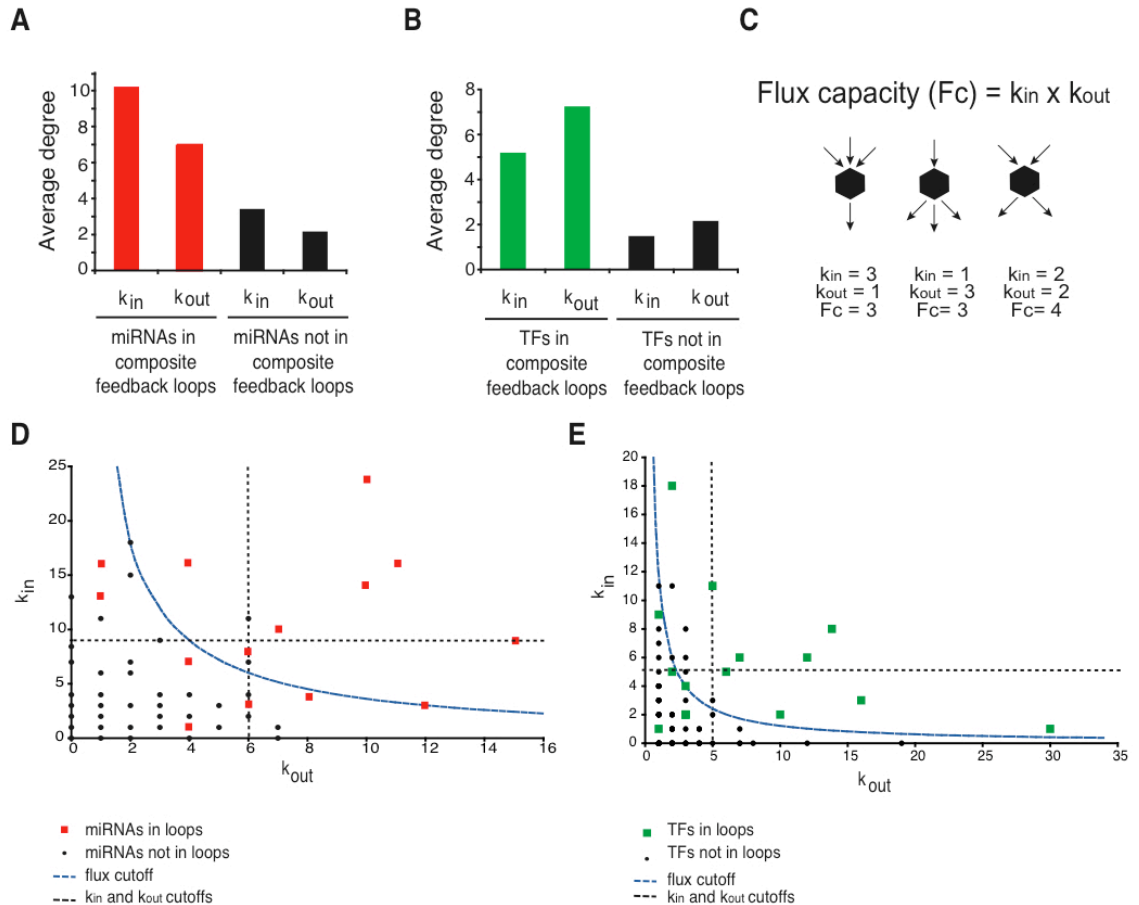


**Figure II-5.** *DAF-3* regulates the *let-7* family of miRNAs. (A) Example of a higher order network sub-graph containing multiple composite miRNA $\leftrightarrow$  TF feedback loops. Black arrows – transcriptional interactions; dashed red arrows – post-transcriptional interactions. Red rectangles – miRNAs, green circles – TFs. Repressive interactions are indicated by blunt arrows, interactions for which the functional consequence is unknown are indicated by dotted arrows. (B) Y1H assay demonstrating the interaction between *Pmir-795* and DAF-3. P – permissive media; S – selective media; B –  $\beta$ -Galactosidase assay; AD – empty vector.

**Table II-8.** *Network motif analysis.*

Composite miRNA $\leftrightarrow$ TF feedback motif	Randomization method	Number of loops in real network	Average number of loops in randomized network	P-value	Number of randomizations
Type I	Edge switching	23	13.71	0.004	25,000
	Node replacement I	23	10.61	0.0039	25,000
	Node replacement II	23	10.62	0.0002	25,000

Figure II-6



**Figure II-6.** *TFs and miRNAs in composite feedback loops are characterized by a high flux capacity ( $F_c$ ).* (A) Average in- and out-degree of miRNAs that participate in loops (red) or that do not (black). (B) Average in- and out-degree of TFs that participate in loops (green) or that do not (black). (C) The  $F_c$  of a node is defined by the product of the in- and out-degree. As the example indicates, nodes with the same total number of edges can have a different flux. (D, E) Plot of in-degree ( $k_{in}$ ) versus out-degree ( $k_{out}$ ) for each miRNA (D) and TF (E) in the integrated network. Red squares – miRNAs involved in composite feedback loops; green squares – TFs involved in feedback loops; black circles – miRNAs (D) and TFs (E) not involved in composite feedback loops. Dashed lines represent cut-offs for  $k_{in}$ ,  $k_{out}$  and  $F_c$  for the 15% most highly connected nodes.



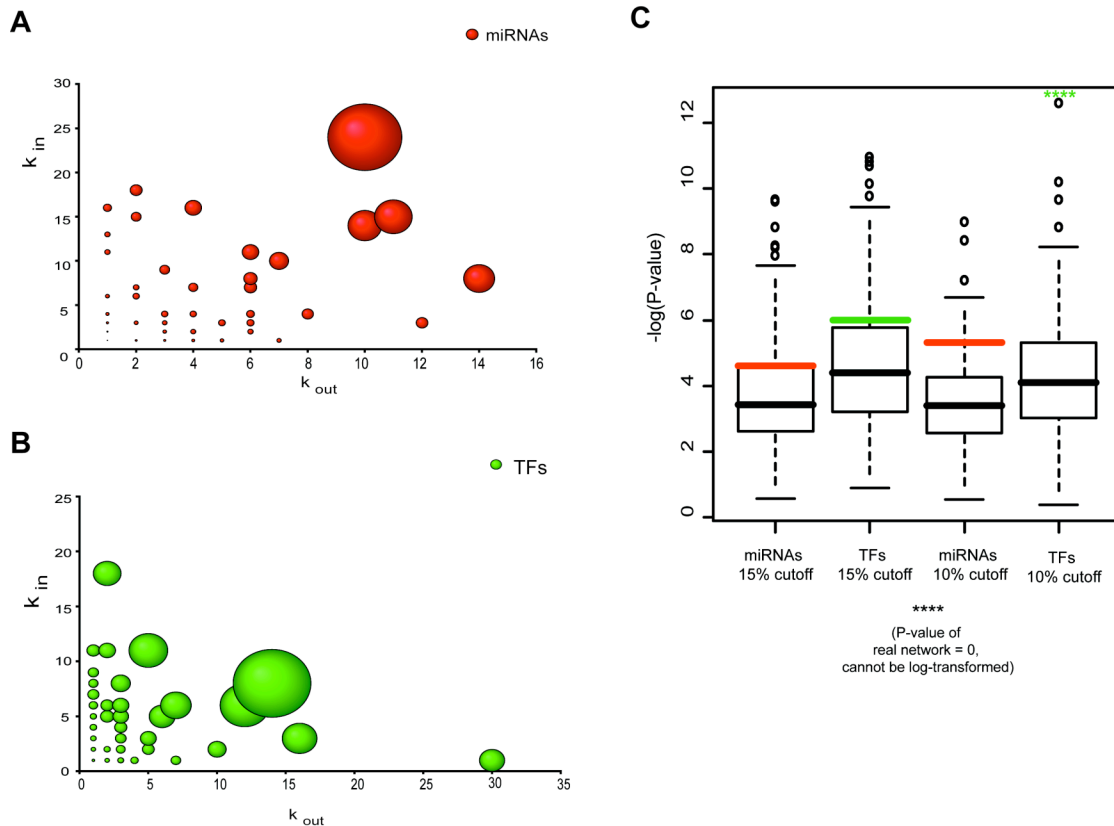
**Table II-9. Separation between nodes involved and not involved in composite feedback loops according to different cut-offs of  $k_{in}$ ,  $k_{out}$  and  $F_c$ .**

**Table S8. Separation between nodes involved and not involved in composite feedback loops according to different cutoffs of  $k_{in}$ ,  $k_{out}$  and flux.**

<b>Nodes with degree 0 included</b>						
top 15% miRNAs	$k_{in} \geq 9$	$k_{in} < 9$	$k_{out} \geq 6$	$k_{out} < 6$	$flux \geq 36$	$flux < 36$
In loops	7	7	9	5	8	6
Not in loops	6	64	9	61	4	66
P-value	0.000780*		0.000147*		0.000025*	
<b>Nodes with degree 0 removed</b>						
top 10% miRNAs	$k_{in} \geq 13$	$k_{in} < 13$	$k_{out} \geq 8$	$k_{out} < 8$	$flux \geq 48$	$flux < 48$
In loops	6	8	6	8	7	7
Not in loops	3	67	2	68	1	69
P-value	0.000469*		0.000171*		0.000005*	
top 15% of TFs	$k_{in} \geq 5$	$k_{in} < 5$	$k_{out} \geq 5$	$k_{out} < 5$	$flux \geq 12$	$flux < 12$
In loops	8	8	8	8	10	6
Not in loops	10	90	9	91	7	93
P-value	0.000473*		0.000141*		0.000001*	
top 10% of TFs	$k_{in} \geq 6$	$k_{in} < 6$	$k_{out} \geq 7$	$k_{out} < 7$	$flux \geq 18$	$flux < 18$
In loops	6	10	6	10	9	7
Not in loops	8	92	6	94	3	97
P-value	0.004173*		0.001520*		0.000000*	
<b>Nodes with degree 0 removed</b>						
top 15% miRNAs	$k_{in} \geq 11$	$k_{in} < 11$	$k_{out} \geq 8$	$k_{out} < 8$	$flux \geq 42$	$flux < 42$
In loops	6	8	6	8	7	7
Not in loops	4	41	2	43	2	44
P-value	0.0079*		0.0014*		0.0002*	
top 10% miRNAs	$k_{in} \geq 15$	$k_{in} < 15$	$k_{out} \geq 9$	$k_{out} < 9$	$flux \geq 66$	$flux < 66$
In loops	4	10	6	8	5	9
Not in loops	2	43	1	44	1	43
P-value	0.024*		0.0004*		0.002*	
top 15% of TFs	$k_{in} \geq 6$	$k_{in} < 6$	$k_{out} \geq 5$	$k_{out} < 5$	$flux \geq 20$	$flux < 20$
In loops	6	11	8	9	9	8
Not in loops	8	49	3	54	2	55
P-value	0.0058*		0.000179*		0.000009*	
top 10% of TFs	$k_{in} \geq 8$	$k_{in} < 8$	$k_{out} \geq 7$	$k_{out} < 7$	$flux \geq 30$	$flux < 30$
In loops	4	13	6	11	8	9
Not in loops	4	53	1	56	0	57
P-value	0.075*		0.00040*		0.000001*	

\*Fisher exact test. P-value for same or stronger association

Figure II-7

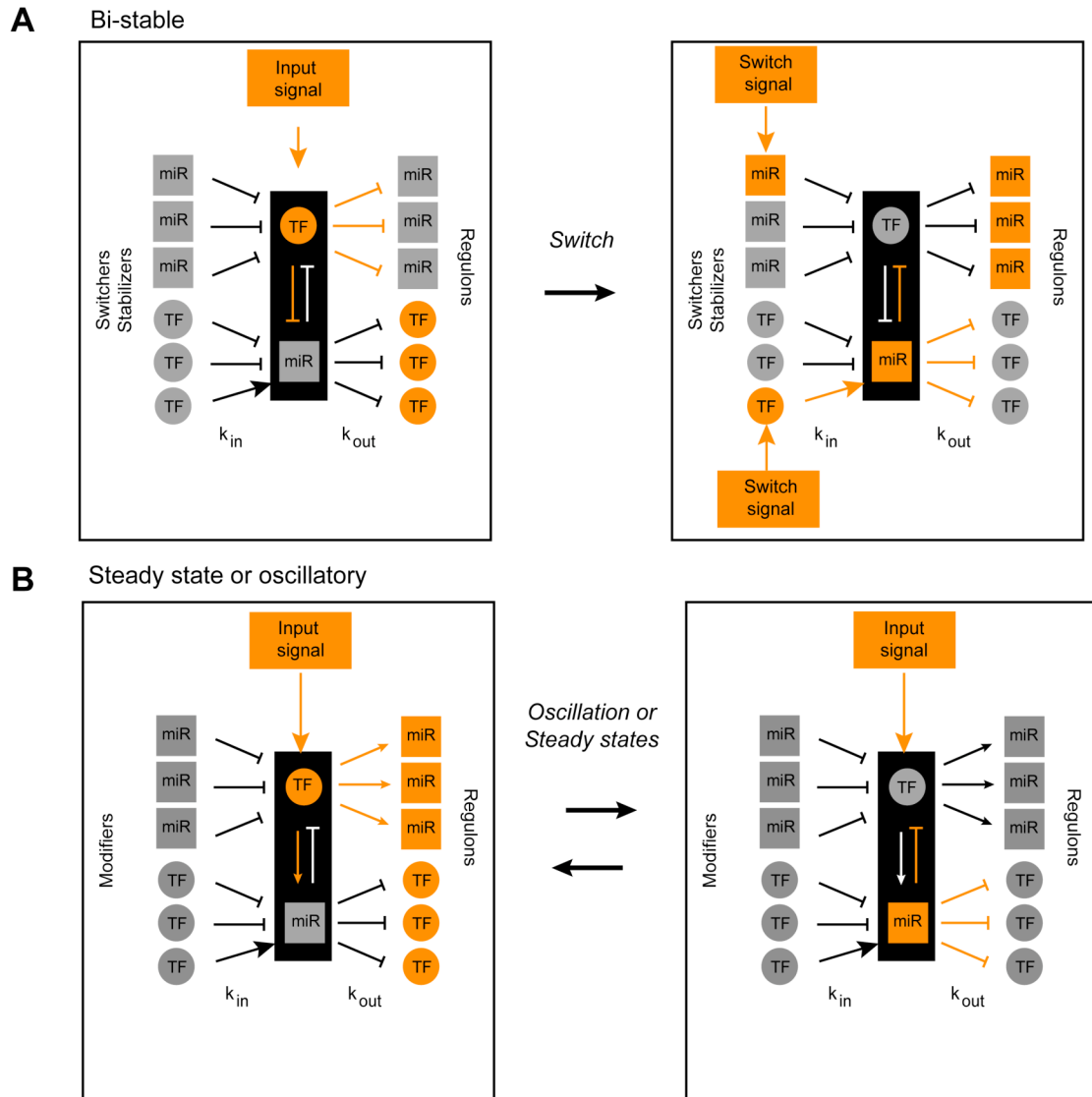


**Figure II-7.** *A high flux capacity correlates with composite feedback loops.* (A, B) Plot of in-degree ( $k_{in}$ ) versus out-degree ( $k_{out}$ ) for each miRNA (A) and TF (B) in 400 randomized networks. The size of the node reflects the ratio of the number of times the node was involved in loops versus the number of times it was not involved in loops. (C) Boxplot of P-values for association between participation in loops and high Fc (top 10% and 15% cut-offs) in real versus 400 randomized networks (generated by Edge switching). Red lines – values for miRNAs in real network, green line – value for TFs in real network. All statistics were done by Fisher Exact test for each individual randomized network.

**Table II-10.** *Introduction of 10% false negative TF=>miRNA interactions does not affect the enrichment of miRNA<=>TF feedback loops.*

	Number of loops	Average number of loops in randomized networks	P-value	Randomizations
Real network	23	13.71	0.003	25,000
1	20	12.85	0.02	25,000
2	21	12.53	0.006	25,000
3	19	11.63	0.013	25,000
4	21	13	0.011	25,000
5	21	12.35	0.007	25,000
6	19	13.14	0.044	25,000
7	22	12.46	0.001	25,000
8	20	12.5	0.013	25,000
9	21	12.75	0.008	25,000
10	20	12.44	0.012	25,000

Figure II-8



**Figure II-8.** Model for the function of composite feedback loops in gene expression programs. (A) Bi-stable systems are generated by double negative feedback loops. (B) Steady state or oscillatory systems can be generated by single negative feedback loops. For each type of loop an example is shown in which orange indicates nodes or edges that are “on”, and grey indicates nodes or edges that are “off”.

### PREFACE TO CHAPTER III

The work presented in the following chapter describes the generation of a resource to study miRNA expression at the genome-scale level. This work is also part of a collaboration with the Ambros lab and embodies the joint effort of several people: M. C. Ow contributed to the generation of transgenic strains and part of the data on Figure III-2; J. S. Reece-Hoyes contributed to the annotation of GFP expression patterns; M. I. Barrasa contributed to some of the analysis in Figure III-5 and myself (Figures III-1 through III-6). A.J. Walhout and I wrote the manuscript.

This chapter has been published separately in:

Martinez, N. J.\*, Ow, M. C.\*, Reece-Hoyes, J. S., Barrasa, M. I., Ambros, V. R., and Walhout, A. J. 2008. Genome-scale spatiotemporal analysis of *Caenorhabditis elegans* microRNA promoter activity. *Genome Res* **18**: 2005-15.

\*These authors contributed equally to this work

**CHAPTER III****Genome-scale spatiotemporal analysis of *Caenorhabditis elegans*****microRNA promoter activity**

**Abstract**

The *Caenorhabditis elegans* genome encodes more than 100 microRNAs (miRNAs). Genetic analyses of miRNA deletion mutants have only provided limited insights into miRNA function. To gain insight into the function of miRNAs, it is important to determine their spatiotemporal expression pattern. Here, we use miRNA promoters driving the expression of GFP as a proxy for miRNA expression. We describe a set of 73 transgenic *C. elegans* strains, each expressing GFP under the control of a miRNA promoter. Together, these promoters control the expression of 89 miRNAs (66% of all predicted miRNAs). We find that miRNA promoters drive GFP expression in a variety of tissues and that, overall, their activity is similar to that of protein-coding gene promoters. However, miRNAs are expressed later in development, which is consistent with functions after initial body plan-specification. We find that miRNA members belonging to families are more likely to be expressed in overlapping tissues than miRNAs that do not belong to the same family, and provide evidence that intragenic miRNAs may be controlled by their own, rather than a host gene promoter. Finally, our data suggest that post-transcriptional mechanisms contribute to differential miRNA expression. The data and strains described here will provide a valuable guide and resource for the functional analysis of *C. elegans* miRNAs.

## Introduction

Differential gene expression can be regulated at many levels and by various *trans*-acting factors. MicroRNAs (miRNAs) and transcription factors (TFs) are pivotal regulators of metazoans gene expression. While TFs physically interact with *cis*-regulatory DNA elements to activate or repress gene expression, miRNAs mainly repress gene expression post-transcriptionally by imperfect base pairing to sequences located in the 3'UTR of their target mRNAs [reviewed in: (Bartel 2004; Walhout 2006)]. Like TFs, many miRNAs are highly conserved between related species and even across phyla. Typically, miRNAs are transcribed by RNA polymerase II into a primary transcript (pri-miRNA) that is further processed by Drosha into a ~60 nt long precursor (pre-miRNA), and subsequently by Dicer into a mature ~23 nt long miRNA [reviewed in: Bartel 2004]. The two founding miRNAs, *lin-4* and *let-7*, were identified genetically as temporal regulators of development in the nematode *Caenorhabditis elegans* (Lee et al. 1993; Reinhart et al. 2000). MiRNAs regulate a broad range of biological processes in animals and plants, including patterning of the nervous system, cell death, cell proliferation and development (Ambros 2004; Stefani and Slack 2008). In addition, as for TFs, there is increasing evidence that mammalian miRNA expression may also be regulated at the post-transcriptional level (Obernosterer et al. 2006; Thomson et al. 2006; Viswanathan et al. 2008; Wulczyn et al. 2007).



Genome-wide genetic analyses in many organisms have demonstrated a myriad of critical roles that TFs play in controlling gene expression during development, homeostasis and disease. For instance, more than 30% of *C. elegans* TFs confer a detectable phenotype when knocked down by RNAi (291 out of 940 predicted TFs tested) [data obtained from WormBase WS180, (Vermeirssen et al. 2007b)]. In contrast, with the exceptions of *lin-4* (Lee et al. 1993), *let-7* (Reinhart et al. 2000), *lisy-6* (Johnston and Hobert 2003) and *mir-1* (Simon et al. 2008), a single null mutation does not result in an easily detectable phenotype for most *C. elegans* miRNA genes (Miska et al. 2007). The observation that most *C. elegans* miRNAs appear to be individually dispensable may reflect roles in processes that have not yet been readily assayable. Alternatively, there may be considerable genetic redundancy between miRNAs and other regulators such as TFs, other miRNAs, and in some cases members of the same miRNA family. For example, the three related miRNAs *lin-58* (hereafter referred to as *mir-48*), *mir-84* and *mir-241* function redundantly in the control of developmental timing in *C. elegans* (Abbott et al. 2005). One approach that will help to delineate the biological function of miRNAs is by determining when and where they are expressed. The particular pattern of expression of each miRNA gene should help to identify potential genetic interactors that exhibit similar expression patterns, and to design experiments for the delineation of phenotypes of miRNA mutants.

A simple anatomy, invariant cell lineage, transparent body, and high-quality complete genome sequence, make *C. elegans* a highly suitable model to study spatiotemporal miRNA expression. In addition, many biological processes are conserved between nematodes and higher organisms, so the analysis of miRNA function in *C. elegans* may potentially be applicable to other animals. For instance, it has been demonstrated that *mir-1*, a highly conserved miRNA, is expressed and functions in muscle in diverse organisms such as mice, zebrafish, fruit flies and nematodes (Simon et al. 2008; Sokol and Ambros 2005; Wienholds et al. 2005; Zhao et al. 2007). Therefore, the spatiotemporal expression pattern, and perhaps, function of many other miRNAs may also be also conserved.

Previous studies in various organisms have examined miRNA expression by *in situ* hybridization (Aboobaker et al. 2005; Wienholds et al. 2005), Northern blotting (Lau et al. 2001; Lee and Ambros 2001), or small RNA library sequencing from enriched tissues (Landgraf et al. 2007; Ruby et al. 2007). Although powerful, such studies can be limited by a relatively low sensitivity. In addition, these methods do not enable the analysis of spatiotemporal expression patterns in living animals as they depend on animal fixation (*in situ* hybridization) or RNA purification (Northern blotting, sequencing). Reporter genes such as that encoding the green fluorescent protein (GFP) have provided powerful tools for the analysis of gene expression *in vivo*. Indeed, promoter::*gfp* fusions in *C. elegans* have already been used to analyze more than 350 TFs and ~1800 other protein-coding genes (referred to here as “all genes”) (Hunt-Newbury et al. 2007;

Reece-Hoyes et al. 2007). Importantly, this approach faithfully recapitulates known gene expression in the majority of cases examined (Reece-Hoyes et al. 2007).

Here, we present a collection of 73 transgenic *C. elegans* strains, each containing a miRNA promoter::*gfp* fusion construct. We used promoter activity as a proxy for miRNA expression *in vivo*. We examined miRNA promoter activity across all developmental stages, and, frequently, to the level of individual cells. We compared miRNA promoter activity to that of the TF and “all gene” datasets introduced above. We find that miRNA promoters are active in all major tissues and cell types. However, miRNA promoters are active later in development than protein-coding gene promoters, which is consistent with roles for miRNAs after the initial specification of body plan, organs and tissues (Schier and Giraldez 2006; Wienholds et al. 2005). We correlate promoter activity with previously reported Northern blotting data and examine two endogenous pri-miRNAs by RT-PCR. Our data suggest that post-transcriptional regulation of pri-miRNAs provides an additional layer of differential miRNA expression in nematodes. The data and transgenic lines that we present provide a platform for functional miRNA studies to delineate their roles in the development of the animal, and to understand their function in gene regulatory networks.

## Results

### *Generation of transgenic PmiRNA::gfp C. elegans strains*

Of the 134 *C. elegans* miRNA genes currently available in miRBase V9.0, 75 reside in intergenic regions, *i.e.* between protein-coding genes, and can be assigned to their own promoter (Fig. III-1A). An additional 22 intergenic miRNAs are transcribed in a total of nine intergenic operons, with a single promoter regulating each operon. Sixteen miRNAs are embedded within the intron of protein-coding genes in the anti-sense orientation either as single genes (7 miRNAs), or as operons (9 miRNAs into 2 operons)(Fig. III-1A, Table III-1). Twenty-one miRNAs are embedded within the intron of a protein-coding gene in the sense orientation. It has been hypothesized that such miRNAs are under the control of the host gene promoter (Baskerville and Bartel 2005) and, therefore, we largely focus on the set of 113 miRNAs with presumed independent promoters.

We generated miRNA promoter::*gfp* (*PmiRNA::gfp*) fusions by Gateway cloning (Walhout et al. 2000). We used the *PmiRNA* Entry clones we generated previously (Martinez et al. 2008), and a Gateway-compatible GFP destination vector (Dupuy et al. 2004). Promoter sequences were defined as the intergenic genomic sequence upstream of annotated miRNA genes with a minimum length of 300 bp and a maximum length of 2 kb. It has been shown previously that upstream sequences defined by these criteria are often sufficient for rescue of miRNA mutant phenotypes (Chang et al. 2004; Johnson et al. 2003; Lee et al.

1993) and/or to recapitulate miRNA expression (Johnson et al. 2005; Li et al. 2005; Yoo and Greenwald 2005). *PmiRNA::gfp* constructs were then used to transform *unc-119(ed3)* worms by microparticle bombardment as described (Berezikov et al. 2004; Reece-Hoyes et al. 2007).

In total, we generated a collection of transgenic lines for 70 *PmiRNA::gfp* constructs (we will introduce another three below). These 70 constructs together include upstream sequences for 61 single gene miRNAs and nine miRNA operons, corresponding to a total of 86 miRNAs (out of 113 considered, or 76%). On average, we obtained four independent lines per construct. We observed a high transmission rate of the *PmiRNA::gfp* transgene for most of the lines (data not shown). With only one exception, all independent lines for a given construct show similar expression patterns. The exception is the promoter of *mir-227-80*. While one line shows mosaic expression in excretory cells, vulva, body wall muscle and head neurons, two other independent lines show expression in the pharynx and head neurons. All strains were genotyped to verify the presence of *Pmir-227-80* and both expression patterns are provided in our EDGEDb database (Barrasa et al. 2007).

#### *Characterization of miRNA expression patterns*

We examined the activity of miRNA promoters throughout the whole organism and across all developmental stages in living animals, and, when feasible, to the level of individual cells. Specifically, for each transgenic line we

examined GFP expression in a mixed stage population of hermaphrodites. We only recorded the expression pattern of a given *PmiRNA::gfp* reporter strain that was observed consistently in each of the independent *PmiRNA::gfp* transgenic lines (data not shown). Detailed descriptions and representative images can be found in Table III-2 and in our publicly available EDGEdb database.

In total, 90% of the miRNA promoters confer GFP expression (63 out of 70) (Table III-3). The expression rate of *PmiRNA::gfp* fusions is comparable to that of TFs (91%) (Reece-Hoyes et al. 2007) and “all genes” (79%) (Hunt-Newbury et al. 2007) (Fig. III-1B). This demonstrates that the chosen genomic sequences upstream of miRNAs indeed function as promoters. The promoters of seven miRNAs did not drive detectable GFP expression *in vivo*. Two of these miRNAs are conserved in the related nematode *C. briggsae*: *Isy-6*, a well-characterized miRNA involved in neuronal specification (Johnston and Hobert 2003) and *mir-77*, for which a phenotype has not been described, but which has been detected in large-scale sequencing analyses (Ruby et al. 2006). The fact that we did not observe GFP expression for these promoters may be because they lack elements required for expression, or because the transgene is present at a low copy number, which may not suffice for the detection of GFP expression. The other five miRNAs for which we did not detect promoter activity, *mir-257*, *mir-258*, *mir-261*, *mir-267* and *mir-271*, are not conserved in *C. briggsae* and have not been detected by large-scale sequencing (Ruby et al. 2006). We found a significant correlation between GFP expression and conservation (Fisher exact

test,  $P$ -value $<0.05$ ), and between GFP expression and detection by sequencing (Fisher exact test,  $P$ -value $< 0.05$ ). Based on these observations, it is possible that some or all of *mir-257*, *mir-258*, *mir-261*, *mir-267* and *mir-271* are not genuine miRNAs and/or are not transcribed under normal culture conditions.

#### *Temporal PmiRNA::gfp activity correlates with Northern blot analysis*

Northern blots have been extensively used to determine the temporal expression of miRNAs in *C. elegans* (Ambros et al. 2003; Lau et al. 2001; Lee and Ambros 2001; Lim et al. 2003). We searched WormBase (WS180) for information regarding temporal miRNA expression. Of the 81 miRNAs for which we had temporal information based on *PmiRNA::gfp* transgenic lines, equivalent information was available for 58 (Table III-4). We found that the observed temporal GFP expression pattern agrees perfectly with the pattern detected by developmental Northern blots in most of the cases. Four *PmiRNA::gfp* strains did not match the temporal expression pattern. For instance, we only detected *mir-82* promoter activity in the L4 and adult stage while Northern blotting detected mature miRNA in all developmental stages (Fig. III-2A, see also below). These discrepancies may be due to a lack of regulatory elements in the chosen genomic DNA fragment. In other cases, the temporal pattern partially agrees with previously reported patterns (Table III-4). Twelve *PmiRNA::gfp* strains exhibited earlier expression than reported previously by Northern blotting. There are several explanations for this difference. For instance, the DNA fragments used as

promoters may lack transcriptional elements that are required for repression of miRNA expression in early developmental stages. Also, GFP transgenics may be more sensitive for detecting spatially restricted miRNA expression in early stages of development. For instance, mature *mir-237* was detected from L3 to adult stages, however, we observed GFP expression in *Pmir-237::gfp* animals as early as the first larval stage (see also: Esquela-Kerscher et al. 2005). We performed additional Northern blotting using StarFire probes to detect the temporal expression of nine miRNAs: *mir-241*, *mir-84*, *mir-48*, *let-7*, *mir-83*, *mir-230*, *mir-240*, *mir-82* and *mir-85*. This allowed the more sensitive detection of low levels of mature miRNAs than traditional oligonucleotide probes (Behlke et al. 2000; Abbott et al. 2005; Ow et al. 2008). For four of these miRNAs (*mir-241*, *mir-84*, *mir-48* and *mir-83*), we detected a weak miRNA signal at earlier stages than previously reported, consistent with our *PmiRNA::gfp* expression data (Table III-4 and Fig. III-2A). We also detected the temporal expression of two additional miRNAs (*mir-59* and *mir-90*) for which there was no information in WormBase and found that it was consistent with promoter activity (Table III-4 and Fig. III-2A).

We have also observed cases in which mature miRNAs were only detected early by Northern blotting, while we detected continuous *PmiRNA::gfp* activity in later stages. This was the case for two miRNA operons: *mir-42-44* and *mir-35-41* (a total of 9 miRNAs). In addition to the aforementioned reasons, the differences observed between mature miRNA expression and miRNA promoter activity may be due to post-transcriptional mechanisms that may regulate



transcript stability or processing of either the pri-miRNA, the pre-miRNA or the mature miRNA (see below).

Taken together, the temporal expression in *Pmir::gfp* animals was consistent with the expression determined by Northern blotting for 65% (39/60) of the miRNAs in our dataset. For only 7% (4/60) of the miRNAs, the expression determined by Northern blotting does not agree with promoter activity, while the remaining 28% (17/60) partially agrees (Fig. III-2B).

#### *Post-transcriptional mechanisms contribute to differential miRNA expression*

Our data suggest that post-transcriptional mechanisms affect transcript processing or stability of several miRNAs. For instance, while *mir-61* is detected by Northern blotting in all developmental stages, *mir-250* is only detected starting from the L1 stage (Lee and Ambros 2001; Lim et al. 2003), even though both miRNAs are likely transcribed from the same promoter (*Pmir-61-250*). This suggests that post-transcriptional mechanisms may either prevent processing of the *pre-mir-250* transcript or affect mature *mir-250* stability. Similarly, three other miRNAs that are expressed from a single operon and are thus controlled by one promoter, *mir-42*, *mir-43* and *mir-44*, are differentially expressed: mature *mir-42* and *mir-43* are only detected in embryos while *mir-44* is detected not only in embryos but also in larval and adult stages (Lau et al. 2001). Consistent with the expression of mature *mir-44*, we observed promoter activity from *Pmir-42-44* in all developmental stages, suggesting that *mir-42* and *mir-43* might be subject to

post-transcriptional regulation. Lau *et al.* only detected mature miRNAs from the *mir-35-41* operon in the embryo by Northern blotting. However, they detected the precursor of *mir-35* (*pre-mir-35*) both in the embryo and at the L4 stage, suggesting that it is downregulated between those stages (Lau *et al.* 2001). We detected *Pmir-35-41* activity (by GFP fluorescence) not only in embryos and L4 stages but also in the other larval stages and in adults (Supplemental Table S3 and EDGEdb). To test whether this may be the result of GFP stability rather than promoter activity, we used RT-PCR to detect the endogenous *mir-35-41* primary transcript (*pri-mir-35-41*). We observed *pri-mir-35-41* in embryos and L4, where mature and pre-miRNAs are detected, as well as in L1, L2 and L3 stages, where neither mature nor pre-miRNAs from this cluster were detected. This suggests that post-transcriptional mechanisms regulate the processing or stability of the *mir-35-41* primary transcript, pre-miRNAs, or mature miRNAs during L1 to L4 stages (Fig. III-2C). We also compared the expression of the *let-7* primary transcript to the expression of mature *let-7* as described previously (Bracht *et al.* 2004). We detected mature *let-7* by Northern blotting starting at the L3 stage, which is in agreement with previous observations (Reinhart *et al.* 2000)(Fig. III-2A). However, we detected *pri-let-7* by RT-PCR as early as the embryonic stage, consistent with the GFP expression observed in *Plet-7::gfp* strains (Fig. III-2C and EDGEdb). Similar observations have been made for *let-7* in mammalian systems, where *let-7* processing is selectively blocked in embryonic stem cells (Viswanathan *et al.* 2008; Wulczyn *et al.* 2007). Our results suggest that post-

transcriptional mechanisms likely regulate *pri-let-7* processing at early stages (Fig. III-2C).

Taken together, our results show that miRNA promoter activity largely overlaps with mature miRNA expression and that post-transcriptional mechanisms likely contribute to differences in primary and mature miRNA expression.

#### *The promoters of miRNAs are active later in development*

We detected GFP expression conferred by miRNA promoters in all developmental stages. Representative examples of embryonic promoter activity are shown in Figure III-3A. We compared the temporal expression conferred by promoter of miRNAs to that of TFs and “all genes” and noticed that miRNA promoters overall tend to be less active in embryos (P-value <0.05)(Fig. III-3B). In addition, the majority of miRNA promoters that confer embryonic expression tend to do so at later embryonic stages, on average, than do promoters of TFs (Fig. III-3C; this analysis could not be done for the “all genes” dataset as analogous temporal expression information was not available). This observation is in agreement with previous studies in other organisms that suggest that miRNAs are involved in tissue differentiation and maintenance rather than the establishment of body plan, organs and tissues (Schier and Giraldez 2006; Wienholds et al. 2005).

*Most miRNA promoters drive GFP expression in a tissue-specific manner*

We found that miRNA promoters drive expression in all major tissues and cell types, except the germline (Table III-3)(Fig. III-4A). Representative examples of miRNA promoter activity in various parts of the somatic gonad and neuronal cells are shown in Figures III-4B and III-4C. Microparticle bombardment has been reported to be the method of choice when germline expression is desired (Praitis et al. 2001). However, none of the miRNA promoters are able to direct GFP expression in the germinal gonad. Thus, it is possible that the miRNAs assayed here are exclusively expressed in somatic tissues. However, promoters of protein-coding genes also generally fail to drive GFP reporter expression in the germline (Hunt-Newbury et al. 2007; Reece-Hoyes et al. 2007). Thus, we think that it is more likely that the GFP transgene is silenced in the germinal gonad. Future studies that use germline-specific deep sequencing of miRNAs will reveal whether any of the miRNAs are expressed in this tissue.

To enable the comparison at a tissue level between miRNAs and protein-coding genes, we re-annotated the TF and “all genes” datasets according to a systematic spatiotemporal expression scheme that we devised. We defined 23 categories (hereafter referred to as “tissues”), including intestine, vulva, head neurons, *etc.* (see *Methods* and Table III-5 for precise definitions). Some of these are highly specific (*e.g.* distal tip cells), and others are broader (*e.g.* head neurons). We observed that most miRNA promoters confer GFP expression in only a few tissues or cell types. For instance, almost 50% of the promoters confer

expression in three or fewer tissues while only less than 5% of promoters confer ubiquitous somatic expression (*lin-4*, *let-7* and *mir-53*). A high degree of tissue specificity has also been observed for miRNAs in other organisms, including chicken and zebrafish (Xu et al. 2006; Wienholds et al. 2005). The promoters of TFs and “all genes” drive GFP expression with a similar degree of tissue specificity (Fig. III-4D). We recently obtained a genome-scale miRNA transcriptional network (Martinez et al. 2008) that reveals a similar overall network architecture as protein-coding gene networks (Deplancke et al. 2006; Martinez et al. 2008; Vermeirssen et al. 2007a). Together, these observations indicate that the regulation of miRNA gene promoters is not fundamentally different from that of protein-coding gene promoters.

#### *Members of miRNA families can be expressed in distinct or overlapping patterns*

MiRNAs can be classified into families according to sequence similarities (Bartel 2004). Sixty percent of *C. elegans* miRNAs (78 out of 134) can be classified into 24 families, each containing between two and eight members (Ruby et al. 2006). Members of a given family are predicted to share target mRNAs and may function redundantly (Abbott et al. 2005; Miska et al. 2007). For instance, the *let-7* family members *mir-48*, *mir-84* and *mir-241* function together to regulate the L2 to L3 cell fate transitions in the hypodermis (Abbott et al. 2005). Redundancy among miRNAs from the same family can occur if miRNA family members are (partially) co-expressed. We examined the extent to which

spatiotemporal promoter activity of members of a miRNA family overlap. We were able to compare the expression patterns for ten complete miRNA families, as well as for two families for which we have expression patterns for most, but not all of the members (Fig. III-5A). Interestingly, we observed that some families are expressed with a high degree of overlap, whereas other families exhibit largely non-overlapping spatiotemporal expression. For instance, miRNAs from the *mir-35* family (*mir-35-36-37-38-39-40-41* cluster and *mir-42*) are expressed throughout all stages and in overlapping tissues, including the vulva, seam cells, head neurons and the rectum (Fig. III-5B). Similarly, members of the *lin-4* (*lin-4*, *mir-237*) and *mir-46* (*mir-46*, *mir-47*) families are expressed in overlapping tissues (Table III-3). In contrast, miRNAs from the *mir-75* family (*mir-75* and *mir-79*) are expressed in different tissues. While *Pmir-75* confers GFP expression exclusively in the intestine, *Pmir-79* drives expression in the hypodermis (Fig. III-5C). Similarly, members of the *mir-232* (*mir-232*, *mir-357*) and *mir-251* (*mir-251*, *mir-252*) families exhibit distinct expression patterns (Table III-3).

We introduce a “tissue overlap coefficient” (TsOC) as the number of tissues shared between two miRNAs divided by the smallest of the total number of tissues where either miRNA is expressed (Fig. III-5D). This coefficient is similar to the topological overlap coefficient (TOC) that is used for network modularity analysis (Vermeirssen et al. 2007a). We used TsOC as a measure to determine if the overlap in expression between miRNAs from the same family is different than the overlap in expression between miRNAs from distinct families.

We calculated TsOCs for all pairs of miRNAs from distinct families, as well as pairs of miRNAs from the same family (see *Methods*). We found that the distribution of TsOCs for pairs of miRNAs from the same family is significantly different than the distribution of TsOCs for pairs of miRNAs from distinct families (Fig. III-5E; Fisher exact test P-value = 0). Pairs of miRNAs from the same family tend to have a higher TsOC compared to pairs of miRNAs from distinct families. Taken together, the degree of overlapping expression varies per miRNA family, however, miRNAs from the same family do tend to exhibit overlapping expression patterns. Thus, it is likely that the lack of phenotypes for individual miRNAs can be explained (at least partly) by familial redundancy and that, in addition, many miRNAs may have a synthetic genetic interaction with other miRNAs, or perhaps with protein-coding genes.

### *Intragenic miRNAs*

MiRNA genes that are located within the intron of a protein-coding gene in the sense orientation are thought to be under the control of the host gene promoter (Baskerville and Bartel 2005). We generated *PmiRNA::gfp* constructs using the immediate upstream sequence of three of these intragenic miRNAs: *mir-58*, *mir-2* and *mir-82*, which are embedded in the intron of Y67D8A.1, *ppfr-1* and T07D1.2, respectively (Fig. III-6). We found that the region upstream of *mir-58* does not confer GFP expression (data not shown). Surprisingly, however, sequences upstream of both *mir-82* and *mir-2* drive tissue-specific GFP

expression (Fig. III-6). In addition, the annotation of *lin-4* has recently changed; rather than being located in an intergenic region (WS140), it is now annotated to be located in an intron of F59G1.4 (WS180). We and others have shown that the genomic fragment immediately upstream of *lin-4* does function as a promoter (Esquela-Kerscher et al. 2005; Ow et al. 2008; this study). It has been previously shown that internal promoters in operons are a common feature in the *C. elegans* genome (Huang et al. 2007). It is tempting to speculate that internal miRNA promoters located in the introns of protein-coding genes might be common as well. In contrast to *C. elegans* miRNAs, a large proportion of human miRNAs are located within introns (Rodriguez et al. 2004). It will be interesting to see if the genomic sequences upstream of these miRNAs can function as promoters in mammals or whether this is specific to nematodes.



## Discussion

We present here the generation and analysis of transgenic animals for 73 *PmiRNA::gfp* constructs that represent the expression of 89 *C. elegans* miRNAs. Several lines of evidence indicate that the majority of these transgenic animals likely recapitulate endogenous miRNA transcription. First, it has been demonstrated previously that a 2 kb fragment upstream of the translational start site of protein-coding genes accurately drives gene expression in the majority of cases examined (Reece-Hoyes et al. 2007). Secondly, the majority of *PmiRNA::gfp* lines completely or partially recapitulate previously reported temporal expression of miRNAs detected by Northern blotting (Ambros et al. 2003; Lau et al. 2001; Lee and Ambros 2001; Lim et al. 2003). Thirdly, for a handful of miRNAs it has been shown that such a fragment is sufficient for miRNA rescue and in other expression experiments (Chang et al. 2004; Johnston and Hobert 2003; Johnston et al. 2005; Lee et al. 1993; Li et al. 2005; Yoo and Greenwald 2005).

We compared miRNA promoter activity to mature miRNA expression determined by Northern blotting. While in the majority of cases promoter activity exactly agrees with mature miRNA expression, there are cases in which they only partially agree. We have shown in the case of *let-7* and the *mir-35-41* operon that this partial agreement is likely due to post-transcriptional mechanisms that contribute to differential miRNA expression. Such mechanisms can in principle control pri-miRNA, pre-miRNA or mature miRNA stability and/or

processing. It has been shown previously that mammalian miRNAs can be regulated post-transcriptionally (Obernosterer et al. 2006). Viswanathan and colleagues have identified LIN-28 as a developmentally regulated RNA binding protein that selectively blocks the processing of *pri-let-7* in embryos (Viswanathan et al. 2008). In the future, it will be important to dissect the factors that play a role in post-transcriptional regulation of *C. elegans* miRNAs.

We found that miRNAs are expressed in a variety of tissues. In zebrafish and fruit flies, previous studies have also shown a broad expression for many miRNAs (Aboobaker et al. 2005; Wienholds et al. 2005). We also found that miRNAs are expressed relatively late in development, which is in agreement with results obtained in zebrafish and likely reflects a function of miRNAs in tissue differentiation and maintenance, rather than in tissue establishment (Wienholds et al. 2005).

Most miRNAs do not confer a detectable phenotype when deleted (Miska et al. 2007). It is likely that the lack of phenotypes for individual miRNAs can be explained not only by familial redundancy but also by genetic interactions with miRNAs from other families, or perhaps by interactions with protein-coding genes, such as TFs. The spatiotemporal miRNA expression patterns will provide an important tool for the identification of genes with which they may act redundantly, and hence will be an important tool that can be used toward understanding the cellular functions of each miRNA.

Our study provides some important advantages over other studies of miRNA expression. First, our method is non-invasive, which means that expression can be studied in living animals. Secondly, in contrast to methods such as Northern blotting or sequencing, we can frequently annotate miRNA expression to the single cell level. Thirdly, we will provide all the strains to the *C. elegans* community, which should help to delineate the expression patterns at greater levels of resolution. Fourthly, the transgenic lines will enable the study of miRNA expression under different (experimental) conditions, including dauer, stress, *etc*, and in males. And finally, the transgenic lines will be available for other studies. For example, they can be used to identify or validate upstream regulators of miRNA expression. We recently mapped a genome-scale miRNA regulatory network by high-throughput yeast one-hybrid assays (Deplancke et al. 2004; Deplancke et al. 2006) and used several of the transgenic lines described here for *in vivo* validation of the interactions obtained (Martinez et al. 2008, Ow et al. 2008).

*Note added in proof*

During the review of this paper, we have generated an additional *C. elegans* transgenic strain containing *Pmir-71::gfp*. This strain was not included in the analysis, but information regarding GFP expression is available in Tables III-1 to III-6 and EDGEdb. *mir-71* is an intragenic miRNA, annotated in the intron of

*ppfr-1*, the same intron where *mir-2* is annotated. Sequence upstream of *mir-71* drives GFP expression in vivo.

## Methods

### *Generation of Pmir::gfp constructs*

For our network study (Martinez et al. 2008) we used the 115 miRNA predictions available in WormBase WS130 (<http://www.wormbase.org>) and miRNA registry V4.0 (<http://microna.sanger.ac.uk>) (Griffiths-Jones et al. 2006; Ambros et al. 2003; Lim et al. 2003). We completed these with 19 recently discovered miRNAs (WormBase WS175 and miRBase V9.0). A miRNA promoter is defined as the intergenic region upstream of the predicted stem-loop sequence or from the mature miRNA as annotated in miRBase V4.0 (Table III-1). We used a minimal length of 300 bp and a maximal length of 2 kb. In total, 93 promoters (that control 113 miRNAs) were selected. Seventy three promoters (controlling 89 miRNAs) were successfully cloned into pDEST-DD04 by Gateway cloning as described (Dupuy et al. 2004; Walhout et al. 2000). Constructs were verified by DNA sequencing using either GFP Fw (5'-TTCTACTTCTTTTACTGAACG) or GFP Rv (5'-CTCCACTGACAGAAAATTTG) primers.

The following *PmiRNA::gfp* constructs were generated by conventional restriction enzyme-based cloning into the pPD97.75 vector (see Table III-1 for information on restriction sites used): *Pmir-257*, *Pmir-51*, *Pmir-2*, *Pmir-228*, *Pmir-54*, *Pmir-81*, *Pmir-235*, *Pmir-227-80* and *Pmir-234*, *Plet-7*, *Plin-4*, *Pmir-48*, *Pmir-237*, *Pmir-241*, *Pmir-84*.

### *C. elegans strains*

Routine *C. elegans* maintenance and culture were done as described (Brenner 1974). The DP38 strain (*unc-119(ed3)*) was cultured in liquid media for microparticle bombardment as described (Reece-Hoyes et al. 2007) or in egg-plates (Wood 1988).

#### *Transformation of C. elegans by microparticle bombardment*

Transgenic *PmiRNA::gfp* animals were generated as described previously (Berezikov et al. 2004; Reece-Hoyes et al. 2007).

#### *Genotyping*

The genotype of each transgenic line was confirmed by single animal PCR (Williams et al. 1992) using GFP Fw and GFP Rv primers (see above) as described, followed by DNA sequencing to confirm the identity of the miRNA promoter in the *PmiRNA::gfp* transgene.

#### *Characterization of GFP expression patterns*

Mixed populations of hermaphrodites were examined by fluorescence microscopy using a Zeiss Axioskop 2 plus microscope equipped with a FITC filter. We recorded the expression pattern conferred by each miRNA promoter that was consistent in each of the independently derived transgenic lines (except for *Pmir-227-80*, see main text). Fluorescence photographs representative of each expression pattern were taken using a Hamamatsu Orca-ER /1394 video

camera and Axiovision Rel. 4.5 software and stored in the EDGEdb database (Barrasa et al. 2007). For each genotype, we stored up to three independent lines into frozen stocks. These lines were chosen based on highest transmission level and/or GFP expression (data not shown). These lines will be made available through the CGC.

#### *PmiRNA::gfp expression pattern annotation*

We devised a standardized temporal and spatial annotation to record the expression pattern of each *PmiRNA::gfp*. Temporal expression patterns were classified into eight stages: early, mid and late embryo, all four larval stages and adult stage. We defined early embryo as the pre-comma stage, mid-embryo as comma stage and late embryo as two and three-fold embryos. Spatial expression patterns were classified into 23 categories that correspond to tissues, cell types, organs, and, when feasible, to individual cells (*i.e.* coelomocytes and distal tip cells) (Table III-5). For GFP expression analysis purposes, temporal and spatial expression was standardized into a binary code, where 1 represents expression detected and 0 represents no expression detected (Table III-3).

#### *Other datasets*

GFP expression patterns driven by “TFs” and other protein-coding genes (“all genes”) were obtained from Reece-Hoyes et al (Reece-Hoyes et al. 2007) and Hunt-Newbury *et al.* (Hunt-Newbury et al. 2007), and converted into our

binary annotation scheme. Specifically, “all genes” patterns were classified as follow: BM (body wall muscle); BN (body neurons, lateral nerve cords/commissures, ventral nerve cord); C (coelomocytes); DTC (distal tip cell); GS (gonad sheath cells); HH (hypodermis); HN (amphids, dorsal nerve cord, head neurons, labial sensilla, nerve ring, pharyngeal neurons); I (intestinal, intestinal muscle); O (other: amphid socket cells, developing gonad, head mesodermal cell, mechanosensory neurons, pvt interneuron, unidentified body, unidentified cells, unidentified tail, unidentified head, uterine-seam cell, other); P (arcade cells, pharynx); PG (pharyngeal gland cells); PIV (pharyngeal-intestinal valve); R (anal depressor muscle, anal sphincter, rectal epithelium, rectal gland cells); S (developing spermatheca, spermatheca); SC (seam cells); TN (phasmids, tail neurons); U (developing uterus, uterine muscle, uterus); USV (spermatheca-uterine valve); V (developing vulva, vulva other, vulval muscle); X (excretory cells, excretory gland cells). For comparison analyses, several tissues/systems were fused in one or more of the datasets to allow the same category types in all three datasets: HH and BH categories were fused into one category, H (hypodermis); HM and BM categories were fused into one M (muscle); PG and P were fused into P (pharynx) and I and PI were fused into I (intestinal).

### *Northern blot analyses*



Total RNA was extracted using TRIzol reagent (Invitrogen) and analyzed by Northern blotting using 5 µg of RNA from each stage as described before (Ow et al. 2008).

#### *RT-PCR analyses*

Total RNA was extracted as above and digested with RNase-Free DNase Set (Qiagen) following the manufacturer's recommendations. First strand cDNA synthesis was performed using 2.5 µg of total RNA, random primers and SuperScript II (Invitrogen) following manufacturer's recommendations.

Primer sequences used in the PCR reactions:

*mir-38-RT-2*: 5'-GGGCTCTCGGTATATCAGG-3'

*mir-35-PCR-4*: 5'-GGAAATGGTCCATTCAGTCATC-3'

*fat-4 L*: 5'-TGTTTCTATCTTGTTGGAGG

*fat-4 R*: 5'-GGTAAACCATTTGCTGCTGC

Primers used to detect the *let-7* primary transcript are A62, A127 and A63 (Bracht et al. 2004).

#### *TsOC analysis*

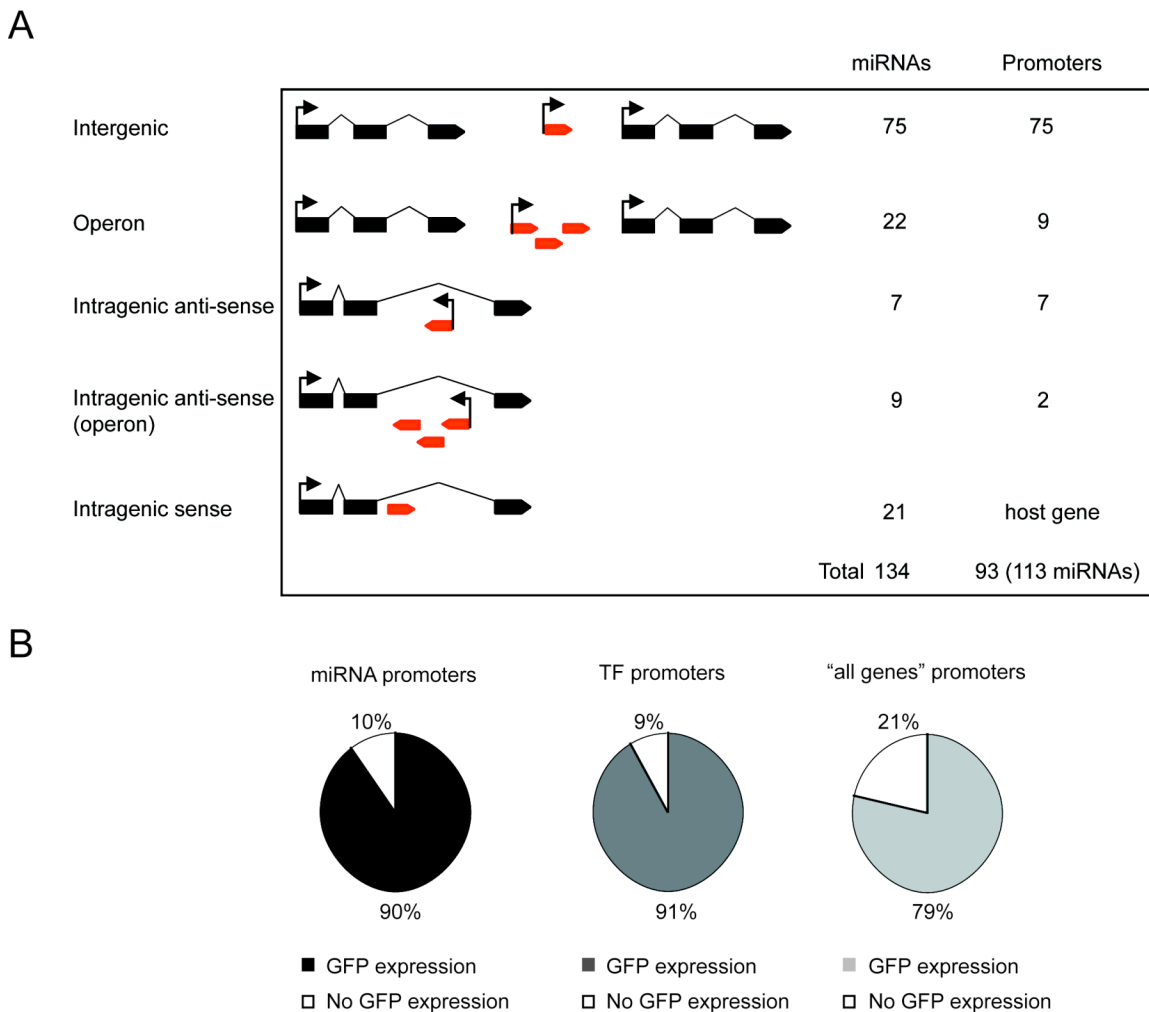
The tissue overlap coefficient (TsOC) between any two miRNAs was defined as the number of tissues where both miRNAs are expressed divided by the smallest of the total number of tissues in which either miRNA is expressed (see Fig. III-5D). In case of operons, where several miRNAs are expressed from

a single promoter, the same expression pattern was assigned to all miRNAs in the operon. We calculated a TsOC for all individual pairs of miRNAs from different families (3160 total pairs) and all pairs of miRNAs from the same families (80 total pairs). We grouped the TsOCs into four bins ( $0 > \text{TsOC} \geq 0.25$ ;  $0.25 > \text{TsOC} \geq 0.5$ ;  $0.5 > \text{TsOC} \geq 0.75$  and  $0.75 > \text{TsOC} \geq 1$ ) and calculated if the distribution of TsOCs within families was significantly different from the distribution between families using a Fisher exact test for 4-rows x 2-columns tables (Freeman and Halton 1951).

## Acknowledgments

We thank members of the Walhout and Ambros laboratories and Job Dekker for advice and critical reading of the manuscript. We thank the *Caenorhabditis* Genetics Center (CGC), which is funded by the NIH National Center for Research Resources (NCRR), for providing the *unc-119 (ed3)* strain. We thank A. Fire for providing pPD95.75. M.C.O. was supported in part by NIH postdoctoral fellowship GM070118-02. This work was supported by NIH grants DK068429 to A.J.M.W., and GM348642 to V.A.

**Figure III-1. *MiRNA promoters and expression rates.***



**Figure III-1. *MiRNA promoters and expression rates.*** (A) Number of miRNA genes and promoters considered according to genome annotation. Black boxes – protein-coding gene exons; red boxes – miRNA genes. (B) Expression rate of *Pmir::gfp* constructs compared to TFs and “all genes”.

Table III-1. Promoter details and primer sequences.

Promoter (miRNA registry V4.0 and *V9.0)	Location WS130 and *WS175	Ch	<sup>a</sup> GW-Forward Primer	<sup>b</sup> GW-Reverse Primer	Forward Primer	Reverse Primer	Comment	Transgenic lines
<i>Plet-7</i>	Intergenic	X	NA	NA	ttcattgtcgact atctttaatgtct cac	cctactgtcgac ccggatccaca gtgtagaccgctc	<sup>c</sup> <i>Sall</i> sites used for cloning are in italics	yes
<i>Plin-4</i>	Intergenic	II	NA	NA	accgtaccgg gctattagtattg caccagac	ggctccccgg gacaggccgg aagcataaact c	<sup>c</sup> <i>Xmal</i> sites used for cloning are in italics	yes
<i>Plsy-6</i>	Intergenic	V	TTTAAACTT ACAGTTTCG G	GCTTATTTTT CAGAAATTA G	NA	NA		yes
<i>Pmir-1</i>	Intergenic	I	ATGACATCT GCAAGAAAC	ATGGGCATG TAAGGAAGT ATGC	NA	NA		yes
<i>Pmir-228</i>	Intergenic	IV	GAGATTCATA TTCGATTTTT CCC	TACTTCATTA ACCTCATCA CC	ggtgaccccc gggaaaaatgt ttattgtgacgtc atactcttg	catgcagtccc ggggcgaacg ggataaggag gaaaatgtctc	<sup>c</sup> <i>Xmal</i> sites used for cloning are in italics	yes
<i>Pmir-230</i>	Intergenic	X	ACACATTCA CAGTTTTAC TGG	ACATCAGAC AAAAAATA CAAGC	NA	NA		yes
<i>Pmir-231</i>	Intergenic	III	GTTTACCATT TTAGAATTC CGC	TGTGGTGCT AACTTTGAA AATA	NA	NA		yes
<i>Pmir-232</i>	Intergenic	IV	CATACAATTT TACATCATAA CG	CAAACCTGC CTGGAATCA A	NA	NA		yes
<i>Pmir-234</i>	Intergenic	II	TGCATCCAA TACAAATATC AG	TGTGTTTCA ACTTACATAG	atgtcttagtga ctcactgaagg tggcctatagc agttgtg	ttctcgaggtcg actagattgcg gactgacgcatt tttattatc	<sup>c</sup> <i>Sall</i> sites used for cloning are in italics	yes
<i>Pmir-235</i>	Intergenic	I	TTTTTTGCA GCTTCGTTT CAGCG	TAATTTTCAC GGTGAACAA TTTTGC	tcgcttggtcga caacaacctgt agctggatcttt ggtcac	gggagagtgc gactttaatttc acggggaaca attttgcaat	<sup>c</sup> <i>Sall</i> sites used for cloning are in italics	yes
<i>Pmir-236</i>	Intergenic	II	CTTTAAATTA AAACGGTGA G	ACCGAAGAA TACGACAGT AGC	NA	NA		yes
<i>Pmir-237</i>	Intergenic	X	NA	NA	cagtgatggcc agacttcactca cgtaggttc	aattctggcca ccacgcaatgt agaagtttg	<sup>c</sup> <i>MscI</i> sites used for cloning are in italics	yes
<i>Pmir-238</i>	Intergenic	III	AAAAGAACG TTGAGTTAC	GGAGAATTG CCATTTTCTT G	NA	NA		yes

<i>Pmir-239a</i>	Intergenic	X	GAGCTTTTG AGGAAATTT	TGTAGAGTT ATTGACTG AG	NA	NA		no
<i>Pmir-239b</i>	Intergenic	X	AAACGTCGC ATTTTGTAT	CTGTAACCTT TTCAGAATAA TAAGG	NA	NA		no
<i>Pmir-240- *786</i>	Intergenic/o peron	X	TTATCTATCC GACCATAACC	CGAAATAGC AATAGACAG GA	NA	NA		yes
<i>Pmir-241</i>	Intergenic	V	NA	NA	<i>ctcccagtcgac</i> <i>cactcatttctact</i> <i>ttctctc</i>	<i>gcacctcccg</i> <i>gactttgacacc</i> <i>cccgcggttg</i>	<sup>c</sup> <i>Sall</i> and <i>Xma</i> I sites used for cloning are in italics	yes
<i>Pmir-242</i>	Intergenic	IV	TCAACGCGT ACACTTTG	CGAGCCAC GAGGGAAC ATTC	NA	NA		yes
<i>Pmir-243</i>	Intergenic	IV	ATCGTGTGA CAATGACGT TG	GTGAACAGA ATTTCTGTG AGTAAC	NA	NA	Annotated in WS175 as intronic (intron of R08C7.12).	yes
<i>Pmir-244</i>	Intergenic	I	CCCCCCTTT GACAATCGA ATA	ATGGAGGGA AAAAATTAAA ATTAAAAGT G	NA	NA		yes
<i>Pmir-245</i>	Intergenic	I	ATTCCAACC GAATTTAAAT C	ATAGCTCTC CAACATTGA ATTC	NA	NA		yes
<i>Pmir-246</i>	Intergenic	IV	GAGTGTCTGA GCGGGAAA C	TGCTGAATAT TGTGTAGC	NA	NA		yes
<i>Pmir-247- *797</i>	Intergenic/o peron	X	AATTTGGAA GAAGAAAA TC	ATTTGTAGA CAACTAGTG TGC	NA	NA		yes
<i>Pmir-249</i>	Intergenic	X	AGTTTAACTT CAGCTCTTG	GAGTATAGAT ACACAATAT GC	NA	NA		no
<i>Pmir-251</i>	Intergenic	X	TCCTATTCAA TTGATTGTT CG	GCAAGGGAT CATTTTCAC	NA	NA		yes
<i>Pmir-252</i>	Intergenic	II	AGATGATATT CAAGACATT ATCG	TGGAATAGT AAGATCAGA ATGAC	NA	NA		yes
<i>Pmir-255</i>	Intergenic	V	AACGAGTTG TGTGAGTCT TG	CACTGAAGA ATTAACAG	NA	NA		yes
<i>Pmir-256</i>	Intergenic	V	TCTCGTGGA GAAAATGGC	TTCTCAATAT TCATCTAGA CAGC	NA	NA		no

<i>Pmir-257</i>	Intergenic	V	AATCTCCCT GACGACTCC	TACTTCGC AAACAGCTT C	tactggaacc gggtgacgact cccctcgcgac accagagtg.	tactcctgccc ggagcgggaa atattacttc gcaaacag	Annotated in WS175 in a trasposon; ° <i>Xmal</i> sites used for cloning are in italics	yes
<i>Pmir-258</i>	Intergenic	X	CCTAAATTC CCAAAAAT GG	CCTGCGGAT CGACTATTG	NA	NA		yes
<i>Pmir-259</i>	Intergenic	V	ACTTTAAAA GTCTTCTGG AAAAAG	AACCGAACT TACCACTAT G	NA	NA		yes
<i>Pmir-261</i>	Intergenic	II	TGGTAAACA TCAAAGCAA AAC	ATCCATATTC ACCGTGAAA	NA	NA		yes
<i>Pmir-262</i>	Intergenic	V	CGGTGGAAT ATTGTTAAAA TTG	CTAATTGTAG AGAAACCAT AC	NA	NA		no
<i>Pmir-264</i>	Intergenic	X	GATTTTTGT ACGGGAATG	AGTGCTGCT TCTCGAATG	NA	NA		no
<i>Pmir-265</i>	Intergenic	IV	AGTAGAAAA TAGAAAGTC AGATCC	GTCAGAATA ACCACACAA AATTAG	NA	NA		yes
<i>Pmir-266</i>	Intergenic	X	GACACGGG AGATGTAAC C	CAGCTTCTG ATCCAAGTT A	NA	NA		yes
<i>Pmir-268</i>	Intergenic	V	ATATCCATA TTCACATGC	CTCTCCAAA ATGTTTCG	NA	NA		yes
<i>Pmir-269</i>	Intergenic	IV	AGAACATGT CTGTTTTGA GATAG	CAAGTTTAG CCAACTTCT G	NA	NA		yes
<i>Pmir-270</i>	Intergenic	IV	TCCGCTGCC AAGAAGCTT C	CTTTGTGCC TACATAAAAC C	NA	NA		yes
<i>Pmir-271</i>	Intergenic	X	ATGGTCTTAT CAGCAGAAG C	TGACCTCCT CTTTCTCAC	NA	NA		yes
<i>Pmir-34</i>	Intergenic	X	TAAAGCCCA CTGGTGACT	GCATTGTCC GTTATAAGAA T	NA	NA		no
<i>Pmir-355</i>	Intergenic	II	CATTCTGGA AAAAGTAGT CGG	TATTAATCTC AAACAGCGA C	NA	NA		yes
<i>Pmir-359- *785</i>	Intergenic/o peron	X	TGGTCTTAG TGGTCTTAG TTC	GAGCATTAAT ATCAGTACT GTAG	NA	NA		yes
<i>Pmir-360</i>	Intergenic	X	GGTTATGGG GTTTTATGA	GCCGAATAA TCAATAAGA A	NA	NA		yes

<i>Pmir-392</i>	Intergenic	X	CTTCGTGTA ACTTACGTT TG	TATCCAGAAT AAATTTTGAA G	NA	NA	yes	
<i>Pmir-45</i>	Intergenic	II	GCCGATTGT GATCAAAAA AATC	GGTGCATGA GCAGCTGAA TAC	NA	NA	yes	
<i>Pmir-46</i>	Intergenic	III	GCCACCCAA TTCACGAGC C	TTTGGACCC TTACCTTCA ACTTTTCG	NA	NA	yes	
<i>Pmir-47</i>	Intergenic	X	ATCGTGGAA ACAATAGGT TCT	GATCAACTA ACCTTGTCG G	NA	NA	yes	
<i>Pmir-48</i>	Intergenic	V	NA	NA	<i>gaggaggtcga</i> <i>ccctacattgac</i> <i>ctaatagac</i>	<i>gagcctgtcga</i> <i>cagttcccggg</i> <i>agtttcaattg</i>	<sup>c</sup> <i>Sall</i> sites used for cloning are in italics	yes
<i>Pmir-49</i>	Intergenic	X	ATGCGTTGT ATTGTTATC	GTACACTAG CATCTGAATA AA	NA	NA	no	
<i>Pmir-51</i>	Intergenic	IV	CACAATATAT GGGTGCGCC AGTC	CAAGTGAGC AGGCAAATG TTGG	<i>acgttgaggtcg</i> <i>acatatacgttg</i> <i>agattgactggt</i> <i>agattctg</i>	<i>aggagctagtc</i> <i>gacgacggac</i> <i>ttttcgacaag</i> <i>tgacagagcaa</i>	<sup>c</sup> <i>Sall</i> sites used for cloning are in italics	yes
<i>Pmir-52</i>	Intergenic	IV	GGCCACAT ACAAACTGC	GTTGGAAGA ACATGAATAT G	NA	NA	yes	
<i>Pmir-53</i>	Intergenic	IV	CGGATCCTT AAGTCAGAA ATTG	GAATCGGGA GAAATTTAT	NA	NA	yes	
<i>Pmir-57</i>	Intergenic	II	ATAATTCAAA CGGAACGAT CG	CGATGAGTT CACATACCT TTTTG	NA	NA	yes	
<i>Pmir-59</i>	Intergenic	IV	CTTCTTTCTA CCCATTCT G	ATGTCATATC TAGATCAAC TC	NA	NA	yes	
<i>Pmir-60</i>	Intergenic	II	AGAAGATGA TACACTTTTT	CGAGTAGAA TATGAATACG AA	NA	NA	yes	
<i>Pmir-63</i>	Intergenic	X	GTGTAAACT TTCGATATCA C	TGAATGAATA AAGTTGCTC C	NA	NA	yes	
<i>Pmir-72</i>	Intergenic	II	ACTCAATTC TCACCTCCA TTT	TCAGCAGAA TAGTAGTAGT AG	NA	NA	yes	
<i>Pmir-75</i>	Intergenic	X	GCCAAATAA AAGTTACATT T	AAGTAGAAA AAAAAACCG C	NA	NA	yes	
<i>Pmir-76</i>	Intergenic	III	TCAAAAACC AGGAAAG	TCTTACCTG AGGATTGTA G	NA	NA	yes	



<i>Pmir-77</i>	Intergenic	II	ACTTCGAGT TTTTGAAATC CG	CAAACCTCTAT ACCAATTTG G	NA	NA	yes	
<i>Pmir-78</i>	Intergenic	IV	TAGCGCCTG ATCTTGTCAT TTG	GTTATTTTAC GGACACTAT G	NA	NA	no	
<i>Pmir-79</i>	Intergenic	I	GATTAATTGA CATGATATC	TCTACTAATT TCAACTTAC C	NA	NA	yes	
<i>Pmir-83</i>	Intergenic	IV	AATACGAATT TCCCACC	CCGAGTGGT GCTTTTAAAT TG	NA	NA	yes	
<i>Pmir-84</i>	Intergenic	X	NA	NA	<i>aaataagtcga</i> <i>cactaaacaaa</i> <i>gtcagatcac</i>	<i>catactcccgg</i> <i>ggatgccaca</i> <i>ggcagacgtat</i> <i>g</i>	<sup>c</sup> <i>Sall</i> and <i>Xma</i> I sites used for cloning are in italics	yes
<i>Pmir-90</i>	Intergenic	III	CATTGTGTTT ACAATGCC ATGC	CGCCCGTTT TGTTAGAAAT GG	NA	NA	yes	
<i>Pmir-227-80</i>	Intergenic/o peron	III	TTGATTGGA AAATTGAAA GC	GAGTGTCCA TTGAAAAAG G	<i>cagggaaagtc</i> <i>gacatatgtgg</i> <i>aatgaatcattg</i> <i>gtagactgtg</i>	<i>tcatgtcggtcg</i> <i>acgagcgaac</i> <i>gagtgtccattg</i> <i>aaaaaggaat</i> <i>a</i>	<i>mir-227</i> not in miRNA Registry but present in WormBase. RNA was detected (Ambros <i>et al.</i> , 2003); <sup>c</sup> <i>Sall</i> sites used for cloning are in italics	yes
<i>Pmir-229-64-65-66</i>	Intergenic/o peron	III	GCAAATTGC CGGAATTG	GAGTCTTCC CGCGTTG	NA	NA	no	
<i>Pmir-42-43-44</i>	Intergenic/o peron	II	GATTAATTAT GGATGTGCG	GTTTATGGG GATTTTAGA GG	NA	NA	yes	
<i>Pmir-54-55-56</i>	Intergenic/o peron	X	TGATACCAA ACAATTGA GTG	ACATTACCAA TTCCAGAG	<i>agtggtcagtcg</i> <i>acctctccgac</i> <i>agataattcgat</i> <i>acaaatcag</i>	<i>gaagattagtc</i> <i>gaccaagtctt</i> <i>taaaaaagca</i> <i>atgttctcgtc</i>	<sup>c</sup> <i>Sall</i> sites used for cloning are in italics	yes
<i>Pmir-61-250</i>	Intergenic/o peron	V	TAGAGACAA AATTGATGTT C	ATAGGCAGG GAGTAGCAG	NA	NA	yes	
<i>Pmir-229-64-65-66</i>	Intergenic/o peron	III	AATTTTCTG GCAAATTGC CGG	GAGTCTTCC CGCGTTGG G	NA	NA	NA	
<i>Pmir-73-74</i>	Intergenic/o peron	X	TTGGCCAAT GGGGACGA A	GAATAAATG AATGAAGAG TG	NA	NA	no	
<i>Pmir-124</i>	Intragenic	IV	NA	NA	NA	NA	NA	

<i>Pmir-2</i>	Intragenic	I	TGTTGAATG GGAAAGTGC	AGATTCGTC GAAATACGA TTG	ttgtgaagggtcg acagagatgg aaggcaacac atggcgccgaa g	agctggctgtcg actgaaagtca taatttgaaca catcaacc	Seq upstream <i>mir-2</i> was taken as promoter (intron seq of F16A11.3); <sup>c</sup> Sall sites used for cloning are in italics	yes
<i>Pmir-233</i>	Intragenic	X	NA	NA	NA	NA		NA
<i>Pmir-798</i>	Intragenic/antisense	IV	NA	NA	NA	NA		NA
<i>Pmir-248</i>	Intragenic	X	NA	NA	NA	NA	Annotated in WS175 as intergenic.	NA
<i>Pmir-253</i>	Intragenic	V	NA	NA	NA	NA		NA
<i>Pmir-254</i>	Intragenic	X	NA	NA	NA	NA		NA
<i>Pmir-272</i>	Intragenic	III	NA	NA	NA	NA		NA
<i>Pmir-273</i>	Intragenic	II	NA	NA	NA	NA		NA
<i>Pmir-353</i>	Intragenic	I	NA	NA	NA	NA		NA
<i>Pmir-354</i>	Intragenic	I	NA	NA	NA	NA		NA
<i>Pmir-356</i>	Intragenic	III	NA	NA	NA	NA	located immediately downstream of ZK62.2.	NA
<i>Pmir-50</i>	Intragenic	I	NA	NA	NA	NA		NA
<i>Pmir-58</i>	Intragenic	IV	GACGAGCAT CGGGGGATA GGTC	GACGAGCAT CGGGGGATA GG	NA	NA	Seq upstream <i>mir-58</i> was taken as promoter (intron seq of Y67D8A.1).	yes
<i>Pmir-62</i>	Intragenic	X	NA	NA	NA	NA		NA
<i>Pmir-67</i>	Intragenic	III	NA	NA	NA	NA		NA
<i>Pmir-70</i>	Intragenic	V	NA	NA	NA	NA		NA
<i>Pmir-71</i>	Intragenic	I	CTATTTCTTC TACAGCAGT ACTCTC	ACGGCGAAA AACAGAATA GT	NA	NA		yes
<i>Pmir-82</i>	Intragenic	X	GAGTTCATA AAAACAAC GGC	GAGTTCATA AAAACAAC GG	NA	NA	Seq upstream <i>mir-82</i> was taken as promoter (intron seq of T07D1.2).	yes
<i>Pmir-86</i>	Intragenic	III	NA	NA	NA	NA		NA
<i>Pmir-87</i>	Intragenic	V	NA	NA	NA	NA		NA
<i>Pmir-260</i>	Intragenic/antisense	II	CAGGACCGA CGAAGGCAT	TTACATGGA GGTCCCGC	NA	NA		no
<i>Pmir-267</i>	Intragenic/antisense	II	GCAGCCGA CACATTACG GGATC	GGGTTTTGA AACTTTTATT TTA	NA	NA		yes

<i>Pmir-81</i>	Intragenic/antisense	X	GTGTAGCAA CCGTAGGAG	TTTTATGCAT TGAATAGAA	<i>taggagaacc</i> <i>gggttctggtcg</i> <i>actctaaaaatt</i> <i>atgaatgac</i>	<i>tcacgatgcc</i> <i>ggggaaaagc</i> <i>atctttggattgct</i> <i>ctcagatca</i>	<sup>c</sup> <i>Xma</i> I sites used for cloning are in italics	yes
<i>Pmir-85</i>	Intragenic/antisense	II	TTAACAAGT GTATTCCT	AGAATACGA ACGACGAGA AAA	NA	NA		yes
<i>Pmir-35-36-37-38-39-40-41</i>	Intragenic/antisense/operatoron	II	CCGTCATAT CTGACCCAC	AATAGTTGG GAATGGAAA TTAGG	NA	NA		yes
<i>Pmir-358-357</i>	Intragenic/antisense/operatoron	V	AAACTACAG TAATTCCTTA AATG	GAATGATTG GGAGAAAG GTGAAG	NA	NA		yes
<i>Pmir-789.1*</i>	Intragenic/antisense	IV	NA	NA	NA	NA		NA
<i>Pmir-789.2*</i>	Intragenic/antisense	IV	NA	NA	NA	NA		NA
<i>Pmir-799*</i>	Intragenic/antisense	X	NA	NA	NA	NA		NA
<i>Pmir-787*</i>	Intergenic	X	NA	NA	NA	NA		NA
<i>Pmir-784*</i>	Intergenic	X	CTTGTTGCA CTTAACTACA TAAATC	GAAAAACTG AGAATACGA ATAAAG	NA	NA		yes
<i>Pmir-790*</i>	Intergenic	IV	NA	NA	NA	NA		NA
<i>Pmir-791*</i>	Intergenic	X	NA	NA	NA	NA		NA
<i>Pmir-792*</i>	Intergenic	V	NA	NA	NA	NA		NA
<i>Pmir-793*</i>	Intergenic	X	CATTTAAATT ATTTCTCC CAAAG	CTCCGTAAG AAATATTATC TC	NA	NA		yes
<i>Pmir-794*</i>	Intergenic	I	GAGGGGTTT TTCAGTTTTT C	GACACGCTC AGTTTTTTTT TG	NA	NA		yes
<i>Pmir-795*</i>	Intergenic	I	TTGTACTION GCACACTTA	AGCAGGTAA CGTTTTCTCG	NA	NA		no
<i>Pmir-796*</i>	Intergenic	X	NA	NA	NA	NA		NA
<i>Pmir-800*</i>	Intergenic	X	NA	NA	NA	NA		NA
<i>Pmir-788*</i>	Intergenic	X	TTCATTCAC ACTTAAATGT TTATTC	CTTCCAGCT GAAAATTTG TAATAC	NA	NA		yes

NA: not attempted

<sup>a</sup>attB4 tail: ggggacaactttgtatagaaaagtq

<sup>b</sup>attB1R tail: aaaaactactttttatatacaacttatcat

<sup>c</sup>Restriction sites used to clone promoter into pPD95.75

Ch: chromosome

**Table III-2. Description of *PmiRNA::gfp* expression patterns.**

Promoter	Expression pattern description
<i>Plet-7</i>	Expression detected from late embryos to adults. In late embryos to L1, expression is seen in <i>dnc</i> and intestine. From L2 to adults, expression is detected in almost all tissues except germline; specifically muscle, rectum, vulva, spermatheca, pharynx, intestine, <i>vnc</i> , <i>dnc</i> , nerve ring, hypodermis.
<i>Plin-4</i>	Expression seen from late L1 to adult stages. Weak expression detected ubiquitously (except germline). Stronger in pharynx, vulva, vulval muscle, body wall muscle.
<i>Plsy-6</i>	No expression detected.
<i>Pmir-1</i>	Expression detected from early embryos to adults specifically in pharynx and precursors.
<i>Pmir-2</i>	Expressed from late embryos to adulthood. Strong expression detected in many nerves, nerve ring, <i>dnc</i> , <i>vnc</i> and also nerves in the tail.
<i>Pmir-227-80</i>	Expressed from late embryos to adulthood. Expressed in excretory cells, uterus, vulva, <i>dnc</i> , distal tip cells, posterior intestine, amphid neurons, head and body wall muscle.
<i>Pmir-228</i>	Expressed from mid embryo continuing through adulthood. Expressed in amphids, excretory cells, seam cells, vulva, body neurons, rectum and posterior intestine.
<i>Pmir-230</i>	Expressed from L1 to adulthood. Strong expression in rectum, seam cells (L1-L4) and weak expression in hypodermis (not expressed in <i>vnc</i> and <i>dnc</i> ). Also, expressed in vulva muscles.
<i>Pmir-231</i>	Expressed from early embryo continuing through adulthood. In early embryos, expression is detected on the lateral sides and in mid embryo stages, expression is detected on the posterior part only. In larval and adult stages, expression is seen in posterior areas of intestine, rectum and tail hypodermis. Weak expression in pharynx, <i>vnc</i> , vulva, <i>dnc</i> and head nerves and tail nerves.
<i>Pmir-232</i>	Expression detected from late embryos till adulthood. Expressed in excretory cells and canals.
<i>Pmir-234</i>	Expressed from mid embryo continuing through adulthood. Expression seen in head nerves, specifically in 6 pairs in nerve ring and also <i>dnc</i> . Expressed also in tail nerves, specifically in 2 pairs and one single neuron with axon ventral to head. Expression seen in body nerves, specifically in CAN nerves. Also seen in amphid neurons in the pharynx; the cell body is posterior/lateral to posterior pharyngeal bulb (ppb) and axon runs to tip of head. Expression seen in I4, I5 and I6 neurons, cell body in ppb and axon runs to anterior pharyngeal bulb (apb).
<i>Pmir-235</i>	Expressed from late embryos to adulthood. Expression detected in hypodermis, especially at L1-L2. Also detected in vulva, rectum and some amphid neurons.
<i>Pmir-236</i>	Expressed from mid embryo to adulthood. Expression seen in intestine, rectal glands and <i>dnc</i> .
<i>Pmir-237</i>	Expressed from L1 to adulthood. Expressed in distal tip cells from L1 to adults. Also expressed in hypodermis, starting at L2 and strong at L4 and adults. In vulva, expression is seen from L2 till adulthood. Also expressed in head neurons and seam cells.

<i>Pmir-238</i>	Expressed in the intestine from early embryo continuing through adulthood. Also, hypodermal expression from L1 to adults. Bright gfp expression detected in rectal glands.
<i>Pmir-240-786</i>	Expression detected from L3 to adulthood. Expression seen in somatic cells at uterus, ovary, spermatheca, spermatheca/uterus valve and gonadal sheath.
<i>Pmir-241</i>	Expression detected from L1 till adulthood. Expressed in nerves: vnc, dnc, many in nerve ring, pharyngeal nerves and tail nerves. Also, expression seen in muscles from vulva, head, body and rectum and hypodermis.
<i>Pmir-242</i>	Expressed from late embryos to adulthood. Expressed in head nerves more specifically in 3 pairs of amphid neurons, of which, 2 pairs of nuclei are in the anterior to apb and the other pair, posterior to apb. Expression is also seen in pharyngeal/intestine valve and rectal gland.
<i>Pmir-243</i>	Expressed from mid embryo continuing through adulthood. Expressed very strongly in intestine. Also expressed in dnc and valve between spermatheca and uterus.
<i>Pmir-244</i>	Expressed in seam cells from late embryos till adults.
<i>Pmir-245</i>	Expressed from late embryos to adulthood. Expressed in a subset of pharyngeal muscles (pm3, 4 and 7). Expression seen also in head nerves and dnc. Expressed weakly in tails, perhaps in tail hypodermis during adulthood but seen also at earlier stages.
<i>Pmir-246</i>	Expressed from L4 to adults. Expression seen in gonadal sheath and uterine/spermatheca valve.
<i>Pmir-247-797</i>	Expression detected in pharynx from late embryo thorough L1. From L1 to L2/L3, expression is also detected in rectal glands and from L2 to L4, in distal tip cells
<i>Pmir-251</i>	Expression seen from late embryos to adulthood. Expressed in pharyngeal muscles
<i>Pmir-252</i>	Expressed in larval stages to adulthood in the pharyngeal glands, spermatheca and uterine cells, more specifically in a pair of cells closest to vulva. Also in head neurons.
<i>Pmir-255</i>	Expressed from comma stage to adults. Expression seen in pharyngeal muscle cells, specifically in the 3 pm7 cells in the posterior half of the posterior pharyngeal bulb.
<i>Pmir-257</i>	No expression detected.
<i>Pmir-258</i>	No expression detected.
<i>Pmir-259</i>	Expression seen from mid embryos to adults. In mid embryos, expression is detected in few cells in the posterior and anterior part. In larval stages and adults, expression is seen in rectal glands, pharyngeal/intestinal valve. Expression detected also in the reproductive tract, specifically in somatic cells of the uterus, spermatheca and spermatheca-uterus valve.
<i>Pmir-261</i>	No expression detected.
<i>Pmir-265</i>	Weak expression detected in head neurons and somatic cells of the uterus from L2 to adults.
<i>Pmir-266</i>	Very weak expression detected in dnc from L1 to adulthood.
<i>Pmir-267</i>	No expression detected.
<i>Pmir-268</i>	Expression seen from late embryos to adulthood in dnc, vnc, other head/body and tail nuclei nerves as well as pharyngeal nerve in anterior bulb.

<i>Pmir-269</i>	Expression detected in L4 to adult worms in head neurons.
<i>Pmir-270</i>	Expression detected from L1 to adults in the intestine and a couple of head neurons. Mosaic expression.
<i>Pmir-271</i>	No expression detected.
<i>Pmir-35-41</i>	Expressed from early embryo continuing through adulthood. Expression is strong in the earlier stages and becomes weaker in the later stages. Expressed in muscles: vulva, head and rectal muscles. Also, expressed in seam cells, hypodermis, partial intestine and nerves, including dnc, head and tail nerves.
<i>Pmir-355</i>	Expression detected from mid embryo to adulthood. Strong embryonic expression seen in lateral cells and intestine precursors. In larval stages, expression is detected in the intestine and in adult stages, only in the posterior part of the intestine.
<i>Pmir-357-358</i>	Expressed in somatic cells of the uterus and precursor and distal tip cells. Expression starting at L3 stage and continuing through adulthood.
<i>Pmir-359-785</i>	Expressed from L1 to adulthood. Expression seen in spermatheca and vulval muscles and somatic cells of the uterus, including the dtc. In addition, expression is detected in hypodermis, dnc, rectal muscles and anterior body wall muscles.
<i>Pmir-360</i>	Expressed from comma stage to adults. Expression seen in marginal cells of pharynx.
<i>Pmir-392</i>	Expression detected from L1 to adults. Expression seen in nerves whose processes run dorsally to the tail tip and ventrally to nerve ring in the head. In addition, expression seen in posterior intestine and nerves in vulva.
<i>Pmir-42-44</i>	Expression seen in early embryos and continuing through adulthood. In the embryo, expression is seen as stripes on the outside part of the embryo. In the mid embryo stage, gfp is seen on dorsal and ventral part of the embryo, including head. Late embryos show complex expression. Strong expression in seam cells and vulva. Weaker expression in hypodermis. Also seen in posterior intestine and rectum and dnc.
<i>Pmir-45</i>	Expression seen in all stages in the intestine, although it varies from anterior and/or posterior part. From late embryo till L1, strong expression is seen in the posterior part of the pharynx. Expresses also in dnc and a pair of nerves in the head at each side of the posterior pharyngeal bulb, in all stages and two amphids from L2 on. The late embryo stage shows a complex pattern of expression especially in lateral stripes. Weak expression in head muscle is also seen.
<i>Pmir-46</i>	Expressed from early embryo continuing through adulthood. In the embryonic stage, expression is detected on the lateral sides. Later on, expression is seen in hypodermis and seam cells
<i>Pmir-47</i>	Expression detected from late embryos to adults. In embryos, expression is detected on lateral sides and in larval stages on, in all hypodermis, rectum and vulval cells.
<i>Pmir-48</i>	Expressed from L1 to adulthood. Expression seen in vulval cells, seam cells, head and tail neurons.

<i>Pmir-51</i>	Expression detected from mid embryos to adults. In late embryos, expression is seen on one side of the embryo, ventral and mostly anterior. From late embryo to L1, expression is detected in canal cells and canal nerves. Also, from late embryos to adults, expression is detected in several nerves, including dnc and vnc. In addition, expression is detected in head muscles, coelomocytes and intestine.
<i>Pmir-52</i>	Expression detected in L4 continuing to young adults only. Expressed in somatic cells of the uterus and subset of neurons at the vnc.
<i>Pmir-53</i>	Expression seen ubiquitously (except germline) from pre-comma stage to adulthood.
<i>Pmir-54-56</i>	Expression seen from mid embryos to adulthood. Expressed strongly in nerve in head (amphid) and tail. Weak expression in external vulval cells, dnc and vnc is also detected.
<i>Pmir-57</i>	Expression seen in embryos starting at the comma stage and continuing throughout adulthood. Posterior regions: intestine, ventral nerve cord, rectum, ventral muscles and tail hypodermis. Also seen in vulval muscles from L2 till adulthood.
<i>Pmir-58</i>	No expression detected.
<i>Pmir-59</i>	Expressed from late embryo stage till adults. Expression seen in nerves, dnc, vnc, two nerves in tail whose processes are ventral to the head, two nerves in head which are anterior to the posterior pharyngeal bulb. Also expressed in the posterior part of the intestine and rectal muscles. In adults, expression is seen also in vulval muscles.
<i>Pmir-60</i>	Expressed in the intestine from early embryos to adults.
<i>Pmir-61-250</i>	Expression detected from early embryos to adults. Detected in dnc, two head nerves whose processes run ventral to each other and laterally towards the tip of the head. Also seen in tail nerves and posterior intestine. In L2, expression is also detected in vulval cell precursors.
<i>Pmir-63</i>	Expression detected from comma stage to adults. Expression seen in intestine.
<i>Pmir-71</i>	Expression detected from L1 to adult. Expression becomes stronger in later stages and in adults is almost ubiquitous. Strongest expression seen in nerve ring, hypodermis, vulva and pharynx.
<i>Pmir-72</i>	Expression seen in late embryos continuing to adulthood. Expression seen in pharynx and amphids, specifically 3 pairs of nerves with cell bodies just posterior to the anterior pharyngeal bulb.
<i>Pmir-75</i>	Expression detected from mid embryos till adults. Expression seen exclusively in intestine.
<i>Pmir-76</i>	Expression seen from late embryo to adult stages. Strong expression in earlier stages. Detected in pharyngeal neurons; two cell bodies in the posterior and two in the anterior part of the bulb. Expression seen also in body neurons, canal nerves, and tail neurons.
<i>Pmir-77</i>	No expression detected.
<i>Pmir-784</i>	Expression detected from L1 to adults in head nerves some of which could be pharyngeal nerves. Also, expressed from L3 on in a set of vulva cells (somatic gonad).
<i>Pmir-788</i>	Expression detected from late embryos to adults. Expression seen in all hypodermal cells and rectal gland.

<i>Pmir-79</i>	Expressed from mid embryo continuing through adulthood. In mid embryos, expression is detected on lateral sides. Later in development, expression is detected in hypodermis.
<i>Pmir-793</i>	Expressed from late embryo stage till adults. Expression seen in head and tail neurons.
<i>Pmir-794</i>	Expressed from late embryos to adulthood. Expressed in head and body wall muscle and intestine (mosaic).
<i>Pmir-81</i>	Very weak expression detected in head neurons from L1 on.
<i>Pmir-82</i>	Expression detected from L4 to adults in the pharyngeal muscle, developing spermatheca, subset of ventral nerve cord and a subset of
<i>Pmir-83</i>	Expressed from early embryo continuing through adulthood. Expression detected in many nerves from head and tail, the most obvious are amphids and phasmids. Also, in <i>dnc</i> , <i>vnc</i> , canal nerves, intestine, rectal gland, spermatheca/uterine valve.
<i>Pmir-84</i>	Expression seen from L1 to adult stages. Strong expression seen in pharynx, vulva and rectal glands. Weak expression seen in hypodermis and seam cells. Also seen in canal nerves, spermatheca/uterine valve and head neuron. Posterior intestine expression also detected.
<i>Pmir-85</i>	Expression seen from L2 to adulthood. Somatic cells at the uterus and spermatheca; also seen in rectal glands and gonadal sheath.
<i>Pmir-90</i>	Expression seen in mid embryos, specifically on the lateral sides of the embryo, continuing to adulthood. GFP expression is seen in head and body wall muscles, amphids and phasmids, including other nerve nuclei in head but cannot see processes. Also, expressed in <i>vnc</i> and lateral nerves of body. Expression seen also in vulval muscles and cells at uterus, proximal to the vulva.





5	high	<i>Pmir-35-41</i>	1 1 1 1 1 1 1 1	0 1 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 1 0 1 1 1 0 1
3	low	<i>Pmir-355</i>	0 1 1 1 1 1 1 1	1 0
2	high	<i>Pmir-357-358</i>	0 0 0 0 0 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
4	low	<i>Pmir-359-785</i>	0 0 0 1 1 1 1 1	0 0 0 0 0 1 1 1 0 0 0 1 1 1 0 0 1 0 1 0 1 0 1
7	high	<i>Pmir-360</i>	0 1 1 1 1 1 1 1	0 0 1 0
4	high	<i>Pmir-392</i>	0 0 0 1 1 1 1 1	0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0
2	low	<i>Pmir-42-44</i>	1 1 1 1 1 1 1 1	0 1 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0
3	high	<i>Pmir-45</i>	1 1 1 1 1 1 1 1	1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0
2	high	<i>Pmir-46</i>	1 1 1 0 0 1 1 1	0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
8	high	<i>Pmir-47</i>	0 0 1 1 1 1 1 1	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
2	high	<i>Pmir-48</i>	0 0 0 1 1 1 1 1	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0
1	high	<i>Pmir-51</i>	0 1 1 1 1 1 1 1	1 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 1 0 1 0
5	high	<i>Pmir-52</i>	0 0 0 0 0 0 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0
5	high	<i>Pmir-53</i>	0 1 1 1 1 1 1 1	1 0 1
5	high	<i>Pmir-54-56</i>	0 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
5	high	<i>Pmir-57</i>	0 1 1 1 1 1 1 1	0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0
2	high	<i>Pmir-58</i>	0 0 0 0 0 0 0 0	0 0
2	high	<i>Pmir-59</i>	0 0 1 1 1 1 1 1	0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
8	high	<i>Pmir-60</i>	1 1 1 1 1 1 1 1	0 1 0
4	high	<i>Pmir-61-250</i>	1 1 1 1 1 1 1 1	0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0
2	high	<i>Pmir-63</i>	0 1 1 1 1 1 1 1	1 0
4	high	<i>Pmir-71</i>	0 0 0 1 1 1 1 1	1 1
6	low	<i>Pmir-72</i>	0 0 1 1 1 1 1 1	0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
3	high	<i>Pmir-75</i>	0 1 1 1 1 1 1 1	1 0
3	high	<i>Pmir-76</i>	0 0 1 1 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
11	high	<i>Pmir-77</i>	0 0 0 0 0 0 0 0	0 0
3	low	<i>Pmir-784</i>	0 0 0 1 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0
14	high	<i>Pmir-788</i>	0 0 1 1 1 1 1 1	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2	high	<i>Pmir-79</i>	0 1 1 1 1 1 1 1	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2	high	<i>Pmir-793</i>	0 0 1 1 1 1 1 1	0 1 1 0 0
7	high	<i>Pmir-794</i>	0 0 1 1 1 1 1 1	1 0 1 1
2	high	<i>Pmir-81</i>	0 0 0 1 1 1 1 1	0 1 0 0 0
4	high	<i>Pmir-82</i>	0 0 0 0 0 0 1 1	0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 0 0 0
6	high	<i>Pmir-83</i>	1 1 1 1 1 1 1 1	1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0
2	high	<i>Pmir-84</i>	0 0 0 1 1 1 1 1	0 1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0
8	high	<i>Pmir-85</i>	0 0 0 0 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0
3	high	<i>Pmir-90</i>	0 1 1 1 1 1 1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 1

TR: Transmission Rate

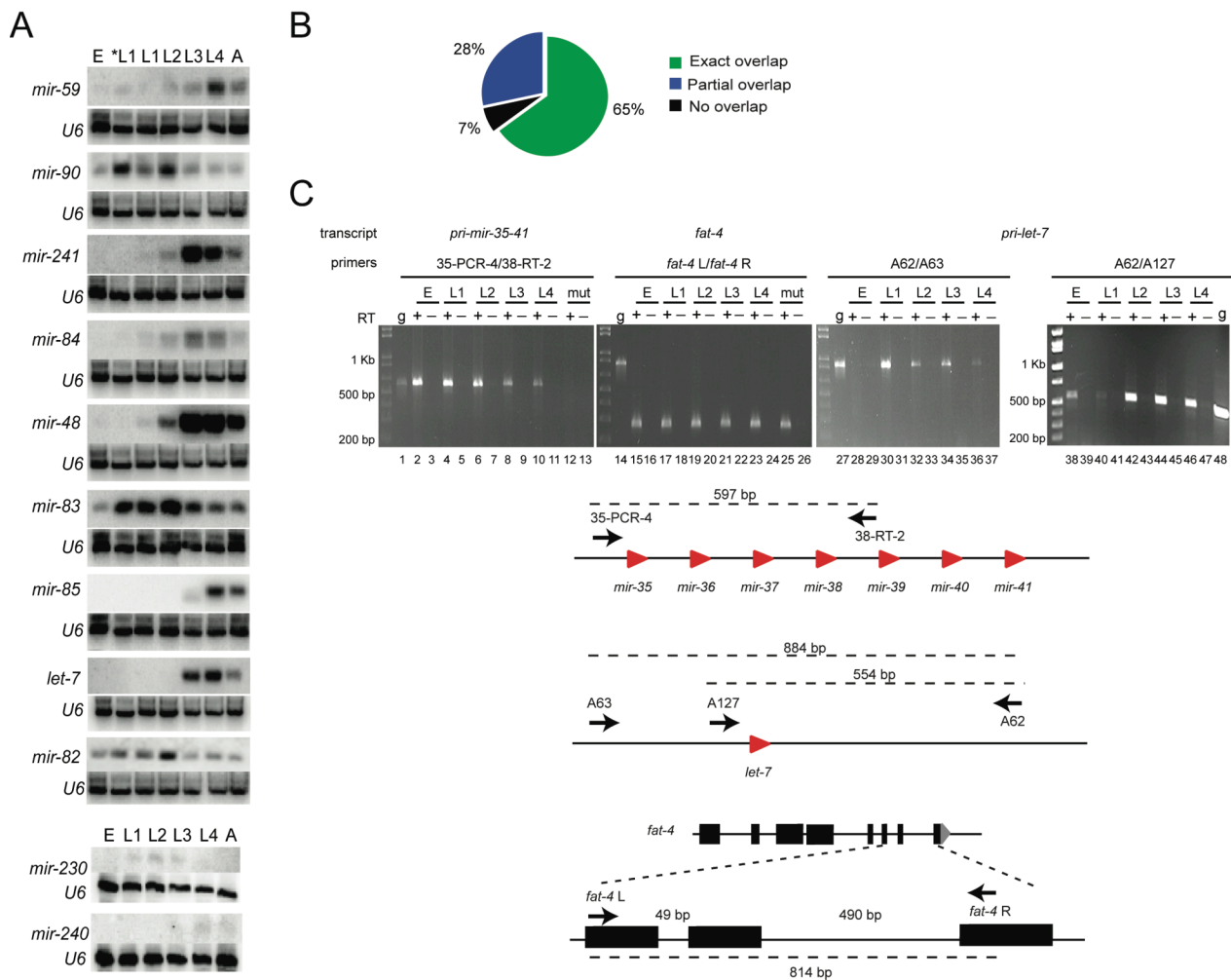
**Table III-4. Temporal expression available for each miRNA.**

<i>PmiRNA</i>	Northern blot temporal expression from Wormbase (WS180)	Northern blot temporal expression from this work	Observed <i>PmiRNA::gfp</i> temporal expression
<i>Plet-7</i>	L3 to adult	L3 to adult	all stages
<i>Plin-4</i>	L1 to adult		L1 to adult
<i>Plsy-6</i>	NA		no expression
<i>Pmir-1</i>	all stages		all stages
<i>Pmir-2</i>	all stages		all stages
<i>Pmir-227-80</i>	all stages ( <i>mir-227</i> is now <i>mir-80</i> )		all stages
<i>Pmir-228</i>	all stages		all stages
<i>Pmir-230</i>	all stages	L1 to L3	L1 to adult
<i>Pmir-231</i>	all stages		all stages
<i>Pmir-232</i>	all stages		all stages
<i>Pmir-234</i>	all stages		all stages
<i>Pmir-235</i>	all stages		all stages
<i>Pmir-236</i>	all stages		all stages
<i>Pmir-237</i>	L3 to adult		L1 to adult
<i>Pmir-238</i>	all stages		all stages
<i>Pmir-240-786</i>	L4 to adult-NA	L4 to adult-NA	L3 to adult
<i>Pmir-241</i>	L3 to adult	L1 to adult	L1 to adult
<i>Pmir-242</i>	all stages		all stages
<i>Pmir-243</i>	all stages		all stages
<i>Pmir-244</i>	all stages		all stages
<i>Pmir-245</i>	L4 to adult		all stages
<i>Pmir-246</i>	L4 to adult		L4 to adult
<i>Pmir-247-797</i>	L3 to adult-NA		all stages
<i>Pmir-251</i>	L1 to adult		all stages
<i>Pmir-252</i>	L1 to adult		L1 to adult
<i>Pmir-255</i>	NA: mature miRNA not detected, only precursor detected in dicer mutants		all stages
<i>Pmir-257</i>	all stages		no expression
<i>Pmir-258</i>	all stages		no expression
<i>Pmir-259</i>	all stages		all stages
<i>Pmir-261</i>	postembryonic expression		no expression
<i>Pmir-265</i>	NA		L2 to adult
<i>Pmir-266</i>	NA		L2 to adult
<i>Pmir-267</i>	NA		no expression
<i>Pmir-268</i>	NA		all stages
<i>Pmir-269</i>	NA		L4 to adult
<i>Pmir-270</i>	NA		L1 to adult
<i>Pmir-271</i>	NA		no expression
<i>Pmir-35-36-37-38-39-40-41</i>	expressed only in embryogenesis		all stages
<i>Pmir-355</i>	NA		all stages
<i>Pmir-357-358</i>	NA-NA		L3 to adult
<i>Pmir-359-785</i>	NA-NA		L1 to adult
<i>Pmir-360</i>	NA		all stages
<i>Pmir-392</i>	NA		L1 to adult

<i>Pmir-42-43-44</i>	<i>mir-42</i> and <i>mir-43</i> expressed predominantly in embryos while <i>mir-44</i> is expressed in all stages		all stages
<i>Pmir-45</i>	all stages		all stages
<i>Pmir-46</i>	all stages		all stages
<i>Pmir-47</i>	all stages		all stages
<i>Pmir-48</i>	L3 to adult	L1 to adult	L1 to adult
<i>Pmir-51</i>	all stages		all stages
<i>Pmir-52</i>	expressed in high levels in L1 and weakly in the rest of the stages		L4 to adult
<i>Pmir-53</i>	all stages		all stages
<i>Pmir-54-55-56</i>	all stages-all stages-NA		all stages
<i>Pmir-57</i>	all stages		all stages
<i>Pmir-58</i>	all stages		no expression
<i>Pmir-59</i>	NA	all stages	all stages
<i>Pmir-60</i>	all stages (highly expressed at L1)		all stages
<i>Pmir-61-250</i>	all stages- L1 to adult		all stages
<i>Pmir-63</i>	all stages		all stages
<i>Pmir-71</i>	all stages		L1 to adult
<i>Pmir-72</i>	all stages		all stages
<i>Pmir-75</i>	all stages		all stages
<i>Pmir-76</i>	NA		all stages
<i>Pmir-77</i>	L1 to adult		no expression
<i>Pmir-784</i>	NA		L1 to adult
<i>Pmir-788</i>	NA		all stages
<i>Pmir-79</i>	all stages		all stages
<i>Pmir-793</i>	NA		all stages
<i>Pmir-794</i>	NA		all stages
<i>Pmir-81</i>	expressed highly at L1		L1 to adult
<i>Pmir-82</i>	all stages	all stages	L4 to adult
<i>Pmir-83</i>	L2 to adult	all stages	all stages
<i>Pmir-84</i>	L2 to adult	L1 to adult	L1 to adult
<i>Pmir-85</i>	highly expressed at L3 and L4	L3 to adult	L2 to adult
<i>Pmir-90</i>	NA	all stages	all stages

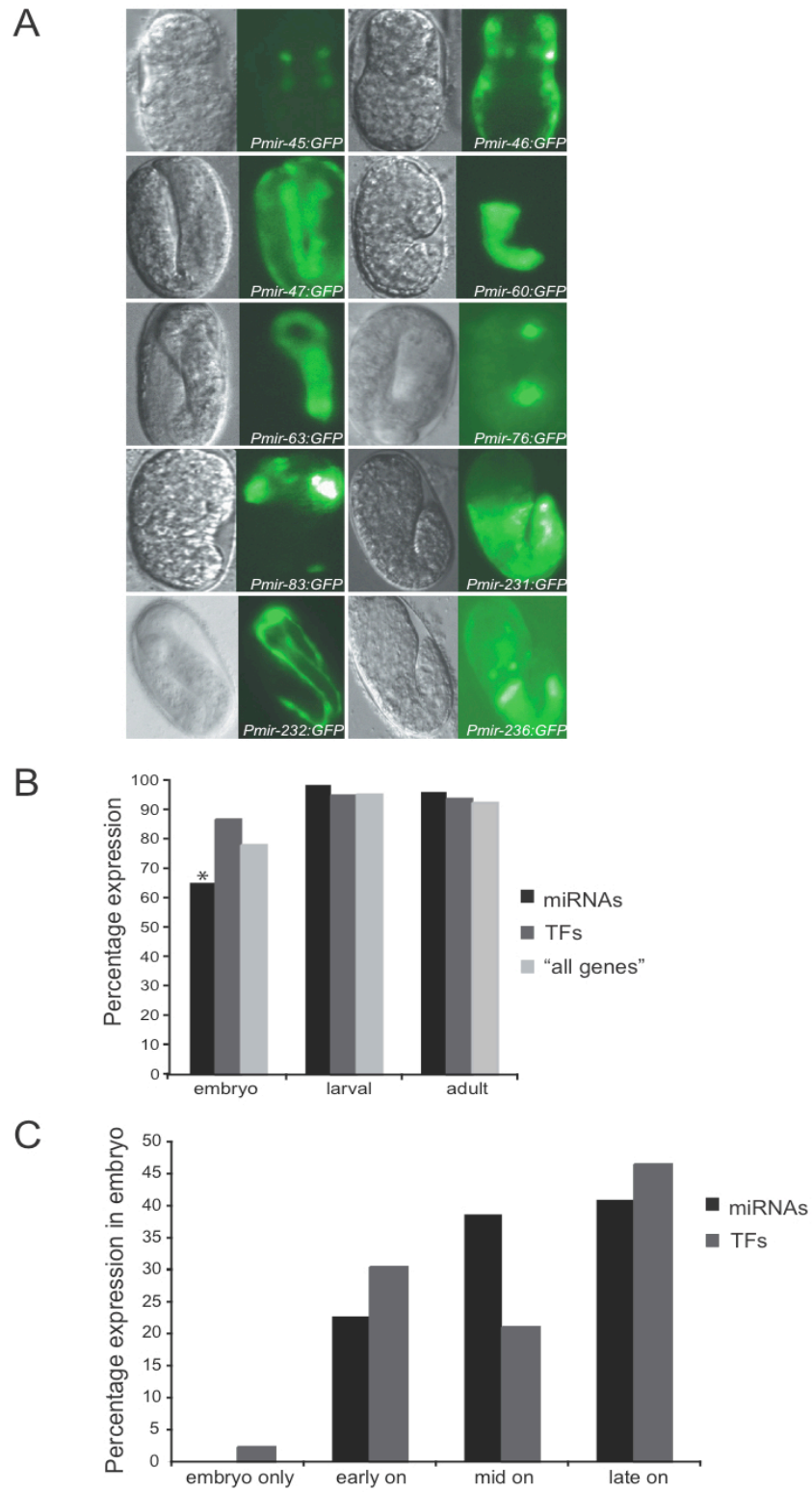
NA: No information available

**Figure III-2.**



**Figure III-2.** *Temporal PmiRNA::gfp activity correlates with Northern blot analysis and uncovers possible post-transcriptional mechanisms that control miRNA expression.* (A) Northern blot analyses using StarFire probes detect temporal expression of mature miRNAs. E – embryo; L – larvae; \*L1 – starved L1, A – adult. Probe against the U6 snRNA was used as control. (B) Comparison between miRNA expression determined by Northern blotting and promoter::gfp reporters. (C) Detection of *mir-35-41* and *let-7* primary transcripts by RT-PCR. As control, we used primers to amplify a protein-coding mRNA (*fat-4*). Total RNA from N2 embryos, L1, L2, L3 and L4 stages and total RNA from VC514 *mir-35-41* mutant embryos (mut) were subjected to reverse transcription (+RT, lanes 2, 4, 6, 8, 10, 12, 15, 17, 19, 21, 23, 25, 28, 30, 32, 34, 36, 38, 40, 42, 44 and 46) or mock reactions (-RT, lanes 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 24, 26, 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47). Genomic DNA was used as size marker (g, lanes 1, 14, 27 and 48). Cartoons indicate the predicted size of PCR amplicons from *mir-35-41* primary transcript, *let-7* primary transcript and *fat-4* mRNA and indicate the primers that were used. Note that *fat-4* L and R primers amplify a product of different size when genomic DNA (lane 14) or cDNA (lanes 15 - 26) was used as template.

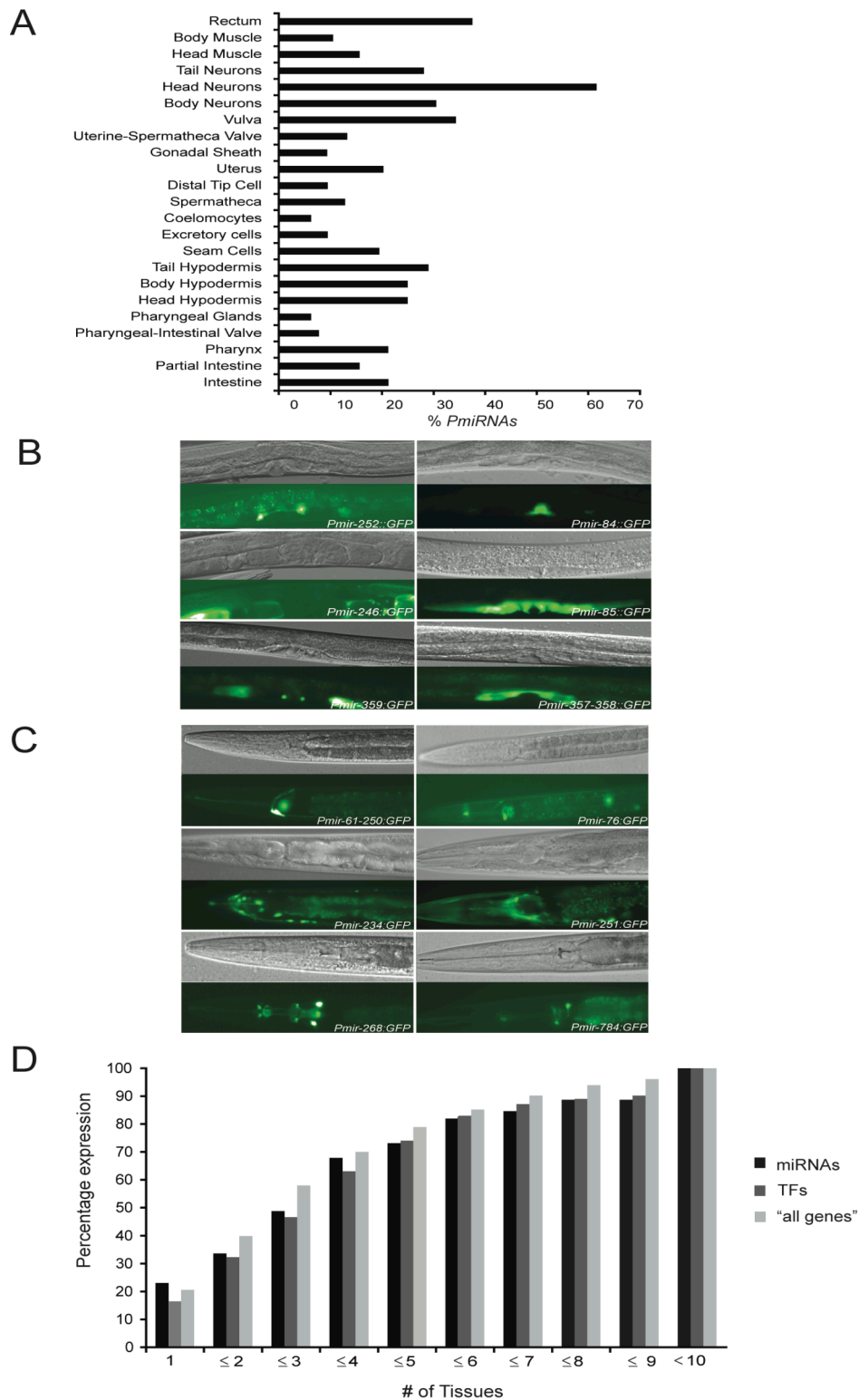
Figure III-3.



**Figure III-3. Temporal miRNA promoter activity.** (A) Representative images of miRNA promoters that drive GFP expression in the embryo. Left – DIC images; right – GFP fluorescence. Additional images can be found in the EDGEdb database (Barrasa et al. 2007). (B) miRNAs tend to be expressed later in development compared to TFs and “all genes”. The asterisk indicates a significant difference ( $P < 0.05$ ). (C) Percentage of miRNA and TF promoters that drive expression in the embryonic stage: embryo only, early, mid and late embryonic stages.



Figure III-4.



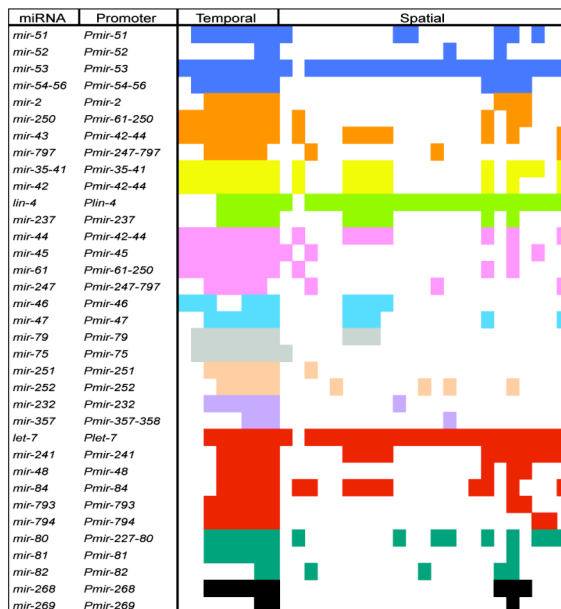
**Figure III-4. Spatial miRNA promoter activity.** (A) Percentage of miRNA promoters that drive expression in a survey of tissues. (B) Multiple miRNA promoters drive expression in various parts of the somatic gonad, including gonadal sheath, vulva and uterus. Top – DIC images; bottom – GFP fluorescence. (C) Multiple miRNA promoters drive GFP expression in the nervous system. Top – DIC images; bottom – GFP fluorescence. Additional images can be found in the EDGEdb database (Barrasa et al. 2007). (D) Most miRNAs as well as TFs and “all genes” promoters confer GFP expression in a tissue-specific manner.

**Table III-5.** *Explanation of the spatiotemporal annotation scheme.*

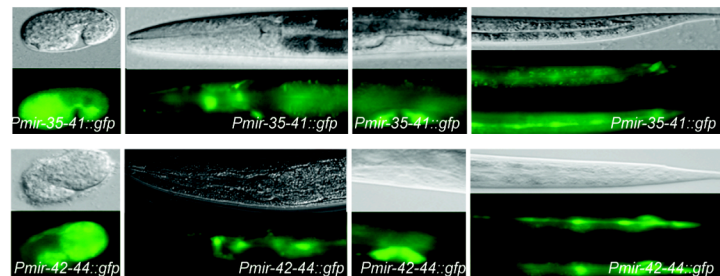
Code	Definition	Comments
0	Early Embryo	
1	Mid Embryo	
2	Late Embryo	
3	L1 larvae	
4	L2 larvae	
5	L3 larvae	
6	L4 larvae	
7	Adult	
I	Intestine	
PI	Partial Intestine	Include potentially spurious posterior/anterior intestine
P	Pharynx	Includes pharyngeal muscle, bulbs, isthmus and arcade
PIV	Pharyngeal-Intestinal Valve	
PG	Pharyngeal Glands	
HH	Head Hypodermis	
HB	Body hypodermis	
TH	Tail Hypodermis	
SC	Seam Cells	
X	Excretory System	Includes excretory cells and canal cells
C	Coelomocytes	
S	Spermatheca	
DTC	Distal Tip Cells	
U	Uterus	
GS	Gonadal Sheath	
USV	Uterine-Spermatheca Valve	
V	Vulva	Includes vulva muscle, external vulval cells (vulva hypodermis)
BN	Body Neurons	Neurons in which cell body lies along the body. Includes ventral nerve cord, canal nerves and vulva neurons.
HN	Head Neurons	Neurons in which cell body lies in the head. Includes dorsal nerve cord, amphids, phasmids, pharyngeal nerves, nerve ring
TN	Tail Neurons	Neurons in which cell body lies in the tail.
HM	Head Muscle	
BM	Body Muscle	
R	Rectum	Includes rectal glands and muscle

Figure III-5.

A

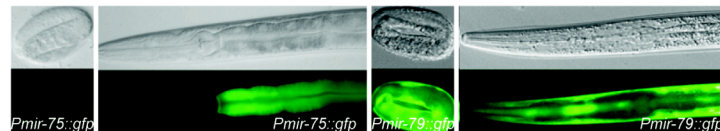


B



*mir-35 family*

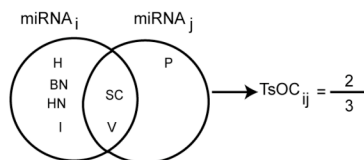
C



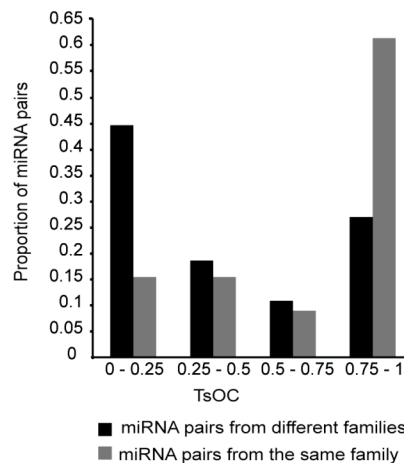
*mir-75 family*

D

$$TsOC_{ij} = \frac{\text{number of shared tissues/cell types}}{\text{minumum total number of tissues/cell types}}$$

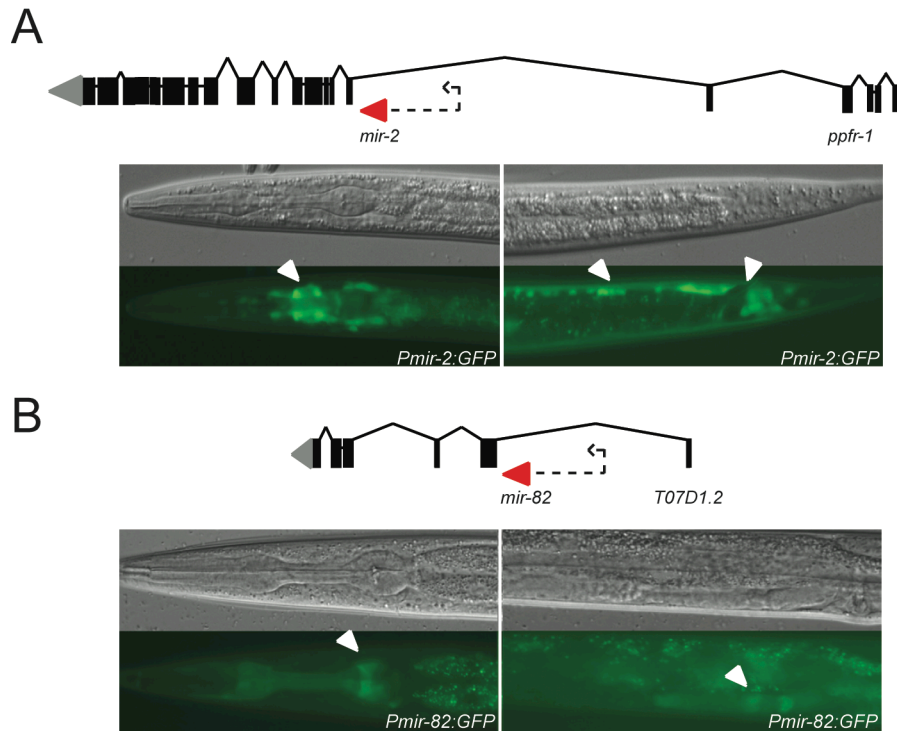


E



**Figure III-5.** *MiRNAs from a given family can have overlapping as well different spatiotemporal expression patterns.* (A) Cartoon depicting expression patterns of nine complete and two incomplete (*let-7* and *mir-80* families shown at the bottom) miRNA families. Each color represents a family. Spatiotemporal expression is as in Supplementary Table S3. (B) miRNAs from the *mir-35* family are expressed in overlapping tissues/cell types. *Pmir35-41::gfp* and *Pmir-42-44::gfp* are shown. (C) miRNAs from the *mir-75* family are expressed in different tissues/cell types. Top – DIC images; bottom – GFP fluorescence. (D) Definition and example of TsOC between any two miRNAs. (E) Distribution of TsOC among miRNA pairs from the same or different family.

Figure III-6.



**Figure III-6.** *Upstream sequences of intragenic miRNAs can drive GFP expression in vivo.* (A) *Pmir-2* drives expression in the nerve ring (left), ventral nerve cord and tail neurons (right). (B) *Pmir-82* drives expression in pharyngeal muscle and head neurons (left), and developing spermatheca (right). Top – DIC images; bottom – GFP fluorescence. Arrows indicate expression. Dotted arrows indicate sequence used as promoter.

## PREFACE TO CHAPTER IV

The work presented in the following chapter describes the characterization of the FLYWCH family of TFs and their role in the expression of miRNAs in the *C. elegans* embryo. This work is also part of a collaboration with the Ambros lab and part of this chapter was published separately in:

Ow, M. C., Martinez, N. J., Olsen, P. H., Silverman, H. S., Barrasa, M. I., Conradt, B., Walhout, A. J., and Ambros, V. 2008. The FLYWCH transcription factors FLH-1, FLH-2, and FLH-3 repress embryonic expression of microRNA genes in *C. elegans*. *Genes Dev* **22**: 2520-34.

This chapter embodies the joint effort of several people: M. C. Ow is the principal author of this work and the person who carried out the bulk of the experiments and wrote the original manuscript; M. I. Barrasa contributed to the identification of an FLH-1 and FLH-2 consensus binding sites and Figure IV-9 (not in published manuscript). I included additional figures as well as extensive changes in the text and tables that are not part of the published manuscript. Overall, I contributed to Figures IV-3, IV-7 A, IV-8 B, IV-8 C, IV-10, IV-12 and IV-13 and Table IV-3.

## CHAPTER IV

**The FLYWCH transcription factors FLH-1, FLH-2 and FLH-3 repress  
embryonic expression of microRNA genes in *C. elegans***



**Abstract**

MicroRNAs are small non-coding RNAs that mainly repress gene expression post-transcriptionally by antisense base pairing to target mRNAs. Although miRNAs are involved in a variety of biological functions, little is known about their transcriptional regulation. Using yeast one-hybrid assays, we found that transcription factors with a FLYWCH Zn-finger DNA-binding domain bind to the promoters of several *Caenorhabditis elegans* miRNA genes. We found that three FLYWCH transcription factors, FLH-1, FLH-2 and FLH-3, function redundantly to repress embryonic expression of *lin-4*, *mir-48*, and *mir-241*, miRNA genes that are normally expressed only post-embryonically. Although single mutations in either *flh-1*, *flh-2* or *flh-3* genes result in a viable phenotype, double mutation of *flh-1* and *flh-2* results in early larval lethality and an enhanced derepression of their target miRNAs in embryos. Mutations of *lin-4* or *mir-48&mir-241* do not rescue the lethal *flh-1; flh-2* double-mutant phenotype, suggesting that the early lethality phenotype is not solely the result of precocious expression of these miRNAs.

## Introduction

MicroRNAs (miRNAs) are an abundant class of small non-coding regulatory RNAs found in plants and animals. This ancient class of regulatory RNAs modulates a variety of biological processes including developmental timing, metabolism, and cell fate through base-pairing with the 3' untranslated region (UTR) of their target mRNAs (Ambros 2004; Bushati and Cohen 2007). Most animal miRNAs are transcribed by RNA polymerase II as part of longer primary transcripts (pri-miRNAs) that are then processed in a stepwise manner by protein complexes that include the RNase III enzymes Drosha and Dicer to produce the mature 21–22-nt miRNAs (for review, see Kim 2005).

Studies in mammals have shown that the biogenesis of some miRNAs can be regulated at the transcriptional level. For example, the proto-oncogene c-Myc directly activates the transcription of the *mir-17-92* cluster, and c-Myc-induced overexpression of *mir-17-92* induces tumor angiogenesis (O'Donnell et al. 2005; Coller et al. 2007). In contrast, c-Myc expression in lymphoma cells results in the transcriptional repression of a broad repertoire of miRNAs (Chang et al. 2008). Repression of *mir-124a* transcription by the RE1 silencing transcription factor (REST) contributes to the maintenance of neuronal identity (Conaco et al. 2006). Also, the myogenic transcription factors myogenin and myogenic differentiation 1 (MyoD) have been implicated in regulating the expression of two muscle-specific miRNAs, *mir-1* and *mir-133*, during myogenesis (Rao et al. 2006).

The founding member of the miRNA class of small RNAs is the product of

the *Caenorhabditis elegans lin-4* gene (Lee et al. 1993). Expression of *lin-4* is first detected in the middle of the first larval stage (L1) (Feinbaum and Ambros 1999), and its up-regulation results in the down-regulation of one of its target mRNAs, *lin-14* (Wightman et al. 1993). Down-regulation of the LIN-14 protein then allows the transition from the L1 to the L2 developmental stage (Ambros and Horvitz 1987).

Two lines of evidence suggest that the temporal regulation of *lin-4* occurs at the transcriptional level. First, Northern blotting analysis of the *lin-4* miRNA in wild-type animals reveals the presence of two transcripts, an ~65-nt and a 22-nt species. The longer transcript is a precursor of the mature 22-nt *lin-4* (Lee et al. 1993). Both RNAs are up-regulated coordinately during the mid-L1 stage (R. Lee and V. Ambros, unpubl.), suggesting that the *lin-4* precursor is activated transcriptionally during the L1 stage and then the mature *lin-4* is rapidly processed from its precursor. Second, *lin-4::gfp* transcriptional reporters containing only DNA sequences upstream of the miRNA recapitulate its temporal expression, indicating that these upstream sequences contain all the transcriptional regulatory elements required for the temporal regulation of *lin-4* (Esquela-Kerscher et al. 2005; Baugh and Sternberg 2006; this study).

In this study, we identify a class of Zn-finger FLYWCH transcription factors that includes FLH-1, FLH-2, and FLH-3 (*FLYWCH* transcription factor-1, *FLYWCH* transcription factor-2, and *FLYWCH* transcription factor-3) that act

redundantly during embryogenesis to repress the transcription of *lin-4* and other miRNAs that are normally expressed post-embryonically.

## Results

### *FLH-1 binds to an upstream region of lin-4*

To identify candidate proteins that could be direct regulators of *lin-4* expression, we conducted yeast one-hybrid (Y1H) screens using an 87-bp fragment from the phylogenetically conserved upstream region of the *lin-4* gene as bait (Lee et al. 1993). This DNA fragment (fragment 365–451) consists of nucleotides 365–451 (as measured 5' from the start of the mature *lin-4*) of a 693-bp Sall *lin-4* rescuing construct (Lee et al. 1993). As preys in the Y1H screens, we used a random-primed and an oligo dT-primed *C. elegans* cDNA library. We screened  $2.1 \times 10^6$  yeast transformants and found several candidates exhibiting fragment 365–451 binding activity. Among these candidates was a 485-bp sequence encoding a portion (residues 105–265) of an uncharacterized ORF, *y11d7a.12*, which encodes a predicted transcription factor with a FLYWCH Zn-finger DNA-binding domain (Dorn and Krauss 2003; Reece-Hoyes et al. 2005). Based on the presence of the FLYWCH motif in the Y11D7A.12 protein, the gene name *flh-1* was assigned to *y11d7a.12*.

### *The FLH-1-binding fragment in the lin-4 promoter is essential for repression of lin-4 in the embryo*

To determine whether sequences contained within fragment 365–451 are necessary for the proper temporal expression of *lin-4 in vivo*, we used a *Plin-4::gfp* reporter, consisting of 2.4 kb of DNA sequences upstream of the mature

*lin-4* fused to GFP. As expected, animals containing the *Plin-4::gfp* (*mals134*) transgene exhibited no GFP expression during embryogenesis (Fig. IV-1A). Consistent with previous reports, GFP expression from *mals134* is first detected at the mid-L1 stage and persists into adulthood in various cell types, including the hypodermis, vulva, pharynx, ventral nerve cord, and muscles (data not shown) (Esquela-Kerscher et al. 2005; Baugh and Sternberg 2006; this study). A *Plin-4::gfp* reporter construct containing a deletion of nucleotides corresponding to the sequence of fragment 365–451 exhibited GFP expression in late-stage embryos (Fig. IV-1A), suggesting that this fragment contains *cis*-acting regulatory sequences necessary for the repression of *lin-4* expression in the embryo, presumably through the binding of FLH-1.

#### *RNAi of flh genes results in precocious embryonic expression of lin-4*

To test whether FLH-1 is required for the repression of *lin-4* expression in the embryo, we assayed *lin-4* RNA levels in embryos produced by hermaphrodites treated with *flh-1*(RNAi). We used *rrf-3(pk1426)* animals that are hypersensitive to RNAi (Simmer et al. 2002). Northern blot analysis of total RNA extracted from *flh-1*(RNAi) embryos detected weak precocious expression of *lin-4* (Fig. IV-1B). This precocious expression was enhanced by simultaneous RNAi of both *flh-1* and *flh-2* (*c26e6.2*), which encodes another of the *C. elegans* FLYWCH family of proteins (Fig. IV-1B,C). RNAi of *flh-2* alone did not result in any detectable precocious *lin-4* (Fig. IV-1B). Similarly, RNAi of *flh-1* or *flh-2* alone

did not lead to appreciable precocious expression of GFP from our *Plin-4::gfp* reporter (data not shown). However, *Plin-4::gfp* was precociously active in double-RNAi, *flh-1(RNAi); flh-2(RNAi)* embryos (Fig. IV-1D). These results suggest a functional redundancy between FLH-1 and FLH-2 in the repression of *lin-4* expression during embryogenesis.

#### *Isolation and characterization of deletion mutations in the flh genes*

To further test the role of the FLYWCH family of proteins in the regulation of miRNA gene expression, we obtained deletion mutations of *flh-1* and *flh-2* by screening a library of ethyl methanesulphonate (EMS)-mutagenized worms using gene-specific PCR primers. The *y11d7a.12* deletion mutation, *flh-1(bc374)*, is an 894-bp deletion that deletes most of exons 2 and 3 and results in the loss of the FLYWCH domain (Fig. IV-2A). The *c26e6.2* mutation, *flh-2(bc375)*, is a 2023-bp deletion that extends from the predicted translation start site to all of exon 4 (Fig. IV-2A), and it also contains an insertion/duplication of *ttttatcagaccgcctgt* at the deletion junction.

Animals homozygous for either *flh-1(bc374)* or *flh-2(bc375)* exhibited a nearly wild-type phenotype with a low penetrance of young larvae with morphological abnormalities (2.8% for *bc374*, *n* = 502) (Fig. IV-2B; Fig. IV-3A). While the single FLYWCH mutants had an almost wild-type phenotype, loss of both *flh-1* and *flh-2* resulted in the complete penetrance of early larval lethality (Fig. IV-2B; Fig. IV-3A). L1 larvae homozygous for *flh-1(bc374)* and *flh-2(bc375)*

had either severe morphological abnormalities and/or appeared necrotic, and 100% ( $n = 92$ ) of these double-mutant L1 larvae died before reaching the L2 larval stage.

Unlike the *flh-1(bc374); flh-2(bc375)* double mutant, animals double mutant for two other alleles of *flh-1* and *flh-2*—*flh-1(tm2118); flh-2(tm2126)*—were viable (Fig. IV-3A). The viability of the *flh-1(tm2118); flh-2(tm2126)* double mutant is consistent with the less severe molecular lesions in these alleles as compared with *flh-1(bc374)* and *flh-2(bc375)* and indicates that *flh-1(tm2118)* and *flh-1(tm2126)* express residual protein and are probably not nulls (Fig. IV-2A; data not shown). Although *flh-1(tm2118); flh-2(tm2126)* larvae appeared superficially normal, adults are uncoordinated and retain more embryos than N2 adults, suggesting defects in egg-laying. Interestingly, animals carrying the likely null *flh-1(bc374)* allele and the less severe *flh-2(tm2126)* allele exhibit larva lethality (Fig. IV-3A and data not shown).

Animals mutants for a third FLYWCH motif-containing protein (*y11d7a.13*), *flh-3(tm3024)* (Fig. IV-1C), exhibited wild-type phenotype. Animals containing mutations in *flh-2* and *flh-3* —*flh2(bc375);flh-3(tm3024)* — appeared superficially normal, although adults retained more embryos than N2 adults (data not shown). Since *flh-1* and *flh-3* are neighbor genes located approximately only 5kb apart on chromosome IV, we were unable to obtain double *flh-1;flh-3* mutant animals. To test whether knock down of *flh-1* and *flh-3* genes confer a phenotype similar to what is seen in *flh1(bc374);flh-2* mutants, we performed double *flh-1;flh-3* RNAi



and scored for larvae with morphological abnormalities and/or that appeared necrotic. We observed a high percentage of abnormal larvae (46%,  $n = 520$ ), comparable to double *flh-1;flh-2* and *flh-2;flh-3* (31%,  $n = 414$ ; 37%,  $n = 359$  respectively) RNAi (Fig. IV-3B). Altogether, these analyses suggest that single *flh* mutants display almost wild-type phenotypes while any double mutant display high penetrance of severe phenotypes, including early larval lethality, and reveal also a functional redundancy among FLH proteins.

#### *Expression pattern of FLH transcription factors*

To visualize the expression pattern of FLH-1, FLH-2, and FLH-3, we made fluorescent translational or transcriptional fusions and examined their expression in transgenic worms. Expression of VENUS (Nagai et al. 2002) from the translational fusion *flh-1::venus* localizes to most cells starting at the gastrulation stage, with its expression diminishing by the L1 stage (Fig. IV-4A). The expression pattern of the translational *gfp::flh-2* transgene (Fig. IV-4A) was somewhat different from that of *flh-1::venus*. In embryogenesis, GFP was detected starting at the gastrulation stage. However, expression in head and tail cells persisted during the larval and adult stages (Fig. IV-4A). Fluorescence from the transcriptional fusion *Pflh-3::gfp* was detected in late-stage embryos and L1 larvae (Fig. IV-4A)

Northern blot analysis of total RNA extracted from populations of staged animals shows that the *flh-1*, *flh-2*, and *flh-3* mRNAs are detected in embryos

and reduced significantly after hatching (Fig. IV-4B; data not shown). While we were unable to observe VENUS from FLH-1::VENUS, we detected the *flh-1* mRNA in L4 animals and adults (Fig. IV-4B), suggesting that additional regulatory elements may be involved in the regulation of *flh-1* expression that are not present in our *flh-1::venus* transgene.

Western blot analyses using polyclonal antibodies against FLH-1 show that FLH-1 is present during embryogenesis and adulthood, while FLH-2 could be detected only in embryos (Fig. IV-4C). Protein expression data, together with Northern blot analysis showing reduction of *flh-1* mRNA after embryogenesis, and with the temporal expression of the fluorescent transgenes, indicate that FLH-1 and FLH-2 function during embryogenesis to repress *lin-4* expression (as well as other target genes, see below) and then are down-regulated soon after hatching, consistent with the post-embryonic up-regulation of *lin-4*.

#### *Precocious expression of lin-4 in flh mutants reduces LIN-14 levels in embryos*

To determine if the elevated expression of *lin-4* in embryos of *flh* mutants could lead to precocious down-regulation of LIN-14, the principal target of *lin-4*, we used Western blots to estimate LIN-14 levels in single and double *flh-1* and *flh-2* mutant embryos (Fig. IV-5A). Single-mutant *flh-1(bc374)* embryos or double-mutant *flh-1(tm2118); flh-2(tm2126)* embryos exhibited an eight- to 10-fold decrease in LIN-14 levels compared with N2, respectively. *flh-2(bc375)* embryos showed no change in LIN-14 levels compared to N2 (Fig. IV-5A).

Interestingly, although LIN-14 levels were significantly lower during embryogenesis in the single *flh-1* and the double *flh-1; flh-2* mutants, the level of LIN-14 in starved L1 larvae remained relatively unchanged compared with N2 (Fig. IV-5A). In L1 larvae, only a slight increase (two-fold) in LIN-14 was detected in the single *flh-1(bc374)* and *flh-2(bc375)* mutants and in the double *flh-1(tm2118); flh-2(tm2126)* mutant (Fig. IV-5A). These results indicate that the precociously expressed *lin-4* in embryos of the *flh-1(bc374)* and *flh-1(tm2118); flh-2(tm2126)* can lead to embryonic repression of LIN-14, but does not significantly affect LIN-14 levels post-embryonically.

*Elevated lin-4 levels in flh mutant embryos do not result in post-embryonic heterochronic defects*

It has been shown previously that elevated *lin-4* results in a precocious expression of L2-adult developmental events (Feinbaum and Ambros 1999). We examined whether up-regulation of *lin-4* in *flh-1* and *flh-2* mutants causes defects in post-embryonic developmental timing. In particular, we used a transgenic *col-19::gfp* reporter to monitor the timing of adult-specific developmental programs in the lateral hypodermis. The *col-19* gene is an adult-specific collagen gene that is under the control of the heterochronic pathway (Liu et al. 1995). Heterochronic mutations that cause precocious development result in the premature expression of *col-19::gfp* during larval stages, whereas mutations that cause retarded development result in the loss of *col-19::gfp* expression in adults (Abrahante et

al. 1998). We did not observe altered timing of *col-19::gfp* expression in *flh-1(bc374)* or *flh-2(bc375)* single mutant as well as in *flh-1(tm2118); flh-2(tm2126)* or *flh-2(bc375); flh-3(tm3024)* double mutants (Table IV-1). We were unable to examine *flh-1(bc374); flh-2(bc375)* double-mutant animals for post-embryonic heterochronic phenotypes because of the early larval lethality of the double-mutant combination.

Animals with a high-copy transgene of the *lin-4* gene display precocious phenotypes, including egg-laying defects, dumpy phenotype, and defects in tail and vulva morphology, that are reminiscent of *lin-14* loss-of-function (*lf*) mutations (Feinbaum and Ambros 1999). Unlike worms that overexpress *lin-4* from a transgene, single *flh-1(bc374)* animals, other than having a low penetrance of abnormally shaped larvae, do not show phenotypes similar to those of *lin-14(lf)* animals. *flh-2(bc375)* mutants, however, do exhibit a low penetrance of dumpy animals (Fig. IV-2B; Fig. IV-3A). In addition, the double *flh-1(tm2118); flh-2(tm2126)* mutants are uncoordinated and have egg-laying defects and an incomplete penetrance of dumpy animals.

An additional phenotype of animals overexpressing *lin-4* is the precocious expression of L2 larvae-specific cell division at the L1 larval stage (Feinbaum and Ambros 1999). To determine whether single or double mutants of *flh-1* and *flh-2* exhibit precocious L2-stage cell division, we examined the hypodermal seam cell of the V lineage (V1–V6). Wild-type L1 larvae hatch with six V-lineage seam cells. These six seam cells divide once in L1 to generate a daughter seam cell

and a hypodermal cell. Because the hypodermal daughter cell does not divide, the number of V-lineage seam cells at the end of the L1 stage remains at six. At the L2 stage, five of the six seam cells generate two daughter seam cells, thus increasing the number of the V-lineage seam cells from six to 11 (Sulston and Horvitz 1977). In L1 larvae deficient for LIN-14, the seams cells undergo aberrant division programs resulting in the production of more than six seam cells (Ambros and Horvitz 1984). We used a seam cell nuclei-specific fluorescent marker (*scm::gfp*) to examine whether the precocious expression of *lin-4* in the *flh* mutants results in an abnormally high number of V-lineage seam cells in the L1 stage. As in wild-type L1 larvae, the number of seam cells remained at six in the single *flh-1(bc374)* and *flh-2(bc375)* mutants as well as in the double *flh-1(tm2118); flh-2(tm2126)* mutant (Table IV-2; Fig. IV-5B), indicating that the precocious expression of *lin-4* during the embryonic stage is not sufficient to confer post-embryonic heterochronic defects.

*In addition to lin-4, FLH proteins can regulate the levels of other miRNAs*

To determine whether the levels of other miRNAs besides *lin-4* changed in *flh* mutants, we performed real-time RT-PCR (miRTaqMan) assays on 107 miRNAs using total RNA isolated from embryos from various *flh* mutants. For *flh-1(bc374)*, we found that in addition to *lin-4*, *mir-48*, *mir-241* and *mir-59* were also significantly increased at least twofold in mutant compared with wild-type embryos. Significant decreases of less than twofold were detected for *mir-51* and

*mir-60* (Pval<0.005; Fig. IV-6A; Table IV-3). In *flh-2(bc375)*, there was a significant decrease of at least twofold in the levels of *mir-38* and *mir-51* but no significant increase in the levels of any of the miRNAs was detected (Fig. IV-6B; Table IV-3). miRTaqMan analysis of animals mutant for *flh-3* showed a significant twofold increase in the levels of five miRNAs, including *mir-34* and *mir-49*, and a significant twofold decrease of six miRNAs (Fig. IV-6D; Table IV-3).

The most marked changes were observed in double *flh* mutants. Analysis of *flh-1(bc374);flh-2(bc375)* embryos showed significantly increased levels in six miRNAs. Among these miRNAs, *lin-4*, *mir-48*, *mir-241* showed an even higher overexpression level than in *flh-1(bc374)* embryos. Five other miRNAs were significantly decreased at least twofold (Fig. IV-6C; Table IV-3). Embryos double mutant for *flh-1(tm2118)* and *flh-2(tm2126)* exhibit increased as well as decreased levels for 14 and 16 miRNAs, respectively (Fig. IV-6E; Table IV-3). Interestingly, *flh-1(tm2118);flh-2(tm2126)* embryos displayed similar increased levels of *lin-4*, *mir-48* and *mir-241* compared to *flh-1(bc374);flh-2(bc375)* embryos. Moreover, among the set of miRNAs that showed changes in expression in *flh-1* and *flh-2* mutants compared to wild-type embryos, *lin-4*, *mir-48* and *mir-241* are the only three miRNAs that change in both, *flh-1(bc374);flh-2(bc375)* and *flh-1(tm2118);flh-2(tm2126)* mutant embryos (Table IV-3). Animals with a combination of the *flh-2(bc375)* and *flh-3(tm3024)* mutations exhibited elevated levels of several miRNAs, including *lin-4*, *mir-48*, and *mir-241* (Fig. IV-6F; Table IV-3). Double *flh-2(bc375)* and *flh-3(tm3024)* embryos also displayed

decreased levels of several miRNAs among which, *mir-244* and *mir-51* were also decreased in single-gene *flh* mutants, although at lower levels (Table IV-3).

Altogether, our TaqMan results indicate that FLH proteins can directly or indirectly regulate the levels of several miRNAs. While single-gene *flh* mutations have only moderate or undetectable effects on miRNA levels, double *flh* gene mutations do have significant effects on miRNA levels, especially for *lin-4*, *mir-48* and *mir-241*, further confirming a redundant relationship between FLH-1, FLH-2, and also FLH-3.

#### *Genome-scale Y1H screens reveal additional interactions between miRNA promoters and FLH proteins*

As part of a genome-scale Y1H screen to identify the TF that can interact with miRNA promoters we tested whether other miRNA promoters besides *lin-4* could be directly bound by FLH-1 and FLH-2. We detected binding of FLH-1 to the promoters of *lin-4*, *mir-241*, *mir-48*, *mir-53*, *mir-59*, and *mir-358-357*, and binding of FLH-2 to the promoters of *lin-4*, *mir-241*, and *mir-48* (Figure IV7.A;) (Martinez et al. 2008a). Northern blot analysis showed that *lin-4*, *mir-48*, *mir-59*, and *mir-241* are precociously expressed in *flh-1(bc374)* but not in *flh-2(bc375)* or N2 embryos (Fig. IV-7B). No change in expression levels was detected for *mir-53* or *mir-358* in either *flh-1(bc374)* or *flh-2(bc375)* (Fig. IV-7B). We have generated transgenic worms expressing GFP under the control of specific miRNA promoters as part of a genome-scale analysis of miRNA promoter activity

(Martinez et al. 2008b). We crossed *Plin-4::gfp*, *Pmir-59::gfp*, *Pmir-48::gfp* and *Pmir-241::gfp* transgenics to *flh-1(bc374)* mutants. While GFP expression was absent in the wild-type background, precocious embryonic GFP expression was observed in *Plin-4::gfp*, *Pmir-59::gfp*, and *Pmir-241::gfp* in the *flh-1(bc374)* mutant background (Fig. IV-7C). Northern blot analysis of RNA from various developmental stages of *flh-1(bc374)* and *flh-2(bc375)* revealed little or no change in the temporal expression of their miRNA targets during post-embryogenesis (data not shown), indicating a primary role of FLH-1 and FLH-2 in embryogenesis. Altogether, we identified six FLH targets by Y1H assays. Using a combination of Northern blotting, GFP reporters and TaqMan assays we showed that FLH proteins function to repress four of these targets (*lin-4*, *mir-48*, *mir-241* and *mir-59*) at the embryonic stage (Table IV-4).

#### *Identification of an FLH-1 consensus binding site*

We used the sequences of the miRNA promoters that were found to bind FLH-1 in Y1H assays (*Plin-4*, *Pmir-241*, *Pmir-48*, *Pmir-53*, *Pmir-59*, and *Pmir-358-357*) to derive a putative FLH-1 consensus binding site using the Improbizer algorithm (Ao et al. 2004). We found that the predicted FLH-1-binding site contains an a/gGGCGCCG sequence that tends to be located in the first 1 kb upstream of the annotated miRNA (Fig. IV-8A). miRNA promoters that bind FLH-1 by Y1H assays (“Y1H positives”) had higher Improbizer scores than a set of promoters that did not bind FLH-1 (“Y1H negatives”) (Fig. IV-9). Of the six



promoters found to be positive for Y1H interaction with FLH-1, four (*lin-4*, *mir-241*, *mir-48*, and *mir-59*) contain the a/gGGCGCCG, and these four correspond to miRNAs that changed in expression in *flh* mutants as assessed by Northern blots, GFP reporters and TaqMan assays. Only one Y1H negative promoter had an Improbizer score comparable with the Y1H positives. Interestingly, this miRNA (*mir-34*) was upregulated in *flh-3(tm3024)*, *flh-1(bc374)*; *flh-2(bc375)* and *flh-2(bc375);flh-3(tm3024)* mutants, suggesting that we failed to detect FLH binding to this promoter by Y1H assays. Alternatively, FLH proteins might indirectly regulate this miRNA despite the presence of an FLH-1 consensus binding site.

We found that DNA baits containing one or three copies of the consensus binding site sequence can, indeed, interact with FLH-1 in Y1H assays (Fig. IV-8B). Moreover, deletion of the a/gGGCGCCG sequence from the *mir-48* promoter abolishes FLH-1 binding while it does not affect binding of other *Pmir-48* interactors (Fig. IV-8C) (Martinez et al. 2008a).

To further verify the interaction between FLH-1 and the consensus sequence, we performed gel mobility shift assays using total protein extract from N2 embryos with the 87-bp fragment 365–451 from *Plin-4* (Fig. IV-8D) and a 51-bp fragment (fragment 200–251; consisting of nucleotides 200–251 upstream of the mature *mir-48*) from *Pmir-48* that also contains a consensus FLH-1-binding sequence (Fig. IV-8E). Addition of radiolabeled fragment 365–451 to the embryo extract resulted in a shifted complex that was competed away by unlabeled fragment 365–451 (Fig. IV-8D, lanes 3–6). A supershifted complex was formed

upon the addition of anti-FLH-1 serum but not with a control antibody (Fig. IV-8D, lanes 8,9). Likewise, incubation of the embryo extract with a radiolabeled fragment from *Pmir-48* (fragment 200–251) resulted in the formation of a shifted complex (Fig. IV-8E, lane 2). A supershifted complex was detected with the addition of anti-FLH-1 serum (Fig. IV-8E, lane 6) but not with a control antibody (Fig. IV-8E, lane 8). The shifted complex can be outcompeted upon the addition of unlabeled cold fragment 200–251 (Fig. IV-8E, lane 3). Addition of 100-fold excess of unlabeled fragment 200–251 deleted for the consensus site was ineffective in preventing FLH-1 binding to a wild-type radiolabeled fragment 200–251 (Fig. IV-8E, lane 4), consistent with the sequence-specific binding of FLH-1 to the *Pmir-48* 200–251 fragment via the consensus sequence. Taken together, our results suggest that FLH-1 binds directly to its miRNA targets through a a/gGGCGCCG consensus sequence.

*FLH-2 can also interact with the FLH-1 consensus binding site*

Since FLH proteins regulate common set of genes, we explored the possibility that other FLYWCH-containing proteins besides FLH-1, can interact with the FLH-1 consensus binding site. We found that in Y1H assays, DNA baits containing one or three copies of the FLH-1 consensus binding site sequence can also interact with FLH-2 but not FLH-3 or PEB-1, a fourth FLYWCH-containing protein (Fig. IV-10A). We derived a consensus binding site for FLH-2 using the sequences of the miRNA and protein-coding gene promoters that

were found to bind FLH-2 in Y1H assays (*Plin-4*, *Pmir-241*, *Pmir-48*, *Punc-30*, *Pceh-60*, *PF59H6.6*, and *Phlh-30*) (Martinez et al. 2008a; Vermeirssen et al. 2007a; Arda et al. in preparation). Consistently, the FLH-2 consensus binding site is very similar to the FLH-1 binding site, containing a GGCGCCG motif (Fig. IV-10B). In agreement with our observations, a PEB-1 DNA binding site has been previously identified and does not resemble the FLH-1 consensus site (Thatcher et al. 2001). These results suggest that FLH-1 and FLH-2 are able to recognize the same sequence albeit in Y1H assays. We were unable to detect any interaction between FLH-3 and miRNA as well as protein-coding gene promoters tested in Y1H assays (Deplancke et al. 2007; Vermeirssen et al. 2007a; Martinez et al. 2008a; Arda et al. in preparation).

#### *FLH-1 represses the expression of flh-2*

TFs involved in similar processes often engage in cross-factor control, where pairs of factors each bind the promoter of the reciprocal factor (Borneman et al. 2006). To investigate the possibility of cross-factor control between *flh* genes, we used the promoters of *flh-1* and *flh-2* as baits in Y1H assays. Specifically, *flh-1* and *flh-2* promoters (*Pflh-1* and *Pflh-2*) were defined as the intergenic sequences upstream of the ATG with 2 kb in length. While we did not find any interactions for *Pflh-1*, *Pflh-2* interacted with FLH-1 and FLH-2 proteins (Fig. IV-11A). Consistent with this finding, *Pflh-2* has an FLH-1 consensus binding site 1.5 kb upstream of the ATG (Fig. IV-11B). To investigate the

consequence of the interaction between FLH-1 and FLH-2 and *Pflh-2 in vivo*, we quantified the levels of FLH-1 and FLH-2 protein in wild-type as well as *flh-1(bc374)* and *flh-2(bc375)* mutant embryos. While there is little change in the levels of FLH-1 protein in *flh-2(bc375)* mutants (1.4 fold), FLH-2 levels are increased by 7.8 fold in *flh-1(bc374)* compared to wild-type worms (Fig. IV-11C). In sum, these findings suggest that FLH-1 represses the expression of *flh-2* in embryos and this repression is likely direct.

#### *FLYWCH proteins engage in protein-protein interaction*

We performed Yeast two-hybrid (Y2H) assays to investigate the possibility of protein-protein interactions among members of the FLYWCH family (Walhout et al. 2002). We tested binary combinations between each AD-FLH fusion with each DB-FLH fusion (Fig. IV-12A). We found that an AD-FLH-2 fusion can interact with DB-FLH-1, DB-FLH-3 as well as DB-FLH-2 fusions as measured by activation of the HIS3 and  $\beta$ -Galactosidase reporter. Although we only detected activation of the HIS3 reporter gene, AD-FLH-1 was able to interact with DB-FLH-3, which is in agreement with previous reports (Walhout et al. 2002). These results suggest that each FLYWCH protein is capable to engage in protein-protein interactions with the other two FLYWCH proteins and that FLH-2 can also engage in protein-protein interactions with itself (Fig. IV-12B).

*FLH transcription factors likely regulate multiple targets*

To determine whether early larval lethality of the double *flh-1(bc374); flh-2(bc375)* mutation was caused primarily by overexpression of *lin-4*, *mir-241*, or *mir-48* during embryogenesis, we asked whether the lethality of the *flh-1; flh-2* double mutant could be rescued in animals deleted for *lin-4* or *mir-48;mir-241*. We crossed *flh-1(bc374); flh-2(bc375)* to *lin-4(e912)* and to *nDf51 [mir-48(0) mir-241(0)]* mutants. Animals that were homozygous for either *lin-4(0)* or *nDf51 [mir-48(0) mir-241(0)]* did not produce viable progeny that were homozygous for both *flh-1(bc374)* and *flh-2(bc375)* (data not shown). This implies that the inviability of *flh-1(bc374); flh-2(bc375)* animals is not primarily a consequence of excessive levels of *lin-4* or of *mir-48* and *mir-241*, but may be due to the collective up-regulation of miRNA targets of FLH-1 and FLH-2, and/or altered expression of yet to be identified miRNA or protein-coding target genes of FLH-1 and FLH-2.

## Discussion

We report the identification of three previously uncharacterized *C. elegans* transcription factors—FLH-1, FLH-2, and FLH-3—that act redundantly to repress miRNA expression. GFP reporters, mRNA and protein analyses suggest that FLH protein function primarily in the embryo.

By a combination of different assays, including Y1H, GFP reporters, RNAi and TaqMan PCR, we found that FLH-1, FLH-2, and FLH-3, act redundantly to repress the expression of *lin-4*, *mir-48* and *mir-241*, miRNAs that are normally expressed in later stages of development. Y1H assays identified three FLH-1 additional targets (*mir-59*, *mir-357-358* and *mir-59*). One of them, *mir-59*, was also verified by GFP reporters and TaqMan assays. However, we did not detect an increased de-repression in double *flh* mutants, suggesting that *mir-59* might be mainly regulated by FLH-1. Hence, it is feasible that FLH proteins regulate common as well as different set of target genes (Fig. IV-13A). TaqMan PCR assays of single and double *flh* mutants also identified additional miRNAs, besides *lin-4*, *mir-48* and *mir-241*, whose levels changed compared to wild-type embryos (Table IV-3). In the case of *flh-1* and *flh-2* miRNA targets, we searched the promoters of these miRNAs for the presence of an FLH-1/FLH-2 consensus binding site. Except for *Pmir-34*, none of the promoters contain a consensus binding site suggesting that the levels of these miRNAs changed as a consequence of indirect rather than direct FLH-1 regulation (Fig. IV-13A).

FLH-1, FLH-2, and FLH-3 are three of four proteins in *C. elegans* that

contain a FLYWCH DNA-binding domain (Reece-Hoyes et al. 2005). Proteins with FLYWCH domains have also been identified in insects and vertebrates (Dorn and Krauss 2003; Krauss and Dorn 2004; Babu et al. 2006). The FLYWCH motif is a Cys<sub>2</sub>His<sub>2</sub>-type zinc-finger domain characterized by the conserved sequence: F/Y-X(n)-L-X(n)-F/Y-X(n)-WXCX(6–12)CX(17–22)HXH (where X is any amino acid). It was first identified in isoforms of the putative chromatin-modulating protein *modifier of (mdg4)* locus [*mod(mdg4)*] in *Drosophila melanogaster* (Gerasimova et al. 1995; Buchner et al. 2000). Phenotypes of *mod(mdg-4)* mutant flies include effects on position effect variegation, chromatin insulation, nerve cell pathfinding, chromosomal meiotic pairing, and apoptosis (for review, see Dorn and Krauss 2003). It is not known whether *mod(mdg4)* mutations cause defects in miRNA expression in *Drosophila*. *Drosophila* Mod(mdg4) has not been shown to bind to DNA, but the Mod(mdg4) FLYWCH domain does seem to mediate protein–protein interactions, since it can directly interact with the DNA-binding protein Suppressor of Hairy-wing [Su(Hw)] as part of the gypsy insulator (Ghosh et al. 2001).

Using Y1H assays and computational analyses, we show that FLH-1 and FLH-2 can recognize the same consensus binding site. Interestingly, we were unable to detect FLH-3 binding to DNA in Y1H assays even though we have used to date more than 300 miRNA and protein-coding gene promoters as DNA baits (Deplancke et al. 2006a; Vermeirssen et al. 2007a; Martinez et al. 2008a; Arda et al. in preparation). It is likely that FLH-3 does not bind DNA, in a similar

way as the *Drosophila* Mod(mdg4) protein. Although we did not establish a direct *in vivo* interaction between the FLH proteins, Y2H assays detected a physical interaction between FLH-1, FLH-2 and FLH-3. Future studies will determine if the *C. elegans* FLYWCH motif, in analogy to the *Drosophila* FLYWCH motif of Mod(mdg4), is responsible for these protein-protein interactions.

*C. elegans* possesses a fourth protein with a FLYWCH domain, PEB-1 (Thatcher et al. 2001). However, unlike *flh-1*, *flh-2* and *flh-3*, miRNA TaqMan analyses of *peb-1* mutant embryos did not show any aberrance in *lin-4* levels (data not shown). The function and DNA-binding activity of *peb-1* requires the FLYWCH motif as a deletion in this region results in deficiencies in pharynx development and molting (Beaster-Jones and Okkema 2004; Fernandez et al. 2004). DNase-footprinting experiments identified a TGCCGT sequence as the PEB-1 binding site (Thatcher et al. 2001). Consistently, Y1H analysis showed that PEB-1 does not interact with the FLH-1 consensus binding site, although it is able to bind to the promoter of several miRNAs as well as protein-coding genes (Martinez et al. 2008a and Arda et al. in preparation). Future experiments will be required to determine if there is any functional interaction between PEB-1 and the other FLH proteins.

We found that the loss of both FLH-1 and FLH-2 results in early larval lethality. It is apparent from our studies that this phenotype is not due simply to the overexpression of *lin-4*, *mir-48*, or *mir-241* during embryogenesis since a *lin-4(lf)* mutation or deletion of *mir-48* and *mir-241* did not rescue the *flh* lethal



phenotype. Also, the increases in *lin-4*, *mir-241*, and *mir-48* levels observed in the viable *flh* mutants [*flh-1(tm2118)*; *flh-2(tm2126)* or *flh-2(bc375)*; *flh-3(tm3024)*] were approximately similar to the increases in those miRNAs observed in the inviable mutant [*flh-1(bc374)*; *flh-2(bc375)*]. Lethality in the latter case could be attributable to a slight but concerted overexpression of multiple miRNAs. However, we do not favor that option since worms with a mutation in a gene whose product is involved in the general processing and function of all the miRNAs, *alg-1*, also fails to rescue the lethality of the double FLYWCH mutant (data not shown). It is likely, then, that the FLH proteins regulate the transcription of protein-coding genes, as well as miRNA genes, that collectively contribute to the *flh* lethal phenotype.

The functional redundancy between the *flh-1*, *flh-2*, and *flh-3* genes in the repression of miRNA targets during embryogenesis is consistent with a range of possible molecular relationships among the FLH-1, FLH-2, and FLH-3 proteins *in vivo*, including binding common sets of DNA sites, common protein partners, and/or functioning separately in redundant pathways. We did not observe post-embryonic developmental timing defects in *flh* mutants that overexpress in embryos miRNAs known to be developmental timing regulators. For example, although transgenic overexpression in early larval stages of *mir-48* or *lin-4* has been shown to cause precocious expression of later larval cell fates (Feinbaum and Ambros 1999; Li et al. 2005), the viable *flh* mutants that we examined did not display such larval developmental timing phenotypes. This is consistent with our

observation that *mir-48* and *lin-4* are overexpressed during embryogenesis but are normally expressed during development of *flh* mutant larvae (data not shown). It appears that the FLH proteins are particularly involved in inhibiting miRNA expression in embryos, and other regulatory mechanisms, including transcriptional activators, govern up-regulation of these same miRNAs in larvae (Fig. IV-13B).

## Materials and methods

### *Y1H assays*

Fragment 365–451 of *Plin-4* was cloned into pBM2389, upstream of a *GAL1* promoter that drives the expression of *HIS3* (Liu et al. 1993), generating plasmids pBM2389.AF (forward orientation) and pBM2389.AR (reverse orientation). The DNA fragment was also cloned into pSE640, upstream of a *CYC1* promoter driving the expression of *LacZ* (American Type Culture Collection), generating plasmids pSE640.AF (forward orientation) and pSE640.AR (reverse orientation). These reporter plasmids were integrated into *Saccharomyces cerevisiae* strain YM4271. The yeast strains harboring the integrated plasmids were then transformed with two mixed-staged *C. elegans* cDNA libraries, pACT-RB1 (oligo dT primed) and pACT-RB2 (random hexamer primed; gifts from Dr. Robert Barstead). Yeast transformants with pBM2389.AF or pBM2389.AR and an activator plasmid were selected for tryptophan, leucine, and histidine prototrophy and with 40 mM 3-amino-1,2,4-triazole. Plasmids from positive candidates from the histidine screen were extracted and transformed into YM4271 harboring pSE640.AF or pSE640.AR, where the interaction was confirmed using in situ  $\beta$ -Galactosidase assays. All yeast manipulations were done following standard procedures (Adams et al. 1997).

Gateway-compatible Y1H assays were done as described before (Deplancke et al. 2006; Vermeirssen et al. 2007) using 2 kb upstream of the mature miRNA or *flh-1* and *flh-2* genes as baits and FLH-1 and FLH-2 as AD

fusion preys (Martinez et al. 2008a).

### *Plasmid construction*

A 5.6-kb fragment containing a rescuing wild-type copy of the *unc-119* gene was inserted into the promoter-less GFP vector pPD95.75 to generate pMO23. All PCR reactions were done using Platinum *Pfx* DNA polymerase (Invitrogen). The *Plin-4::gfp* reporter plasmid was made by amplifying a 2.4-kb fragment immediately upstream of the mature *lin-4* and inserting it into pMO23. Plasmids for *Pmir-241::gfp*, *Pmir-48::gfp*, *Pmir-59::gfp*, and *Pmir-358-357::gfp* were made by amplifying 1.5–2-kb fragments upstream of the mature miRNA and ligating them into pMO23. The *Plin-4::gfp* plasmid with a deletion of fragment 365–451 was made by the overlap extension PCR method (Ho et al. 1989).

A VENUS translational reporter for *flh-1* (*Pflh-1* [5 kb] *::flh-1::venus::flh-1* 3' UTR [290 bp]) was constructed using a combination of overlap extension PCR and Gateway cloning (Walhout et al. 2000). A GFP translational reporter for *flh-2* (*Pflh-2* [4.8 kb] *::gfp::flh-2::flh-2* 3' UTR [411 bp]) was made using a pBluescript SK(+) vector with GFP and the *unc-119* mini-gene transformation marker from pDP#MM051 (Maduro and Pilgrim 1995). The transcriptional fusion reporter for *flh-3* (*Pflh-3* [4.5 kb] *::gfp::flh-3* 3' UTR [1 kb]) was made by Gateway cloning. Details of all plasmid constructions and primer sequences will be provided upon request.

### *C. elegans strains*

Worms were grown using standard procedures at 20°C on Nematode Growth Medium (NGM) plates (Sulston and Hodgkin 1988). The wild-type strain was *C. elegans* var. Bristol strain N2 (Brenner 1974). Deletion alleles isolated from mutagenesis libraries were backcrossed to N2 at least six to eight times before characterization. All nematode strains used in this study are listed in Table IV-5.

### *C. elegans transformation*

Gold microparticle biolistic bombardment (for review, see Praitis 2006) of DP38 [*unc-119(ed3)*] was used to create transgenic worms carrying fluorescent reporters. We used a transformation procedure described by Berezikov et al. (2004) using a PDS-1000/He system with the Hepta adaptor (Bio-Rad). At least two independent lines were obtained per bombardment.

### *RNAi-by-feeding*

Embryos obtained following hypochlorite treatment of gravid adults were placed on RNAi plates (NGM with 100 µg/mL ampicillin, 15 µg/mL tetracycline, and 1 mM IPTG; seeded with bacteria expressing dsRNA) (Kamath et al. 2001). Once they reached L4, they were transferred onto fresh RNAi plates.

### *Isolation of FLH-1 and FLH-2 deletions*

Deletion alleles *flh-1(bc374)* and *flh-2(bc375)* were isolated from a population of worms mutagenized with EMS using the poison primer method (Edgley et al. 2002). DNA sequencing was performed to assess the nature of the lesions. Sequences of the screening primers will be made available upon request.

### *miRNA TaqMan real-time PCR assays*

One-hundred late-stage embryos were collected into Worm Lysis Buffer (50 mM KCl, 10 mM Tris at pH 8.3, 2.5 mM MgCl<sub>2</sub>, 0.45% NP-40, 0.45% Tween 20, 0.01% gelatin, 30 µg/mL proteinase K), subjected to 10 cycles of freezing and thawing, followed by incubation for 1 h at 65°C and for 20 min at 95°C. A Trizol (Invitrogen) extraction was done, and the RNA template was coprecipitated with glycogen (Ambion). The RNA was used in TaqMan assays following the instructions of the manufacturer using an ABI 7900HT Fast-Real Time PCR System (Applied Biosystems) (Chen et al. 2005). The miRNA C<sub>t</sub> values were analyzed in triplicate from three independent biological samples. The comparative C<sub>t</sub> ( $2^{-\Delta\Delta C_t}$ ) method (Livak and Schmittgen 2001) was used to calculate the average  $\Delta\Delta C_t$  values using the small nucleolar RNA, sn2841 or U18, as the normalization standard. Only those values for which the three independent biological replicates exhibited the same trend (increase or decrease) were considered in our analysis.  $\Delta\Delta C_t$  values were then normalized by

subtracting the average  $\Delta\Delta C_t$  value for all the miRNAs in the experiment. Normalized  $\Delta\Delta C_t$  values for each miRNA assay were averaged across all replicates to generate a  $\Delta\Delta C_t$  final value, and the standard error of the mean was determined. Z-scores were calculated as  $-\Delta\Delta C_t \text{ final} / \text{SD}$ . Z-scores  $\geq 2$  or  $\leq -2$  were considered significant.

#### *Northern blot analysis*

Total RNA was extracted and analyzed (5–20  $\mu\text{g}$ ) by Northern blotting as described by Ambros and Lee (2004) using Starfire probes (Integrated DNA Technologies) complementary to the miRNA or to U6 snRNA. Northern blots for *flh-1*, *flh-2*, and *flh-3* were done as described by Burnett (1997). PCR fragment probes for *flh-1* and *flh-2* were radiolabeled with  $[\alpha\text{-}^{32}\text{P}]\text{dATP}$  using the Decaprime II Random Primed DNA Labeling Kit (Ambion), and hybridized probe was detected using PhosphorImager screens and ImageQuant (Molecular Dynamics). Northern blots used to detect *flh-1* and *flh-2* were reprobed for *flh-3* using a PCR fragment specific for *flh-3* following the instructions of the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche).

#### *Western blot analysis*

Embryos were obtained from the hypochlorite treatment of staged gravid adults. Starved L1s were collected following the overnight hatching of embryos in M9 buffer at 20°C. Worm pellets were resuspended in an equal volume of Lysis

Buffer (4% SDS, 100 mM Tris at pH 6.8, 20% glycerol) and boiled for 20 min. Protein concentration was assessed using the RC DC Protein Assay Reagent (Bio-Rad). Protein extracts were resolved in 10% SDS-PAGE gels and transferred to PVDF membranes. Westerns were done with rabbit antisera against LIN-14 (Hristova et al. 2005),  $\gamma$ -tubulin (Sigma T1450), FLH-1, or FLH-2. Quantification of proteins was done using ImageJ (NIH).

#### *Preparation of anti-FLH-1 and anti-FLH-2*

A histidine-tagged full-length FLH-1 protein was expressed in *Escherichia coli* from the plasmid pHIS.Parallel1 (Sheffield et al. 1999), and purified protein was used to raise polyclonal antibodies in rabbits (Pocono Rabbit Farm and Laboratory). Polyclonal antibodies for FLH-2 were raised in rabbits immunized with a KLH conjugated peptide consisting of the last 20 amino acids (Open Biosystems).

#### *Prediction of an FLH-1 and FLH-2 consensus binding site*

The sequences of the promoters that tested positive for FLH-1 and FLH-2 binding by Y1H were analyzed using Improbizer (Ao et al. 2004) to predict a consensus binding site. In the case of FLH-1, only miRNA promoters were used. Since FLH-2 bound only 3 miRNA promoters, protein-coding gene promoters tested in other Y1H studies were used as well. Two types of input sequences were analyzed, either up to 2 kb or 1 kb upstream of the annotated miRNAs



(Martinez et al. 2008a). We used three types of background sequences: (1) all promoters sequences from the promoterome (Dupuy et al. 2004) ( $\approx$ 20,000 intragenic sequences that range in length from 300 bp to 2.5 kb from the transcription start site); (2) a subset of sequences from the promoterome containing only regions between 2 and 2.5 kb upstream of the translation start site; and (3) the same background as foreground (same set of positive sequences used as background model). The motif shown was the most redundant site among all six searches. Improbizer scores reflect how well a site present in a given promoter fits the position weight matrix (see <http://www.soe.ucsc.edu/~kent/improbizer/improbizer.html> for details). The sequence logo was created using WebLogo (<http://weblogo.berkeley.edu>).

#### *Cloning the FLH-1 consensus binding site*

We cloned the predicted FLH-1-binding site from the promoter of *mir-358*. The sequence tested contained four extra nucleotides on each side of the predicted hit to account for the possibility that flanking nucleotides important for binding may have been missed in the motif searches. Complementary DNA primers were designed to contain one (FLYWCH-1x and FLYWCH-1-y) or three (FLYWCH-3x and FLYWCH-3-y) tandem FLH-binding sites. A Gateway compatible entry vector, pMW#4, was ligated with the annealed primers, and the FLH-1-binding sites were subsequently cloned into pMW#2 and pMW#3 integrated into the genome of *S. cerevisiae* YM4271 and used in Y1H assays as

described previously (Deplancke et al. 2006).

#### *Deletion of the FLH-1-binding site in Pmir-48*

The FLH-1 consensus binding site was deleted from *Pmir-48* using the QuikChange II Site-Directed Mutagenesis kit (Stratagene) following the instructions of the manufacturer and using an entry clone with 2 kb of wt *Pmir-48* as template. The resulting *Pmir-48* deletion entry clone was analyzed by DNA sequencing and subsequently used to generate a Y1H bait as described previously (Deplancke et al. 2006).

#### *Electrophoretic mobility shift assay*

Wild-type gravid hermaphrodites were collected and embryos were harvested by hypochlorite treatment (Sulston and Hodgkin 1988). Embryos (50  $\mu$ L of packed pellet) were washed, resuspended in 650  $\mu$ L of 10 mM HEPES (pH 7.2) supplemented with Halt protease inhibitor cocktail (Pierce) and disrupted by 30 strokes of an ice-cold 3-mL stainless steel dounce homogenizer. Fifty micrograms of the resulting extract were incubated at room temperature in Binding Buffer (10 mM HEPES at pH 7.2, 25 mM NaCl, 1 mM MgCl<sub>2</sub>, 50  $\mu$ M ZnCl<sub>2</sub>, 5% glycerol) with 2  $\mu$ g of BSA and 0.5  $\mu$ g of poly (dI-dC). A 30-min preincubation was done for cold probe chases and supershift assays using anti-FLH-1 or rabbit IgG Ab-1 control antibody (Thermo Scientific). Following the addition of a [ $\gamma$ -<sup>32</sup>P]ATP 5'-end-labeled wild-type or mutant fragments 365–451

(from *Plin-4*) or 200–251 (from *Pmir-48*), samples were incubated for an additional 20 min, immediately loaded into a 5% native gel, and electrophoresed at room temperature.

#### *Y2H assays*

Gateway-compatible Y2H assays using AD and DB fusion proteins were done as described before (Walhout et al. 2000).

#### *Primer sequences*

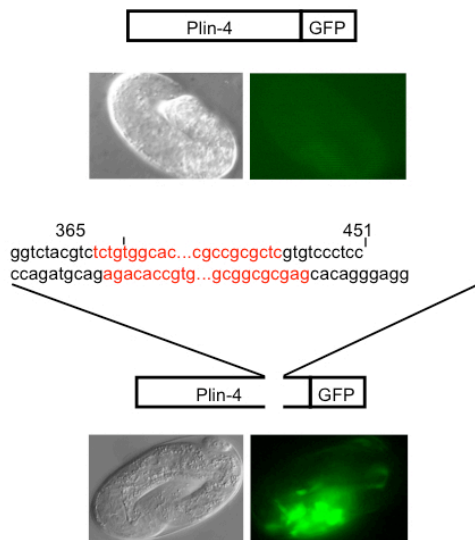
All primer sequences are available upon request

## Acknowledgments

We are grateful to C. Chen from Applied Biosystems for TaqMan assay reagents, S. Mitani from the National Bioresource Project for deletion strains, and C. Mello for providing the facility and supplies to do some of the radioactive experiments. Some nematode strains used in this work were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR). We thank B. Deplancke for help with the FLH-binding site cloning, A. Fire for pPD95.75, and Z. Derewenda for pHIS.Parallel1. M.C.O. was funded in part by NIH post-doctoral fellowship GM070118-02. This work was supported by NIH grants GM348642 and DK068429 to V.A. and A.J.M.W., respectively.

Figure IV-1

A



C

```

Ce-FLH-1 -----MYSPEMNMNIISSPFPSSSLNAPSLADAEVVRSDGGEATSEFPSTVAVPMEI 56
Cb-CBG15060 -----LTRPFLLF 8
Ce-FLH-2 -----MVFVPLPSTFRSNMELAMLA---GKLVAGLLGQPTIV 55
Cb-CBG18201 -----MSTFRQDVIYTTAFPIPLFPKASNNNDELAAVLQLEAGKGLGHRQDPFI 56
Ce-FLH-3 -----KMGATMEWAKVYVSKLNDGQALLQKMYADQAKLASEDEK 55
Cb-CBG15055 -----NDVASSSSAPFSTFKPTLDWATVYVQNTDQORALLQKMYAQEQERKK---EKEK 57

Ce-FLH-1 -----VSQASTAPVQLLAGIDMSVIRKNAWGLLAS-----A 93
Cb-CBG15060 -----RQK-----NOAKALLSG-----N 23
Ce-FLH-2 -----QAVNWS---TRGKHLADIVYVTVGLQK---TN---KAFSLLLDPGQ---TEGL 105
Ce-FLH-3 -----QAVLDG---TITTFVESPDTQIQIAQAKANAFPLGSAIIPFQDPSAISTVFPRLV 114
Cb-CBG15055 -----RMELELHKAQQAILDRLTQGIIFMVDMELNSETSP-----DSQ-----PSSSSSV 101
Ce-FLH-1 -----RQKELH---KQGGKALLDSITGVITVTPSTFP-----DQMGDSASSSSES 107

Ce-FLH-1 -----SAGQFQKNGLTHEKLLTALLANGSTAVAKRDAGSPRTIPLKGLIPLKSGHTMHL 153
Cb-CBG15060 -----TV---GNSIILAKLFGSLNAG---EKSESTSPGQGGKGLKGLIPLKSGHTMHL 153
Ce-FLH-2 -----SEDTRFASSPSPTAVSNSSGQSNATSSSSPEYKFNHVEHVAAGQVINGSEK 165
Cb-CBG18201 -----KEEPCFSSSSPSTTAASSS---SOTVMIFELMHEPEYKFNHVEHVAAGQVINGSEK 173
Ce-FLH-3 -----NSSTHEI---GITVYKLVKIKTKRAGKQ---SPWSSKPVYKFNHVEHVAAGQVINGSEK 164
Cb-CBG15055 -----VQSTEEHEIQLVKDVVLKGFKAQIAAARAKPKYKFNHVEHVAAGQVINGSEK 167

Ce-FLH-1 -----SCVPLKGLKGLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHL 213
Cb-CBG15060 -----SCMPLKGLKGLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHL 137
Ce-FLH-2 -----KCCGKGLKGLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHL 225
Cb-CBG18201 -----KCCGKGLKGLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHL 233
Ce-FLH-3 -----YDMLKGLKGLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHL 216
Cb-CBG15055 -----YDMLKGLKGLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHLIPLKSGHTMHL 224

Ce-FLH-1 -----RVYIKGNVADFNQRKIRKQVADREAAKRLVQOIEHQKQSSSIVAARAAYAGMQ 273
Cb-CBG15060 -----RVYIKGNVDTNLRKIRKQVADREAAKRLVQOIEHQKQSSSIVAARAAYAGMQ 197
Ce-FLH-2 -----YCHRFKSIIRSVWQRKLSKASVNVQVLQCLP-LPMEIDTR---QFOQT-----281
Cb-CBG18201 -----HILSEI---GFQ---OTKMLPSSLTFKRLLEDLTKTEKQKASTHMLL-----263
Ce-FLH-3 -----CVYK---QQNPLQGFQASSELLFPTMLFPRKPRRFRVATP---SPWILH-----273
Cb-CBG15055 -----

Ce-FLH-1 -----NGSIPTSSG---VASFLOQTSAT-----APMTHSMLLAQMFPLNNTKTEMNV 320
Cb-CBG15060 -----STGELSTFPMAAASLFGSAAJAAANAVTSSAFTTQTTIHFALLLAGLPTLW 257
Ce-FLH-2 ----------NMLMLASIN 289
Cb-CBG18201 ----------NMLMLFTI 290
Ce-FLH-3 ----------
Cb-CBG15055 -----

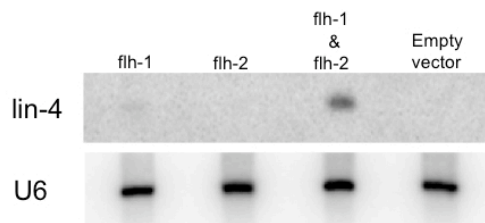
Ce-FLH-1 -----KNEQTFEYVQKMSMLPLQAT-----LAPLIFTENYDSEYVQPAIKKAKELTME 375
Cb-CBG15060 -----KN-ETTAVYVYHISPNMQSATPSTSTIPELLPKTEPLETSSQPSAKRAIDTFDD 316
Ce-FLH-2 -----AAATAAAATAGIS-----PVVQKRVSTVMPITLKRQTTREED 314
Cb-CBG18201 ----------ADAAASQSSP-----PTIQKLTSTCVIPKSFK---SEEDS 329
Ce-FLH-3 ----------SMTNCCGRKPTLRLPDLFKIKSEI 292
Cb-CBG15055 ----------RMVLQDQKPFSLRLPDPNPLFLKAEIS 302

Ce-FLH-1 -----EHLRSDPQOPTLARKRKLWGLNRYVITTTTT-HHFFKSKNDGTHLIVP 434
Cb-CBG15060 -----AEVTSHPVEPTFOAHSEKRLWRGTVRFTMGPSPAFEPFYAKYGAELHIVP 376
Ce-FLH-2 -----SFCASQKQETITTEKSMKMFVTEIGVWQDQD---ALLVSHDNGASVFF 387
Cb-CBG18201 -----AETITAREASVYDTEKMKMKFVTEIGVWQDQD---ALLVSHDNGASVFF 387
Ce-FLH-3 -----DMEMTLTKFLENKRTDLSIRKGLVFFFAHDEL-----VLVADQHNISVQK 347
Cb-CBG15055 -----RISGALTKYELIHREREKAKKHEIFVFFFAKREL-----RQVADYCRGAEHR 357

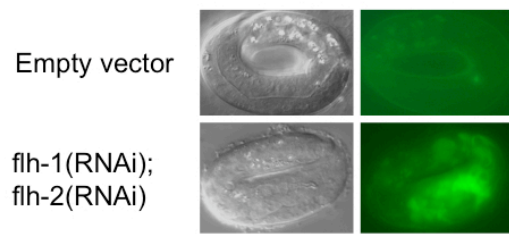
Ce-FLH-1 -----NRHLRDEAKRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 494
Cb-CBG15060 -----CFEIKQKSNRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 436
Ce-FLH-2 -----VWYKQKSNRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 452
Cb-CBG18201 -----VVVLSKTSKSNRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 447
Ce-FLH-3 -----IPIKSNRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 407
Cb-CBG15055 -----LWYKSNRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 417

Ce-FLH-1 -----SKKSNRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 507
Cb-CBG15060 -----SKKSNRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 462
Ce-FLH-2 -----ALRSTQDAPGPELNHV-----470
Cb-CBG18201 -----DYKSTQDAPGDIQQM-----465
Ce-FLH-3 -----NRKSNRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 420
Cb-CBG15055 -----NRKSNRPEKQCCQCLGCLGCFKESWNSPMLNSDNQSTLELID 430
    
```

B



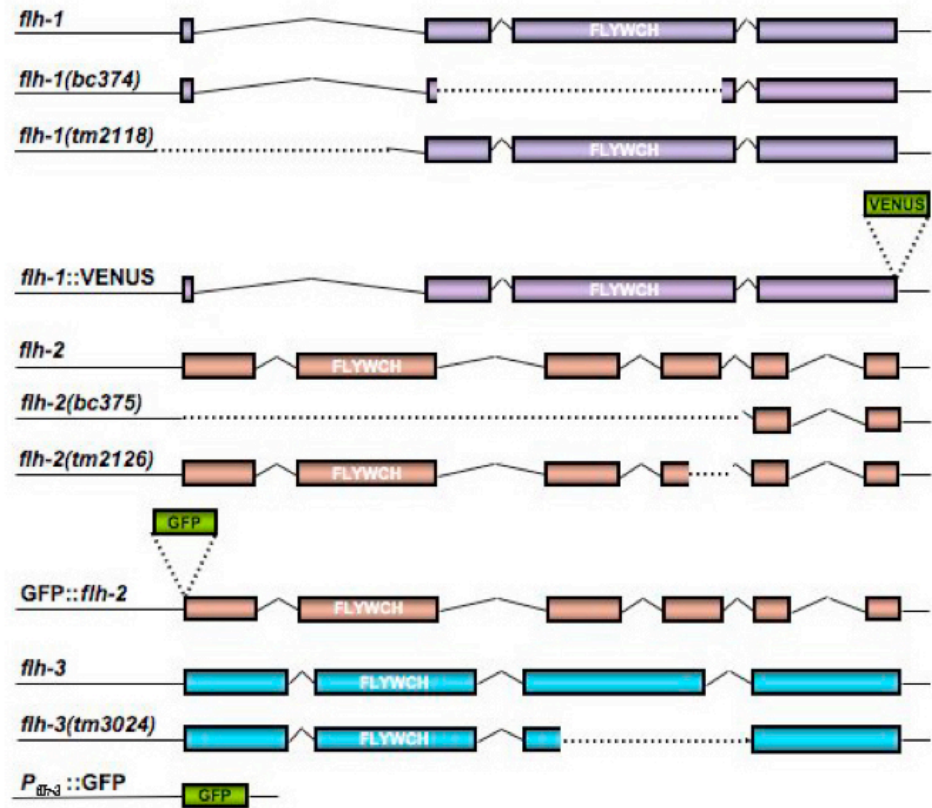
D



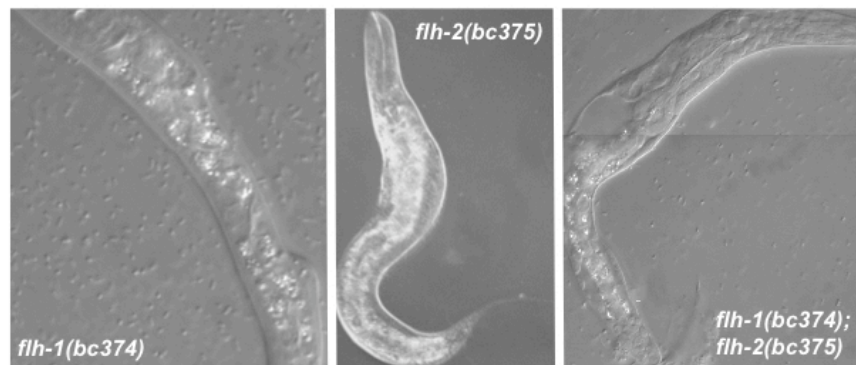
**Figure IV-1.** *FLH-1 is required for lin-4 repression in embryos.* (A) Deletion of the FLH-1-binding site in *Plin-4::gfp* results in precocious embryonic expression. (Top panel) Expression of *mals134* is absent in embryos. (Bottom panel) Animals with a modified *Plin-4::gfp* transgene with a deletion of the FLH-1-binding site (fragment 365–451) show aberrant fluorescent expression in late-stage embryos. (B) RNAi of FLH transcription factors results in increased levels of *lin-4* during embryogenesis. *flh-1(RNAi)*, but not *flh-2(RNAi)*, results in detectable levels of *lin-4* in embryos. *lin-4* levels are further elevated in *flh-1(RNAi); flh-2(RNAi)* embryos. No *lin-4* was detectable in the control RNAi using the empty RNAi vector. The U6 snRNA was used as the loading control. (C) Protein sequence alignment of FLH transcription factors in *C. elegans* and *C. briggsae*. ClustalW alignment of the amino acid sequences of *C. elegans* FLH-1, FLH-2, and FLH-3 and their respective *C. briggsae* orthologs—CBG15060, CBG18201, and CBG15055—shows conservation of the FLYWCH motif and the C terminus. The *flh-1* locus encodes two isoforms, FLH-1a and FLH-1b that differ by three amino acids; residues 339–341 (PLQ) in FLH-1a are absent from FLH-1b. Black highlight indicates identical amino acids and gray boxes indicate similar amino acids. The FLYWCH motif is shown in red. (D) RNAi of FLH transcription factors leads to precocious embryonic expression of *Plin-4::gfp*. Double RNAi-by-feeding of *flh-1* and *flh-2* in animals carrying the *Plin-4::gfp* transgene results in the precocious expression of GFP in late-stage embryos. RNAi in animals using the empty RNAi vector exhibited no GFP in embryos.

Figure IV-2

A



B



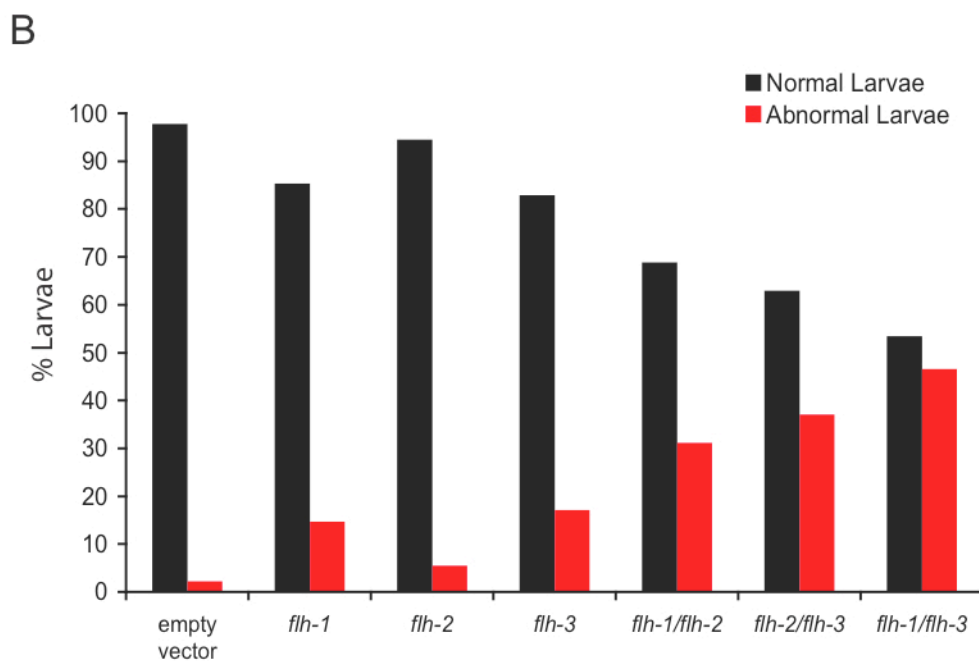
**Figure IV-2. Characterization of FLYWCH mutants.** (A) Schematic of the *flh-1*, *flh-2*, and *flh-3* loci, mutants, and reporter transgenes. The *flh-1* gene encodes two isoforms, FLH-1a and FLH-1b. In this study, we refer to the product of *flh-1a* as “FLH-1.” The nature of the *flh-1(bc374)* and *flh-2(bc375)* mutations is described under Results. The *flh-1(tm2118)* is a 707-bp deletion, from 287 bp upstream to 419 bp downstream of the translational start codon. The *flh-2(tm2126)* lesion is a 348-bp deletion. The *flh-3* locus is immediately upstream of *flh-1* and is transcribed in the opposite orientation. The *flh-3(tm3024)* mutant allele is a 337-bp deletion of most of exon 3. The white letters indicate the location of the FLYWCH domain. Dotted lines delineate deleted regions. The figure is not drawn to scale. (B) Phenotype of FLH-1 and FLH-2 mutants. Animals were observed using Nomarski DIC microscopy. (Left and middle panels) Single mutants of *flh-1(bc374)* and *flh-2(bc375)* exhibit a nearly wild-type phenotype with a low penetrance of larvae with morphological abnormalities. (Right panel) The double *flh-1(bc374); flh-2(bc375)* mutation also results in young larvae with morphological aberrations as well as a complete penetrance of early larval lethality.



Figure IV-3

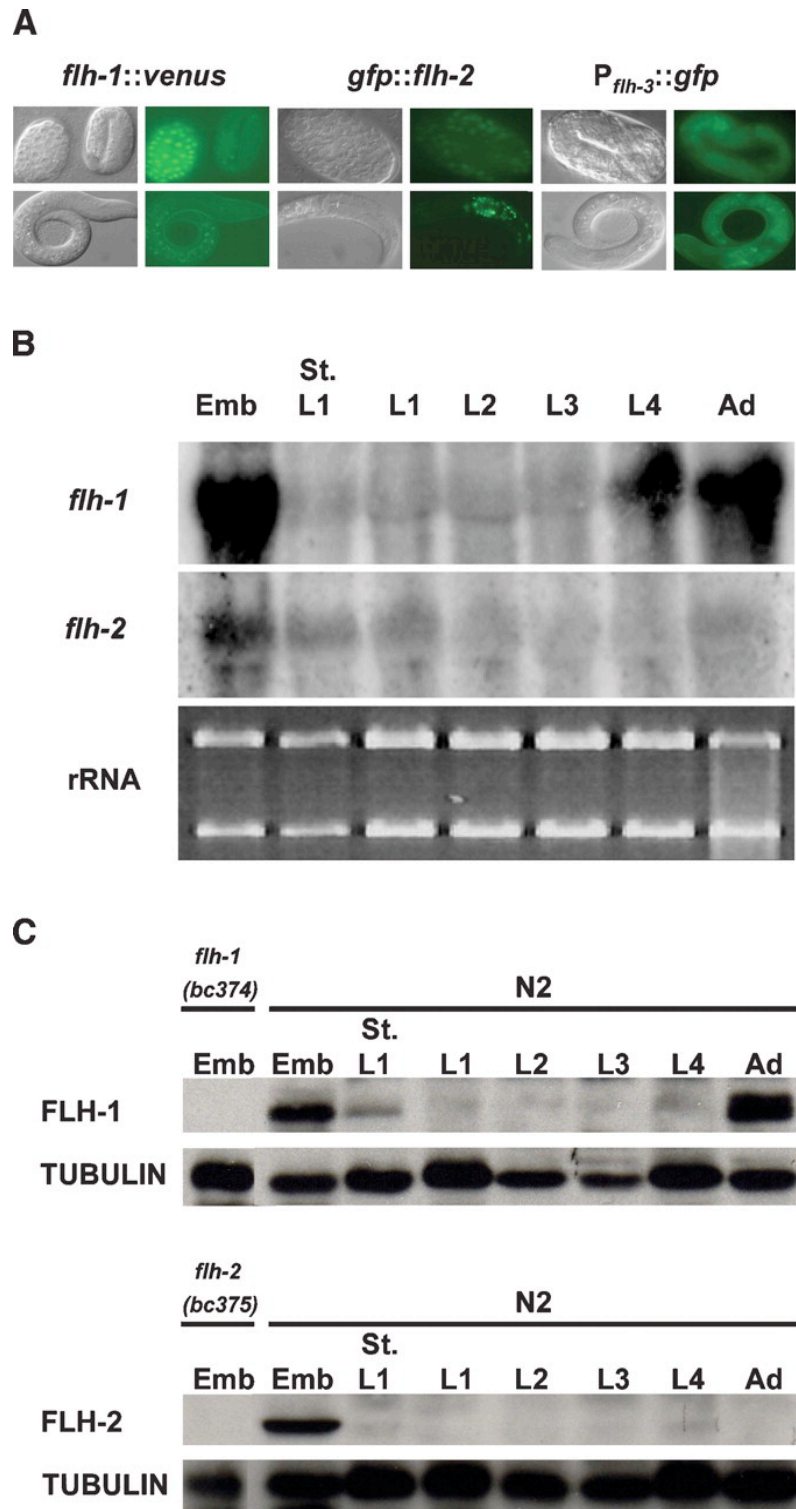
**A**

Mutant	Phenotype
<i>flh-1 (bc374)</i>	~WT, 2.8% abnormal larvae
<i>flh-1 (tm2118)</i>	~WT
<i>flh-1 (RNAi)</i>	11% abnormal larvae
<i>flh-2 (bc375)</i>	~WT, low penetrance of dpy
<i>flh-2 (tm2126)</i>	~WT
<i>flh-2 (RNAi)</i>	2% abnormal larvae
<i>flh-3 (tm3024)</i>	~WT
<i>flh-3 (RNAi)</i>	14% abnormal larvae
<i>flh-1 (bc374); flh-2 (bc375)</i>	Larval lethal
<i>flh-1 (bc374); flh-2 (tm2126)</i>	Larval lethal
<i>flh-1 (tm2118); flh-2 (tm2126)</i>	egl, unc, dpy phenotype
<i>flh-1 (RNAi); flh-2 (RNAi)</i>	28% abnormal larvae
<i>flh-2 (bc375); flh-3 (tm3024)</i>	egl phenotype
<i>flh-2 (RNAi); flh-3 (RNAi)</i>	34% abnormal larvae
<i>flh-3 (RNAi); flh-2 (RNAi)</i>	44% abnormal larvae



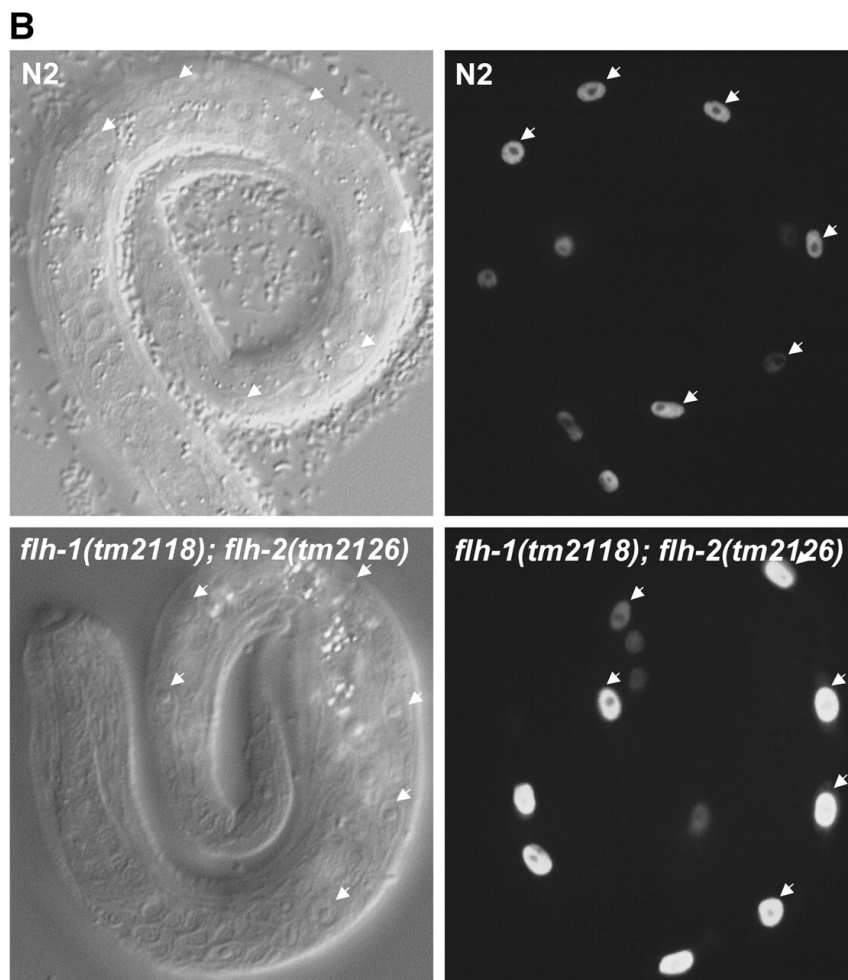
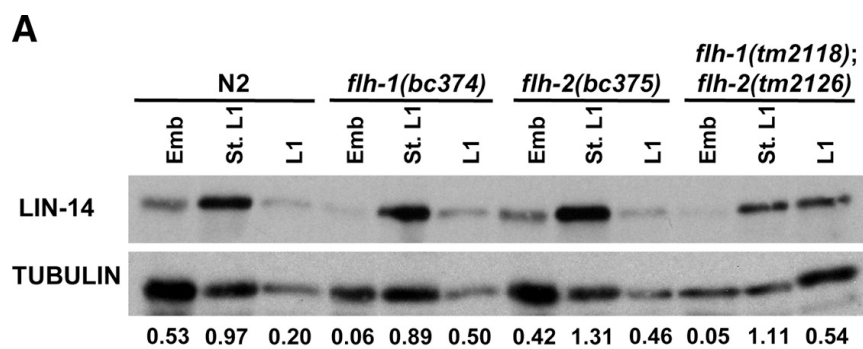
**Figure IV-3.** *Double flh mutants display severe phenotypes compared to single flh mutants.* (A) Summary of phenotypes observed in single and double *flh* mutants and RNAi knockdowns. Egl - egg laying defect; dpy – dumpy phenotype; unc – uncoordinated phenotype (B) Graph depicting the percentage of normal (black) and abnormal (red) larvae observed when *Plin-4::gfp(mals134);rrf-3* animals are fed with single or double combinations of *flh* RNAi food. Food containing empty vector was used as negative control. Food containing *unc-22* RNAi was used as positive control (90% uncoordinated worms, data not shown).

Figure IV-4



**Figure IV-4.** *Expression pattern of FLYWCH TFs.* (A) Expression pattern of FLH-1, FLH-2, and FLH-3. A translational fusion consisting of a VENUS reporter fused to the C terminus of *flh-1* (*Pflh-1::flh-1::venus::flh-1* 3' UTR) displays expression starting from mid-embryogenesis. (*Left panel*) VENUS expression is detected in most cells during the gastrulation stage but is down-regulated during late embryogenesis and is undetectable by L1. (*Middle panel*) GFP expressed from the rescuing translational fusion *gfp::flh-2* (*Pflh-2::gfp::flh-2::flh-2* 3' UTR) is detected in most cells during the gastrulation stage. (*Middle panel*) Unlike *flh-1::venus*, the *gfp::flh-2* reporter shows detectable expression in head and tail cells of larvae and adults. (*Right panel*) The expression of a GFP transcriptional reporter for *flh-3* (*Pflh-3::gfp*) displayed the most intense expression during late stages of embryogenesis and little GFP in mid-stage larvae. (*Right panel*) Expression from *Pflh-3::gfp* was also detected in L1 larvae. (B) Northern analysis of *flh-1* and *flh-2* mRNAs. Total RNA from N2 synchronized animals was analyzed by Northern blotting. Equivalent amounts of the RNA used for the Northern blots were run separately in parallel, and the levels of the rRNAs served as the loading control. The *flh-1* and *flh-2* mRNAs are similar in size to the ribosomal RNAs, and some cross-reactivity may have occurred between the *flh-1* and *flh-2* probes and the rRNAs. (C) Western analysis of FLH-1 and FLH-2. Protein lysates from synchronized animals were analyzed by Western blots with antisera to FLH-1, FLH-2, and tubulin. Protein extracts from the deletion mutants, *flh-1(bc374)* and *flh-2(bc375)*, show that the antibodies are specific to their corresponding antigen. Emb, St, L1, and Ad indicate embryos, starved L1 larvae, and adults, respectively.

Figure IV-5



**Figure IV-5.** *flh* mutants show upregulated levels of LIN-14 protein in embryos but no post-embryonic heterochronic defects. (A) LIN-14 levels in *flh* mutants. Western blots using protein extracts were first probed with anti-LIN-14 serum followed by stripping and reprobing with a tubulin antibody. Embryos that hatched overnight in M9 medium were used as the source of starved L1s. The numerical values represent the ratio of LIN-14 to tubulin. (B) Seam cells number in L1 larvae. Embryos with a seam cell nuclei-specific transgene (*scm::gfp*) were hatched overnight in sterile M9 followed by feeding on seeded NGM plates for 1–2 h. L1 larvae were then scored for the number of V lineage seam cells (V1–V6). Arrowheads point to the V1–V6 seam cells from one plane.

**Table IV-1.** Expression of the *col-19* heterochronic marker in wild-type (N2) and *flh* mutants.

Supplemental Table 1. *col-19::gfp* expression in *flh* mutants

Genotype	Percent of <i>col-19::gfp</i> expression	
	L4 larvae	Adult
N2	0 (n = 499)	100 (n = 726)
<i>flh-1(bc374)</i>	0 (n = 305)	100 (n = 746)
<i>flh-2(bc375)</i>	0 (n = 335)	100 (n = 739)
<i>flh-1(tm2118); flh-2(tm2126)</i>	0 (n = 126)	100 (n = 252)
<i>flh-2(bc375); flh-3(tm3024)</i>	0 (n = 94)	100 (n = 1000)

**Table IV-2.** *Number of seam cells of the V lineage observed in wild-type (N2) and flh mutants.*

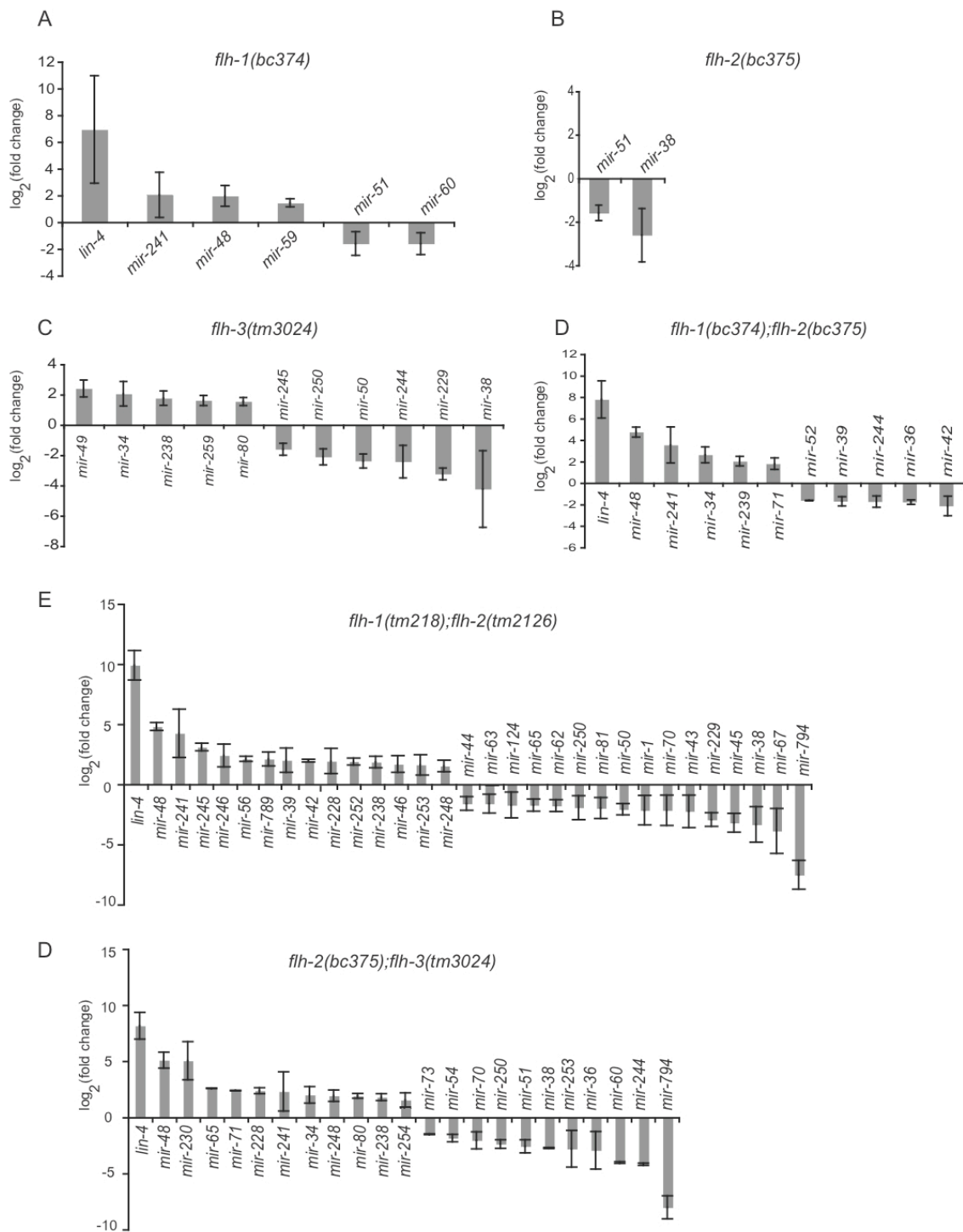
Supplemental Table 2. Number of seam cells<sup>a</sup> in *flh* mutant L1 larvae

Genotype	Percentage of L1 larvae with six seam cells
N2	100 (n = 32)
<i>flh-1(bc374)</i>	100 (n = 34)
<i>flh-2(bc375)</i>	100 (n = 27)
<i>flh-1(tm2118); flh-2(tm2126)</i>	100 (n = 44)

<sup>a</sup> Hypodermal seam cell nuclei of V1-V6 lineage



Figure IV-6

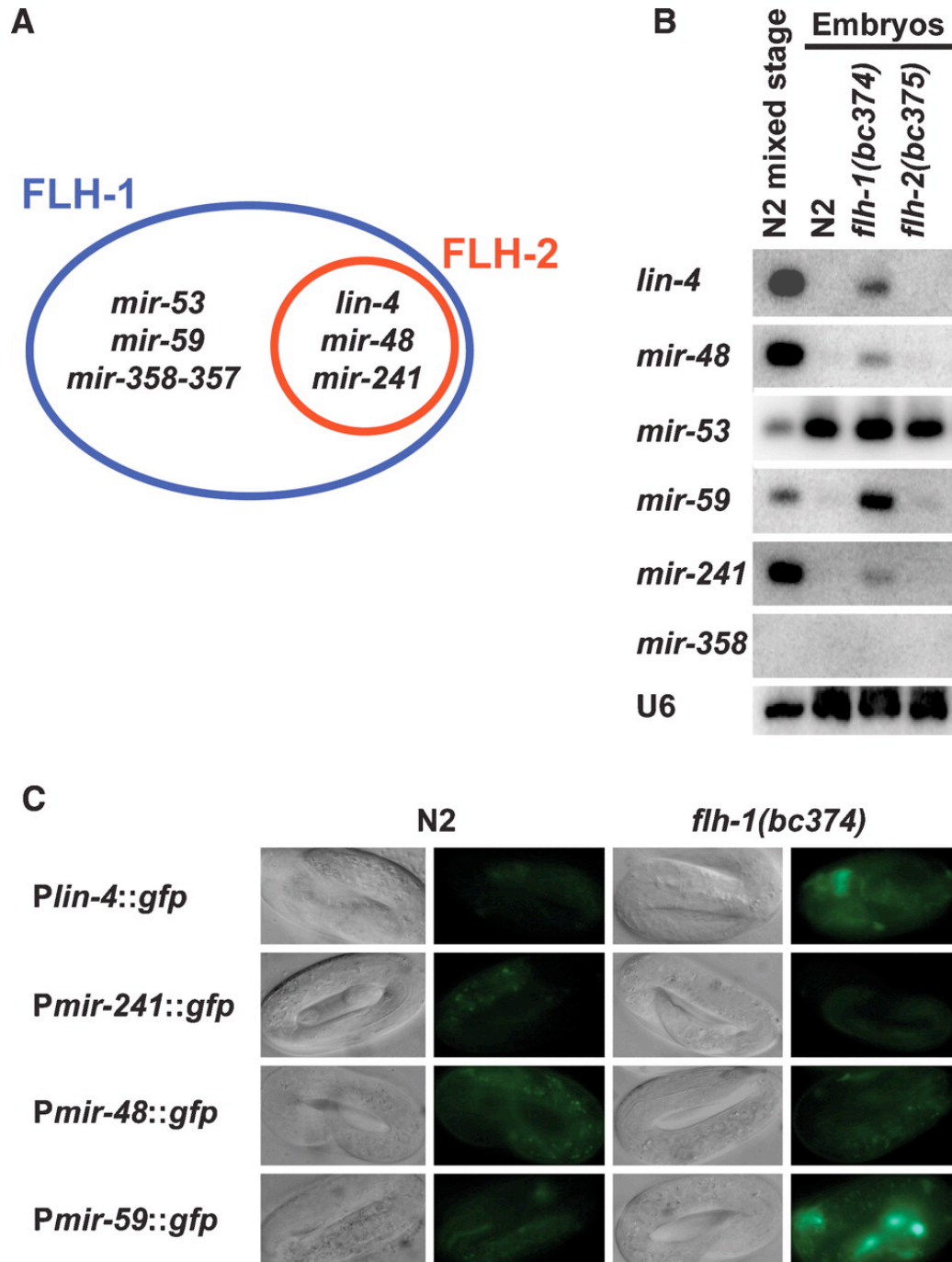


**Figure IV-6.** *MiRNAs that change significantly ( $P_{val} < 0.05$ ) in *flh* mutants compared to wild-type embryos.* Total RNA isolated from 100 late-stage embryos of wild type (N2) and *flh-1(bc374)* (A), *flh-2(bc375)* (B), *flh-3(tm3024)* (C), *flh-1(bc374); flh-2(bc375)* (D), *flh-1(tm2118); flh-2(tm2126)* (E), and *flh-2(bc375); flh-3(tm3024)* (F) were subjected to miRNA TaqMan real-time PCR assays. miRNA levels in mutants compared with the N2 control were determined using the  $\Delta\Delta C$  method and are expressed here as  $\log_2$  (fold change).

**Table IV-3.**  $\text{Log}_2(\text{fold change})$  values of miRNAs that changed significantly ( $P\text{val}<0.05$ ) in *flh* mutants compared to wild-type embryos.

miRNA	<i>flh-1</i> (bc374)	<i>flh-2</i> (bc375)	<i>flh-3</i> (tm3024)	<i>flh-1</i> (bc374); <i>flh-2</i> (bc375)	<i>flh-1</i> (tm2118); <i>flh-2</i> (tm2126)	<i>flh-2</i> (bc375); <i>flh-3</i> (tm3024)
<i>lin-4</i>	6.9282			7.78288075	9.886632969	8.153523109
<i>mir-124</i>					-1.73169581	
<i>mir-228</i>					1.930638476	2.360796955
<i>mir-229</i>			-3.2316		-2.94286208	
<i>mir-230</i>						5.04052662
<i>mir-238</i>			1.76461		1.842751742	1.797670216
<i>mir-239a</i>				2.03759021		
<i>mir-241</i>	2.0317			3.54687009	4.231196852	2.290057125
<i>mir-244</i>			-2.424	-1.72932455		-4.19443997
<i>mir-245</i>			-1.6053		3.091330962	
<i>mir-246</i>					2.388259519	
<i>mir-248</i>					1.509391646	1.909927955
<i>mir-250</i>			-2.1054		-1.95003626	-2.39891045
<i>mir-252</i>					1.883272745	
<i>mir-253</i>					1.604076629	-2.81458845
<i>mir-254</i>						1.536048955
<i>mir-259</i>			1.6142			
<i>mir-34</i>			2.05736	2.62812286		1.997778439
<i>mir-36</i>				-1.76940722		-2.94105038
<i>mir-38</i>		-2.6167	-4.2335		-3.35084592	-2.74977063
<i>mir-39</i>				-1.70075188	2.00318508	
<i>mir-42</i>				-2.12545955	1.96312642	
<i>mir-43</i>					-2.25897237	
<i>mir-44</i>					-1.60904748	
<i>mir-45</i>					-3.20186626	
<i>mir-46</i>					1.672896831	
<i>mir-48</i>	1.9577			4.74766542	4.793043519	5.081371882
<i>mir-49</i>			2.40815			
<i>mir-50</i>			-2.3792		-2.08732652	
<i>mir-51</i>	-1.603	-1.5907				-2.58661005
<i>mir-52</i>				-1.62558979		
<i>mir-54</i>						-1.85481971
<i>mir-56</i>					2.100927075	
<i>mir-59</i>	1.4423					
<i>mir-60</i>	-1.612					-4.03037605
<i>mir-62</i>					-1.788451	
<i>mir-63</i>					-1.61152759	
<i>mir-65</i>					-1.74663342	2.57579462
<i>mir-67</i>					-3.88211436	
<i>mir-70</i>					-2.17096648	-2.05921238
<i>mir-71</i>				1.80306429	-2.16764983	2.374698455
<i>mir-73</i>						-1.52239255
<i>mir-789</i>					2.099770585	
<i>mir-794</i>					-7.54333019	-8.03832255
<i>mir-80</i>			1.54485			1.904101955
<i>mir-81</i>					-1.98651703	

Figure IV-7



**Figure IV-7.** *Y1H assays identify additional miRNA targets.* (A) Venn diagram of FLH-1 (blue) and FLH-2 (red) miRNA targets identified by genome-scale Y1H screens. (B) FLH-1 represses the expression of its miRNA targets during embryogenesis. Total RNA from N2, *flh-1(bc374)*, or *flh-2(bc375)* embryos was extracted and analyzed by Northern blots. RNA from a mixed-stage population of N2 was loaded as reference. *mir-358* was undetectable even after prolonged exposures. All blots were stripped and reprobated for the U6 snRNA as the loading control. (C) Transcriptional reporters for *lin-4*, *mir-241*, and *mir-59* show precocious expression during embryogenesis in an FLH-1 mutant. Late-stage embryos from N2 and *flh-1(bc374)* expressing the *Plin-4::gfp*, *Pmir-241::gfp*, *Pmir-48::gfp*, or *Pmir-59::gfp* transgenes were examined by fluorescence and DIC optics. All images were of identical exposure time and were processed in parallel.

**Table IV-4.** Comparison of various analyses used in this study.

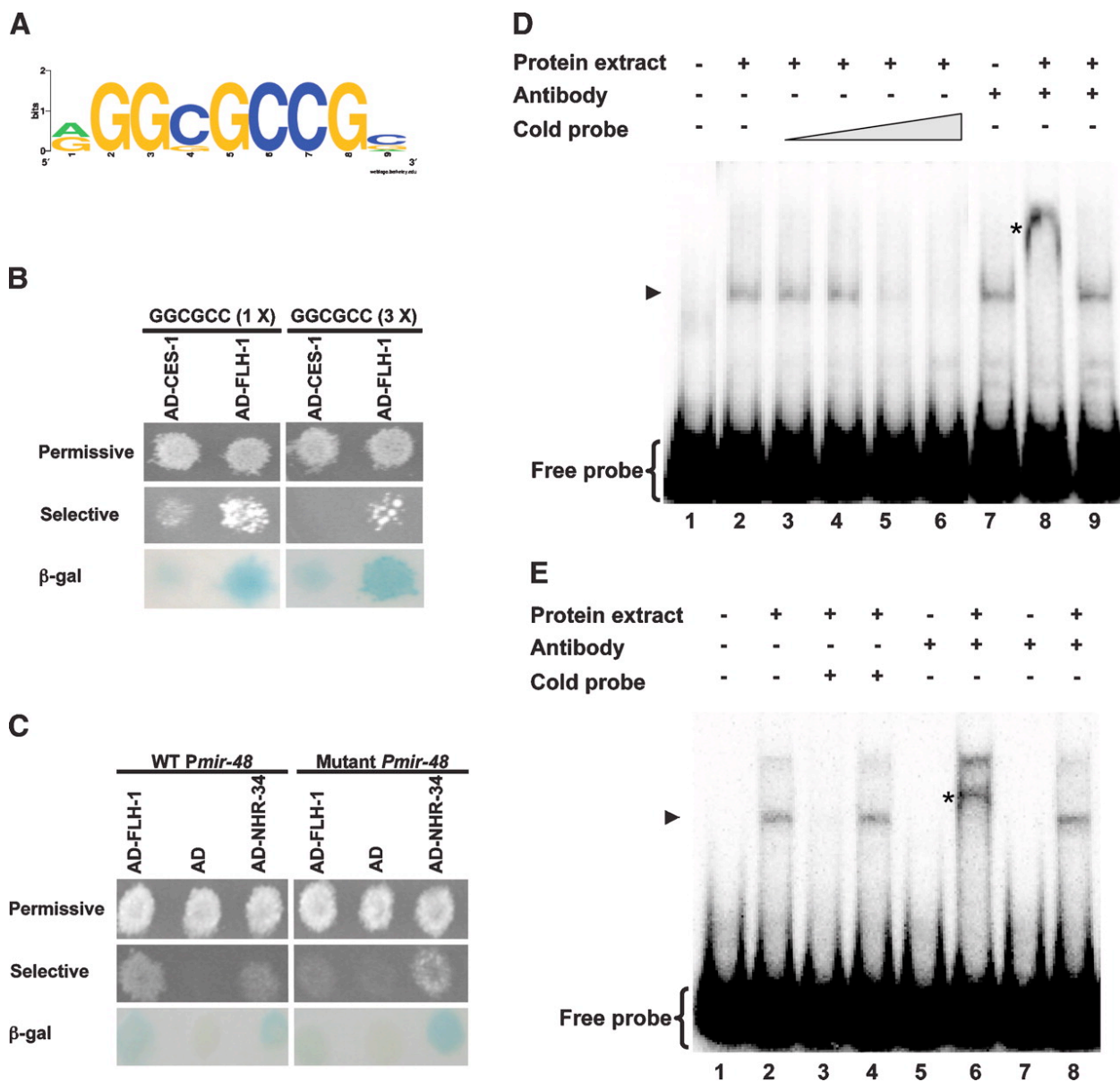
Change in miRNA levels			
FLH-1 Y1H screen	<i>Pmir::gfp</i>	Northern blot	TaqMan assay
<i>lin-4</i>	Up	Up	Up
<i>mir-241</i>	Up	Up	Up
<i>mir-48</i>	Same	Up	Up
<i>mir-53</i>	Same	Same	Same
<i>mir-59</i>	Up	Up	Same
<i>mir-358-357</i>	ND	ND	Same <sup>a</sup>

ND = not detected  
<sup>a</sup> = *mir-357*

Change in miRNA levels			
FLH-2 Y1H screen	<i>Pmir::gfp</i>	Northern blot	TaqMan assay
<i>lin-4</i>	Same	Same	Same
<i>mir-241</i>	Same	Same	Same
<i>mir-48</i>	Same	Same	Same

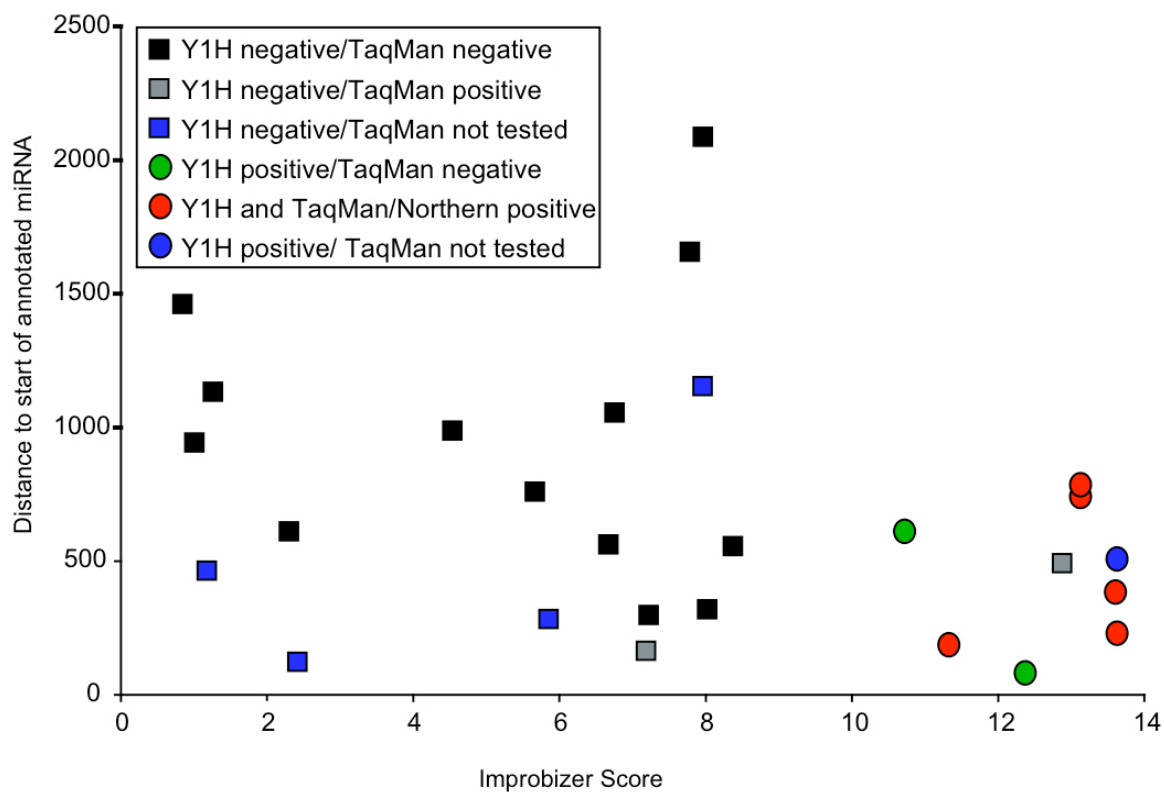
Figure IV-8



**Figure IV-8. Identification of an FLH-1 consensus binding site.** (A) A predicted FLH-1 consensus binding site. The miRNA targets found for FLH-1 were used to predict a consensus binding site of a/gGGCGCCG. (B) Y1H assays show the interaction between AD-FLH-1 and DNA baits containing either one (*left panels*) or three (*right panels*) copies of the FLH-1 consensus binding sites. AD-CES-1 was used as the negative control. (C) Deletion of the predicted consensus binding site in *Pmir-48* abolishes binding by FLH-1 in Y1H assays. (*Left panels*) Y1H assays show the interaction between AD-FLH-1 and DNA baits consisting of a wild-type *Pmir-48*. (*Right panels*) Deletion of the consensus sequence disrupts binding by AD-FLH-1. AD-NHR-34 was used as a positive control. We find AD-NHR-34 binding to wild-type *Pmir-48*, and deletion of the FLH-1-binding site does not disrupt AD-NHR-34 binding. ( $\beta$ -Gal)  $\beta$ -Galactosidase assay. (D) In vitro binding of FLH-1 to the consensus binding site in *Plin-4*. Gel mobility shift assays were performed using a [ $\gamma$ - $^{32}$ P] 5'-end-labeled 87-bp fragment (fragment 365–451) from *Plin-4* incubated with total protein extract from N2 embryos. Competition assays used (lanes 3–6) increasing quantities of unlabeled fragment 365–451. (Lane 8) Antiserum against FLH-1 was added to show that the shifted material contains FLH-1. (Lane 9) Addition of a control rabbit IgG failed to form a supershifted complex. (Lane 1) Free probe only. (Lane 2) Probe incubated with protein extract. (Lanes 3–6) Probe incubated with protein extract and increasing amounts of unlabeled probe (0.1-, 1-, 10-, and 100-fold excess, respectively). (Lane 7) Probe with FLH-1 antibody but without protein extract. (Lane 8) Probe with protein extract and FLH-1 antiserum. (Lane 9) Probe with protein extract and control rabbit IgG. (Lanes 2–5,7,9) Arrowhead points to the shifted probe-FLH-1 complex. (Lane 8) Asterisk denotes the probe-FLH-1-FLH-1 antibody supershifted complex. (E) In vitro binding of FLH-1 to the consensus binding site in *Pmir-48*. Gel shift assays were done using a [ $\gamma$ - $^{32}$ P] 5'-end-labeled 51-bp fragment (fragment 200–251) from *Pmir-48* and total protein extract from N2 embryos. (Lane 2) A shifted complex is seen in the sample incubated with the protein extract. Competition assays were done using both a 100-fold excess of unlabeled fragment 200-251 (lane 3) or unlabeled fragment 200–251 with a deletion of the consensus site (lane 4). (Lanes 6,8) Supershifted complexes were detected upon the addition of FLH-1 antiserum but not with control rabbit IgG. (Lane 1) Free probe only. (Lane 2) Probe incubated with protein extract. (Lane 3) Probe with protein extract and 100-fold excess of cold fragment 200–251. (Lane 4) Probe with protein extract and 100-fold excess of fragment 200–251 with a deletion of the consensus binding site. (Lane 5) Probe with FLH-1 antibody and no extract. (Lane 6) Probe and extract with FLH-1 antibody. (Lane 7) Probe with control rabbit IgG and no extract. (Lane 8) Probe and extract with control rabbit IgG. (Lanes 2,4,8) Arrowhead indicates the shifted probe-FLH-1 complex. (Lane 6) Asterisk shows the probe-FLH-1-FLH-1 antibody supershifted complex.

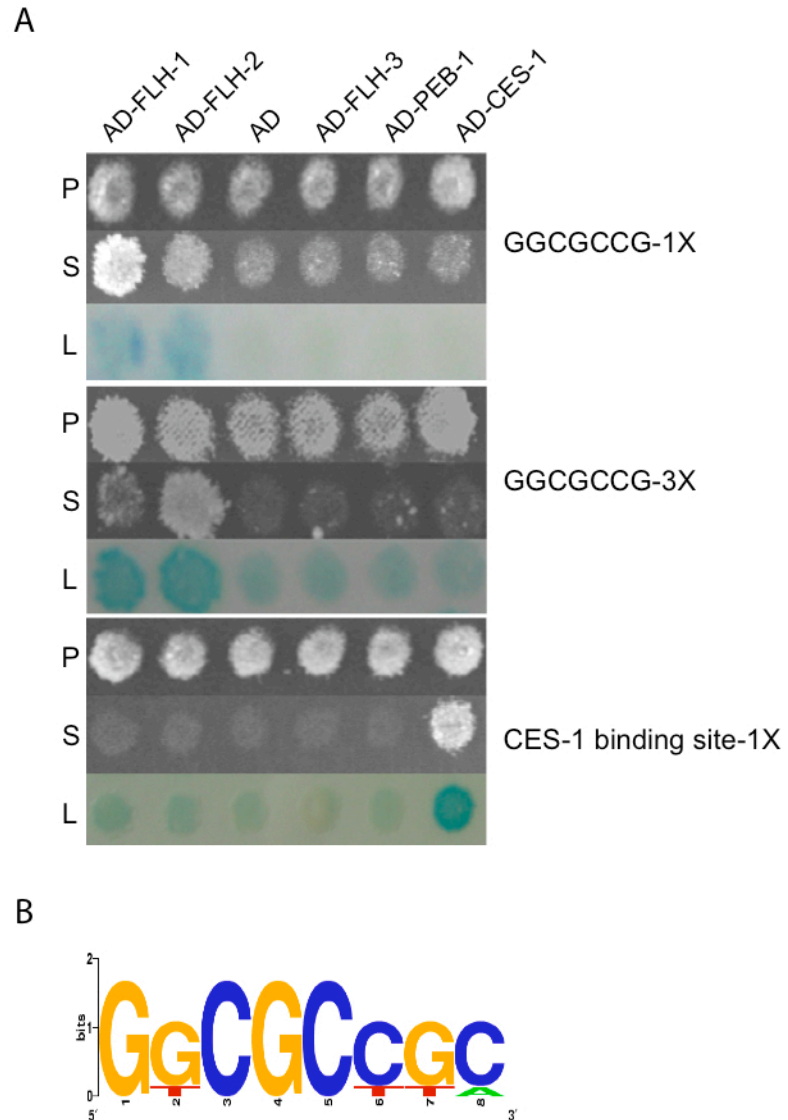


Figure IV-9

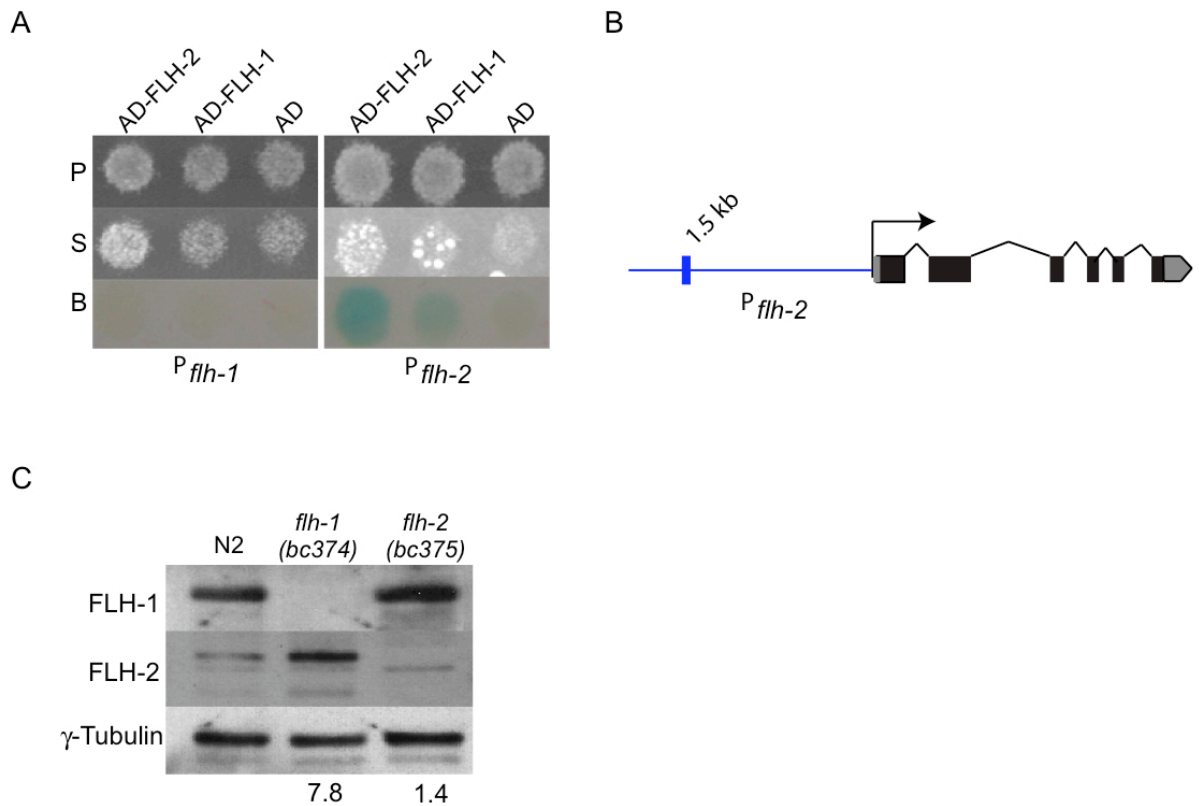


**Figure IV-9.** *Distribution of the FLH-1 consensus binding site.* Plot depicting the distribution of FLH-1 consensus binding sites among a set of miRNA promoters that bind (circles) or do not bind (squares) FLH-1 in Y1H assays according to the distance to the annotated miRNA start site. The promoter of *mir-241* contains two consensus sites. The *lin-4* promoter of *C. briggsae* binds FLH-1 in Y1H (data not shown) and it was included in the binding site search.

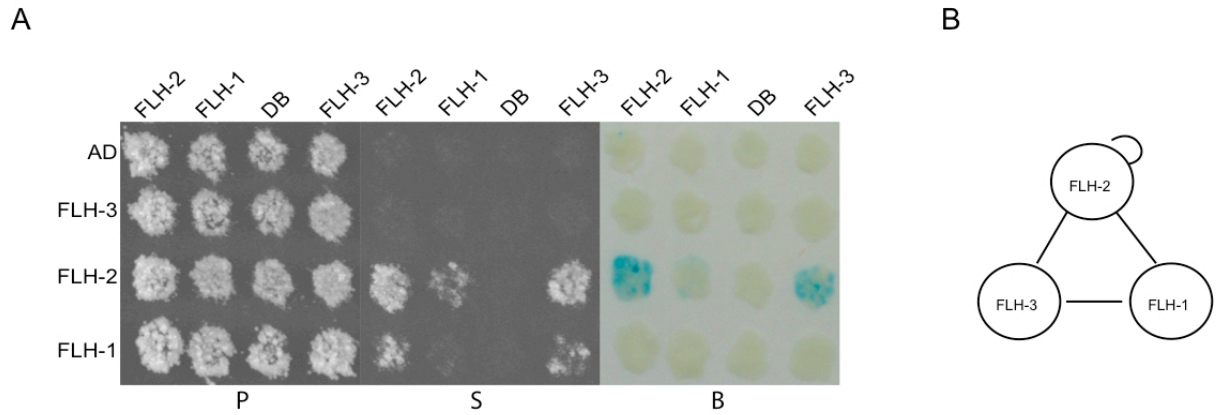
Figure IV-10



**Figure IV-10.** *FLH-1* and *FLH-2* can bind the same consensus binding site. (A) Y1H assays show the interaction between AD-FLH-1 and AD-FLH-2 fusions and DNA baits containing either one (GGCGCCG-1X) or three (GGCGCCG-3X) consensus binding sites. AD-CES-1 and a DNA bait containing CES-1 binding site were used as control. P – permissive media; S – selective media; B –  $\beta$ -Galactosidase assay; AD – empty vector. (B) The miRNA and protein-coding targets found for FLH-2 were used to predict a consensus binding site of GGCGCCGC.

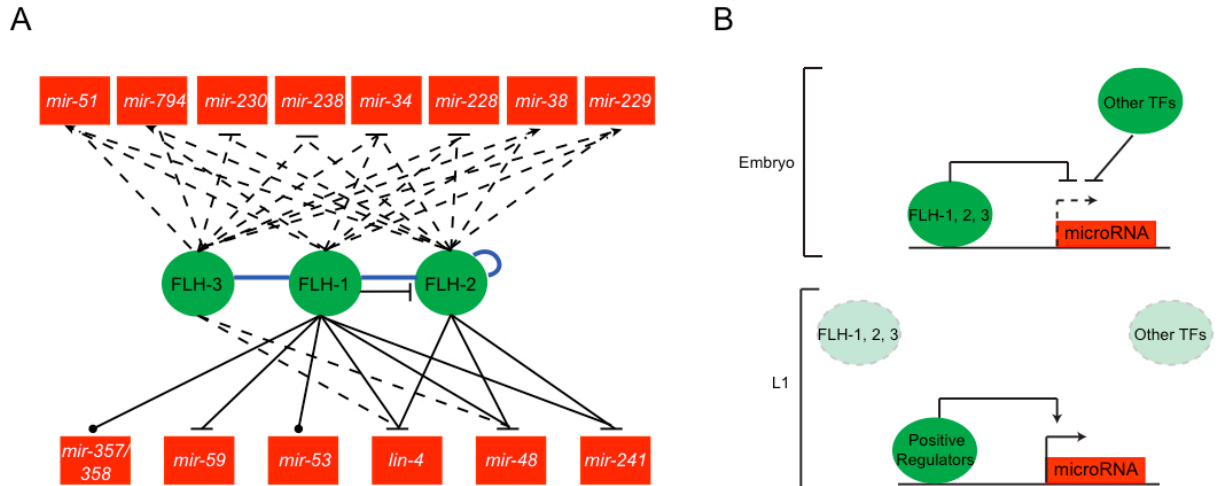
**Figure IV-11**

**Figure IV-11. *FLH-1* represses *flh-2* expression** (A) *FLH-1* and *FLH-2* can interact with *Pflh-2* but not with *Pflh-1* in Y1H assays. P – permissive media; S – selective media; B –  $\beta$ -Galactosidase assay; AD – empty vector. (B) Cartoon showing the presence of a *FLH-1* consensus binding site in *Pflh-2*. (C) Western analysis of *FLH-1* and *FLH-2*. Protein lysates from wild-type (N2), *flh-1(bc374)* and *flh-2(bc375)* animals were analyzed by Western blots with antisera to *FLH-1*, *FLH-2*, and tubulin. For each *flh* mutant, levels of the opposite *FLH* protein were normalized to tubulin control and compared to normalized levels in N2 animals. Fold differences are indicated in the bottom.

**Figure IV-12**

**Figure IV-12.** *FLYWCH* TFs engage in protein-protein interactions. (A) Y2H assays reveal protein-protein interactions between FLH-1, FLH-2 and FLH-3. Rows – AD-FLH fusions; columns – DB-FLH fusions; P – permissive media; S – selective media; B –  $\beta$ -Galactosidase assay; AD and DB – empty vectors. (B) Cartoon depicting interactions between *FLYWCH* transcription factors. FLH-2 interacts with itself and with FLH-1 and FLH-3. FLH-1 and FLH-3 also interact with each other as previously reported (Walhout et al. 2002).

Figure IV-13



**Figure IV-13.** Working model: *FLYWCH* transcription factors regulate miRNA expression. (A) Network depicting regulatory effects on miRNAs (red rectangles) exerted by FLH-1, FLH-2 and FLH-3 (green circles). Solid lines are TF–*Pmir* interactions found by Y1H and/or TaqMan assays; dashed lines represent FLH regulation (direct or indirect) based on miRNAs that change in expression in *flh* mutants using TaqMan assays; dotted lines are TF–*Pmir* physical interactions (no regulatory interaction determined); arrows denote transcriptional activation; blunted arrows denote transcriptional repression; blue lines denote protein-protein interactions found by Y2H. (B) A working model for FLH-1, FLH-2, and FLH-3 TFs. FLH-1, and/or FLH-2, in association with FLH-3, are proposed to directly bind to sequences containing the consensus a/gGGCGCCG in the promoters of *lin-4*, *mir-241*, and *mir-48* during embryogenesis. This binding prevents the premature transcription of these miRNAs. Other TFs could also repress the expression of these miRNAs. Upon hatching, the FLH proteins are down-regulated, releasing the repression on miRNA transcription and allowing the action of other factors, including putative transcriptional activators, to activate miRNA expression during larval development.

**Table IV-5.** List of *C. elegans* strains used in this study.

Strain	Genotype
JR667	<i>wls51 (scm::gfp)</i>
N2	wild-type var Bristol
NL2099	<i>rrf-3(pk1426)</i> II
OK257	<i>peb-1(cu9)/dpy-3(e27) unc-2(e55)</i> X
VT773	<i>mals105 (col::gfp)</i> V
VT1072	<i>unc-119(ed3)</i> III; <i>mals134 (Plin-4::gfp; unc-119+)</i>
VT1095	<i>mals134; rrf-3(pk1426)</i> II
VT1189	<i>unc-119(ed3)</i> III; <i>mals140 (Pmir-241::gfp; unc-119+)</i>
VT1216	<i>rrf-3(pk1426)</i> II; <i>mals140</i>
VT1256	<i>unc-119(ed3)</i> III; <i>mals147 (Plin-4 <math>\Delta</math>fragment 365-451::gfp)</i>
VT1259	<i>unc-119(ed3)</i> III; <i>mals150 (Pmir-48::gfp; unc-119+)</i>
VT1269	<i>rrf-3(pk1426)</i> II; <i>mals150</i>
VT1343	<i>flh-1(bc374)</i> IV
VT1348	<i>flh-1(bc374)</i> IV; <i>mals134</i>
VT1351	<i>flh-1(bc374)</i> IV; <i>mals140</i>
VT1353	<i>flh-1(bc374)</i> IV; <i>mals150</i>
VT1354	<i>flh-1(bc374)</i> IV; <i>mals105</i>
VT1355	<i>flh-1(bc374); wls51</i>
VT1376	<i>flh-2(bc375)</i> III
VT1378	<i>unc-119(ed3)</i> III; <i>mals161 (Pmir-59::gfp; unc-119+)</i>
VT1382	<i>flh-1(bc374)</i> IV; <i>mals161</i>
VT1392	<i>flh-1(tm2118)</i> IV
VT1393	<i>flh-2(tm2126)</i> III
VT1410	<i>flh-2(bc375)</i> III; <i>mals105</i>
VT1411	<i>flh-2(bc375); wls51</i>
VT1469	<i>flh-2(bc375)</i> III; <i>flh-1(bc374)</i> IV/nT1 [ <i>qls51</i> ] (IV;V)
VT1508	<i>unc-119(ed3)</i> III; <i>mals207 (Pflh-2::gfp::flh-2::flh-2 3'UTR; unc-119+)</i>
VT1514	<i>flh-2(tm2126)</i> III; <i>flh-1(tm2118)</i> IV
VT1589	<i>lin-4(e912)</i> II/mnC1; <i>flh-2(bc375)</i> III; <i>flh-1(bc374)</i> IV/nT1 [ <i>qls51</i> ] (IV;V)
VT1590	<i>alg-2(ok304)</i> II; <i>flh-2(bc375)</i> III; <i>flh-1(bc374)</i> IV/nT1 [ <i>qls51</i> ] (IV;V)
VT1593	<i>flh-2(tm2126)</i> III; <i>flh-1(tm2118)</i> IV; <i>mals105</i>
VT1613	<i>lin-4(e912)/mnc1; mals105</i>
VT1645	<i>flh-2(bc375)</i> III; <i>flh-1(bc374)</i> IV/nT1 [ <i>qls51</i> ] (IV;V); <i>mir-48&amp;mir-241(nDf51)</i> V
VT1674	<i>rrf-3(pk1426)</i> II; <i>mals161</i>
VT1846	<i>unc-119(ed3)</i> III; <i>mals304 (Pflh-3::gfp::flh-3 3' UTR; unc-119+)</i>
VT1860	<i>unc-119(ed3)</i> III; <i>mals318 (Pflh-1::flh-1::venus::flh-1 3'UTR; unc-119+)</i>
VT1891	<i>flh-3(tm3024)</i> IV
VT1899	<i>flh-2(bc375)</i> III; <i>flh-3(tm3024)</i> IV
VT1950	<i>flh-1(tm2118)</i> IV; <i>flh-2(tm2126)</i> III; <i>wls51</i>
VT1968	<i>flh-2(bc375)</i> III; <i>flh-3(tm3024)</i> IV; <i>mals105</i>
VT2006	<i>flh-2(bc375)</i> III; <i>flh-1(bc374)</i> IV; <i>mals207</i>

## PREFACE TO CHAPTER V

Part of this chapter has been published separately in:

Martinez, N. J., and Walhout, A. J. 2009. The interplay between transcription factors and microRNAs in genome-scale regulatory networks. *Bioessays* **31**: 435-445.

## **CHAPTER V**

### **Discussion**



## **MiRNA networks: Lessons from genome-scale studies**

### *Global properties of gene regulatory networks involving miRNAs and TFs*

We have recently used Y1H assays to map a genome-scale miRNA transcriptional network (TF=>miRNA) by experimentally identifying the TFs that can interact with *C. elegans* miRNA promoters (Martinez et al. 2008a). We found that the degree distribution of nodes in the miRNA transcriptional network identifies clear TF hubs but not miRNA promoter hubs. (Fig. V-1). This indicates that the topology of miRNA transcription regulatory networks is similar to that of protein-coding gene networks, and that, therefore, miRNA expression overall is regulated in a similar manner as protein-coding genes (Martinez et al. 2008a).

Shalgi and colleagues have analyzed the global degree of a mammalian post-transcriptional miRNA=>target network. This network consisted of computationally predicted interactions between miRNAs and their target genes, which were evolutionarily conserved in four species: human, mouse, rat, and dog. By plotting the in-degree distribution of target genes they showed the existence of hundreds of genes that are subject to extensive miRNA regulation, referred to as “target hubs”. Noticeably, many TFs are among these target hubs, suggesting that “regulating the regulators” is of particular importance (Fig. V-1). Interestingly, the out-degree distribution of miRNA nodes is not scale-free and although multiple miRNAs appear to target many genes, no clear “miRNA hubs” could be identified (Shalgi et al. 2007). Similarly, we have shown that in the predicted *C. elegans* post-transcriptional miRNA=>TF network, the out-degree

distribution of miRNA nodes is not scale-free, indicating the absence of clearly identifiable miRNA hubs (Fig. V-1) (Martinez et al. 2008a). *C. elegans* can tolerate removal of most individual miRNAs without obvious developmental defects (Miska et al. 2007). The lack of clear miRNA hubs and the fact that most of them are not essential for viability both agree with the hypothesis that miRNAs, in contrast to TFs, do not function as master regulators, but rather predominantly function to fine tune gene expression instead of establishing crucial developmental gene expression programs (Bartel and Chen 2004; Hornstein and Shomron 2006). Taken together, transcriptional and post-transcriptional regulatory networks exhibit different in- and out-degree distributions that correlate with the overall biological roles for TFs and miRNAs.

#### *Regulatory circuits containing TFs and miRNAs*

Several lines of evidence suggest that regulatory circuits involving miRNAs and TFs are not isolated instances but are in fact prevalent mechanisms of gene expression at the genome-scale level. First, by using information derived from genome-scale gene expression assays, several bioinformatic studies found that the expression of miRNAs and their targets is often highly correlated or anti-correlated (Farh et al. 2005; Sood et al. 2006; Stark et al. 2005; Tsang et al. 2007). Tsang *et al.* proposed that correlation or anti-correlation in expression between miRNAs and their targets can be explained by the existence of various types of feed-forward and feedback loops involving miRNAs and TFs (Tsang et

al. 2007). However, they did not identify any actual loops in which the participating components are known. Second, Shalgi and colleagues searched for pairs of human miRNAs and TFs that regulate sets of common target genes by identifying putative miRNA sites and TFBSs that co-occur in individual genes. They observed that such miRNA-TF pairs are predicted to regulate each other more frequently than randomly picked pairs, suggesting the existence of feed-forward and feedback loops. Third, Wang *et al.* identified conserved *cis*-regulatory elements surrounding miRNAs in 12 *Drosophila* species (Wang *et al.* 2008). These elements were enriched in known TFBSs. By integrating their predictions with miRNA target interactions they identified putative regulatory feedback loops between miRNAs and TFs. Finally, by integrating the first experimentally defined *C. elegans* TF=>miRNA transcription regulatory network with a predicted miRNA=>TF target network, we uncovered more than 20 miRNA<=>TF feedback loops, where the TF that regulates a miRNA is itself regulated by that same miRNA. We found that such feedback loops occur more frequently than expected by chance (*i.e.* in computationally randomized networks) and, hence, constitute a network motif. This demonstrates that such loops provide a general mechanism of gene expression (Martinez *et al.* 2008a). We have also identified several feed-forward motifs by integrating the upstream regulators of TFs, however, further mapping of transcriptional networks of protein-coding genes will be necessary to determine if miRNA-containing feed-forward loops constitute network motifs at a genome-scale.

Altogether, these network studies revealed not only the existence of reciprocal regulation between miRNAs and TFs but also the existence of extensive coordination in the regulation of shared target genes at the genome-scale level. As a result, genes that are regulated both by transcriptional and by post-transcriptional mechanisms may be tuned at a higher level of precision than could be obtained by either mechanism alone. In support of this idea, a recent computational study found that genes with many TFBSs have a higher probability to be controlled by miRNAs than genes with fewer TFBSs, illustrating the coordinated action that TFs and miRNAs exert on gene expression (Cui et al. 2007).

By examining the degree of the TFs and miRNAs that participate in feedback loops in the integrated *C. elegans* network, we found that most loop participants exhibit both a high in- and out-degree (Martinez et al. 2008a). In most miRNA $\rightleftharpoons$ TF feedback loops, the miRNAs regulate many TFs and are regulated by many TFs, and *vice versa*. The fact that both the miRNAs and the TFs involved in feedback loops have many downstream targets implies that the expression of not only the TF and the miRNA, but also of their respective targets is tightly coordinated. Based on these observations, we propose that gene regulatory effects exerted by TFs and miRNAs can “spread” to large sets of genes. This “regulatory spreading” could in principle be important in regulating large sets of genes, or gene batteries, for instance in different tissues or in response to developmental or environmental cues (Fig. V-2A). By integrating

human protein-protein interactions with miRNA-target interactions, Liang and Li found that proteins regulated by several miRNAs tend to have a higher degree of connectivity in protein-protein interactions networks. The targeting of protein-protein interaction hubs by many miRNAs may consequently affect a large number of interacting proteins as well (Fig. V-2B) (Liang and Li 2007).

In contrast to composite miRNA $\rightleftharpoons$ TF feedback loops, TF $\rightleftharpoons$ TF feedback loops, composed purely of transcriptional interactions, are not overrepresented in gene regulatory networks (Shen-Orr et al. 2002). Why are composite feedback loops preferred over purely transcriptional feedback loops? Transcriptional regulation is generally slow since a TF needs to be transcribed and translated before it can regulate its targets. In case of negative feedback loops, slow responses can lead to instability and noisy oscillations, which may be detrimental for homeostatic systems (Yeager-Lotem et al. 2004; Elowitz et al. 2000). Because of their small size and non-coding nature, miRNA synthesis can be much more rapid than that of TFs. It is tempting to speculate that by mixing a slow transcriptional interaction with a faster pos-transcriptional interaction, the response time could be greatly decreased. Future mathematical and/or experimental approaches will be needed to corroborate this hypothesis.

Although examples of feedback and feed-forward loops have been found experimentally in several organisms, their function and implications on gene expression have been mainly studied by mathematical models or by creating small synthetic networks with loops in *Escherichia coli* (Becskei and Serrano

2000; Elowitz and Leibler 2000; Gardner et al. 2000). Importantly, these synthetic loops consisted only of transcriptional interactions. The functions of these circuits are much more difficult to assess in metazoan networks mostly because these circuits are part of highly interconnected sub-graphs and their disruption may not result in detectable effects. The generation and study of synthetic networks in metazoan systems, containing transcriptional as well as post-transcriptional interactions involving small RNA components, will be essential to further understand how network motifs perform during multicellular development and physiology.

### **Genome-scale analysis of miRNA promoter activity: Lessons and applications**

We have described a set of transgenic *C. elegans* strains, each expressing GFP under the control of a miRNA promoter as a proxy to study miRNA expression *in vivo*. Importantly, the majority of *PmiRNA::gfp* lines completely or partially recapitulate previously reported temporal expression of miRNAs (Ambros et al. 2003; Lau et al. 2001; Lee and Ambros 2001; Lim et al. 2003). Interestingly, our data suggest that post-transcriptional mechanisms contribute to differential miRNA expression. The data and transgenic lines that we generated provide a platform for functional miRNA studies to delineate their roles in the development of the animal, and to understand their function in gene regulatory networks.

### *Spatial and temporal analysis of miRNA promoter activity*

We found that miRNA promoters, similar to protein-coding gene promoters, are active in all major tissues and cell types. These findings are consistent with previous studies in zebrafish and fruit flies that have also shown a wide breadth of expression for many miRNAs (Aboobaker et al. 2005; Wienholds et al. 2005). Taking into account that miRNAs potentially target 10-30% of the genome, our findings are in agreement with the idea that *C. elegans* miRNAs can be involved in a wide range of biological processes.

Several groups have investigated the consequences of globally affecting miRNAs by inactivating enzymes required for miRNA processing. In mouse, *dicer* knockouts are embryonic lethal, however conditional knockout of *dicer* in limbs causes apoptosis without affecting the patterning and differentiation of the limb (Harfe et al. 2005). In zebrafish, *dicer* mutants are devoid of mature miRNAs and have morphogenesis defects, but differentiate multiple cell types and the overall patterning of the body remains normal (Schier and Giraldez 2006; Wienholds et al. 2005). *C. elegans dicer* mutants display developmental timing defects and sterility and mutations in *alg-1*, an Argonaute protein involved in the miRNA pathway, causes a wide range of defects including developmental timing defects, however mutants are viable (Grishok et al 2001). These studies appear to indicate that miRNAs are not obligatory components of all fate specification or signaling pathways but facilitate developmental transitions and have roles after the initial specification of a body plan, organs and tissues (Schier and Giraldez

2006; Wienholds et al. 2005). Consistent with this idea, we find that *C. elegans* miRNA promoters are active later in development compared to protein-coding gene promoters. However, we should not rule out the possibility that miRNAs play roles during early *C. elegans* embryogenesis. Even though, we do not detect miRNA promoter activity in the germinal gonad, we do so in early embryogenesis at least for a minority of promoters. In fact, miRNAs have been implicated in embryo patterning in other organisms, for example, *mir-31* and *mir-196a* have been implicated in the patterning of the *Drosophila* and the mouse embryo, respectively (Leaman et al 2005; Mansfield et al 2004).

#### *Independent transcription of intragenic miRNAs*

It has been hypothesized that intragenic miRNAs located in the sense orientation to host genes are under the control of the host gene promoter (Baskerville and Bartel 2005). Our work provides evidence that *C. elegans* intragenic miRNAs may be controlled by their own, rather than a host gene promoter, since sequences immediately upstream of intragenic miRNAs drive GFP expression *in vivo*. In contrast to *C. elegans* miRNAs, a large proportion of human miRNAs are located within introns (Rodriguez et al. 2004). A recent study identified the proximal promoters of 175 human miRNAs by combining nucleosome mapping with chromatin signatures for promoters (Ozsolak et al. 2008). This study showed that a short (~70-bp) nucleosome-depleted region is observed at the core promoter/initiation region of each miRNA. In agreement with



our findings, approximately one third of the intragenic miRNAs were seen to contain transcription initiation sites independent of their host gene's. Together, these findings suggest that intragenic miRNAs can be transcribed independently from the host gene promoter and that this phenomenon is likely evolutionarily conserved. It will be also important to determine how does the expression of intragenic miRNAs compare to the expression of the host gene.

#### *Conservation of miRNA expression patterns*

Many miRNA genes are highly conserved among diverse animal species. In a few cases it has been shown that miRNAs are not only conserved in terms of sequence but also in terms of their expression pattern. For instance, the *let-7* temporal expression appears to be conserved in a variety of animals, being expressed in later rather than early developmental stages (Pasquinelli et al. 2000; Pasquinelli et al. 2003; Thomson et al. 2006). The spatial expression of the muscle-specific *mir-1* and neuronal-specific *mir-124* appears also to be conserved during animal evolution (Lagos-Quintana et al. 2002; Darnell et al. 2006; Deo et al. 2006; Kloosterman et al. 2006). However, it is not likely that all conserved miRNA genes display conserved expression patterns. A recent study by Ason and colleagues, compared 100 miRNA expression patterns between medaka, zebrafish, chicken and mouse, with only a dozen of miRNAs common in all four species (Ason et al. 2006). This study revealed that while many miRNAs do exhibit a high degree of expression pattern conservation between species,

others exhibit great differences, with variation in miRNA expression being more pronounced the greater the differences in physiology. Differences between species often resulted mainly from changes in temporal rather than spatial miRNA expression. A complete and systematic analysis of miRNA expression pattern conservation will depend on the generation of comprehensive miRNA expression pattern datasets in diverse organisms. The *C. elegans* miRNA expression pattern resource described in this study constitutes a valuable tool to analyze miRNA expression pattern conservation and/or function in animal evolution. As an illustration of this point, a first glance comparison between the expression pattern observed in the Ason *et al.* study and the dataset we generated, already shows a few examples of miRNA expression pattern conservation between chicken, fish and nematodes. For instance, Ason and colleagues have shown that *mir-133* is expressed in muscle in chicken, medaka and zebrafish (Ason et al. 2007). We found that the *C. elegans* homolog of *mir-133*, *mir-245*, is expressed in pharyngeal muscle (Martinez et al. 2008b). This suggests an evolutionarily conserved function for *mir-133/mir-245* in muscle. Similarly, *mir-187* is expressed in neurons in both chicken and medaka, while its *C. elegans* homolog, *mir-76*, is expressed in numerous neurons (Ason et al. 2007; Martinez et al. 2008b).

### *Phenotypic assays*

A recent study in *C. elegans* revealed that most individual miRNAs do not confer a detectable phenotype when deleted (Miska et al. 2007). This high-throughput study focused on relatively rapid phenotypic assays to examine gross morphology, growth, development and behavior. It is then possible that miRNAs are involved in processes that have not yet been assayed. We can utilize the spatiotemporal expression data to design “custom-made” phenotypic assays for each individual miRNA, which can then be used toward understanding its cellular functions. To illustrate this point, take the example of *mir-232*. The promoter of *mir-232* is active exclusively in excretory cells and *mir-232* mutants do not show any phenotype in the standard assays (Martinez et al. 2008b; Miska et al. 2007). The excretory cell is shaped like the letter ‘H’ and forms two canals, one on each side of the worm, and functions to regulate the osmolarity of the organism (Buechner 2002). Mutants that affect the excretory cells usually display abnormalities in the lumen or diameter of the canal, frequently forming vacuoles most often only detected by electron microscopy. To determine if *mir-232* in fact plays a role in excretory cell development and/or physiology, we could assay *mir-232* mutants for tolerance to changes in osmolarity of the growth media and perform electron-dense microscopy to investigate the morphology of the canal and the potential presence of vacuoles.

#### *Combinatorial action of miRNAs*

The lack of phenotypes for individual miRNA deletions can also be explained by genetic interactions with other miRNAs. In this regard, we found that miRNA members belonging to families are more likely to be expressed in overlapping tissues than miRNAs that do not belong to the same family, suggesting a functional redundancy among miRNAs from the same family. Studying the extent to which miRNAs interact in a combinatorial fashion among themselves is an important step for further elucidating the functions of miRNAs. In the future, it will be interesting to investigate whether the miRNAs expressed in a given tissue, developmental stage or environmental condition, work in combination to regulate the expression of common sets of targets.

#### **Case study: FLH transcription factors function together to repress miRNA expression in the embryo**

Both large-scale and high-throughput datasets generated here can be used to derive specific biological hypotheses and in-depth studies, either for individual miRNAs or for TFs. The work presented in Chapter IV constitutes an example of the type of studies that can be derived using the resources we generated. Our Y1H network shows that transcription factors with a FLYWCH Zn-finger DNA-binding domain bind to the promoters of several miRNA genes. By using *promoter::gfp* transgenic strains, as well as additional assays, we found that FLYWCH transcription factors function redundantly to repress embryonic expression of *lin-4*, *mir-48*, and *mir-241*, miRNA genes that are normally

expressed only post-embryonically.

Using Y1H assays and computational tools, we show that FLH-1 and FLH-2 can recognize the same consensus binding site. This consensus binding site can be used to interrogate the *C. elegans* genome to find the full spectrum of (putative) targets. Future analyses, such as PBMs and ChiP, will be also important to determine the DNA binding specificities for FLH-1 and FLH-2 and whether both TFs can bind the same sites *in vivo*.

Interacting proteins often have similar expression patterns and often function in the same biological process (Schwikowski et al. 2000). In agreement with a functional redundancy between FLH proteins, we detected physical interactions between FLH-1, FLH-2 and FLH-3 using Y2H assays. Future analyses, such as co-immunoprecipitations, will be necessary to further confirm the interactions between *C. elegans* FLH proteins. In addition, it will be important to determine what is the functional significance of these protein-protein interactions *in vivo*, as well as possible interactions with other proteins.

Interestingly, we have also identified cross-regulation among the FLH proteins. We showed that the *flh-2* promoter contains a consensus binding site and is bound by FLH-1 in Y1H assays. We found that FLH-2 levels are up-regulated in *flh-1(bc374)* mutants, suggesting that FLH-1 normally represses the expression of *flh-2 in vivo*, likely in a direct manner. Preliminary results indicate that *flh-3* mRNA is also up-regulated in *flh-1(bc374)* mutants, suggesting that FLH-1 is also capable of repressing *flh-3* expression, albeit indirectly (data not

shown). In addition, we have also showed that *flh-1*, *flh-2* and *flh-3* are functionally redundant in repressing miRNA expression in the embryo. These results could reflect a role for FLH-1 in fine-tuning FLH-2 and FLH-3 protein levels in embryogenesis. Alternatively, they could reflect a mechanism by which FLH-1 is required to down-regulate FLH-2 and FLH-3 right before hatching into the L1 stage (Fig. V-3). Interestingly, two of FLH miRNA targets, *mir-48* and *mir-241*, are predicted to target FLH-1 3'UTR. In this model, FLH-1 would relieve miRNA repression by downregulating *flh-2* and *flh-3* expression and would be itself downregulated by its miRNA targets through a regulatory feedback loop (Fig. V-3). This model is agreement with the timing of *mir-48* and *mir-241* expression; both miRNAs start to be expressed in the L1 stage. Further experiments are necessary to corroborate this post-transcriptional interaction. In the future, it will be important to thoroughly study the extent of cross-regulation among all *flh* genes and to identify positive regulators that play a role in the activation of miRNA and *flh* gene expression.

### **Future challenges: Complete, dynamic and integrated networks**

The genes of an organism are differentially expressed through the activity of gene regulatory circuits that we are just beginning to uncover. Network representations of TF and miRNA interactions are undoubtedly powerful when they incorporate reliable, complete and unbiased data. To date, the computationally and/or experimentally mapped gene regulatory networks

available for most genomes are only a small representation of all the interactions that occur. Thus, continued efforts for the experimental mapping of transcriptional networks using a variety of complementary methods such as ChIP and Y1H assays are essential.

Importantly, the scope of regulatory network mapping may turn out to be greater than expected. For instance, miRNA sites that are not conserved or not seed-like, or miRNA sites outside 3'UTRs are usually not taken into account when generating post-transcriptional networks and their interactions may need to be considered. Similarly, as the complete spectrum of TFs in an organism of interest expands they need to be incorporated in regulatory networks. Finally, transcriptional networks have so far used binary information; *i.e.* a TF either regulates a target gene or it doesn't. Longer term, it will be important to include the full spectrum of binding specificities and affinities of all TFs.

Complex cellular and developmental processes depend in part on the precise spatiotemporal expression of genes, which is acquired by regulatory interactions that specifically occur at particular developmental times and/or in a tissue-specific manner. Most gene regulatory networks mapped to date consist of static depictions of all the possible interactions between TFs or miRNAs and their targets that can possibly occur *in vivo*. However, it will be important to integrate all available miRNA, TF and target expression patterns to limit the network to only those interactions that likely occur, and to expand the set of environmental and experimental conditions tested. Together, this information will lead to highly

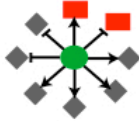



dynamic network models that can help us understand how gene expression relates to development and physiology.

It will also be crucial to integrate TF and miRNA-containing regulatory networks with other functional data, such as protein-protein interactions, on which many efforts have heavily focused, as well as other types of interactions that remain largely uncharacterized, such as those involving RNA-Binding Proteins (RBPs) (Fig. V-4). Such studies will provide insights into how TFs, miRNAs and RBPs together coordinate control of their targets, thereby affecting differential gene expression in a concerted fashion.

Longer term, it will also be important to generate “meta network models” in which different types of nodes and interactions are combined to reveal how signaling networks, regulatory gene expression networks and protein-protein interaction networks function to regulate biological processes that relate to development and homeostasis and how these networks are perturbed in disease (Fig. V-4).

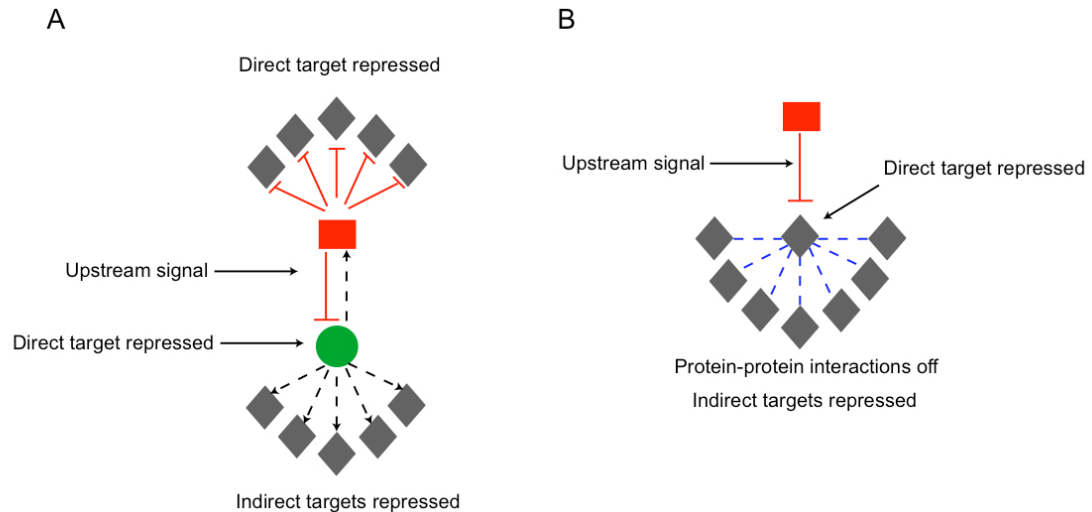


Figure V-1

Node	Degree	Example	Network	Clear hubs
TF	Out		Transcriptional	Yes
Target gene	In		Transcriptional	No
Target gene	In		Post-transcriptional	Yes
miRNA	Out		Post-transcriptional	No

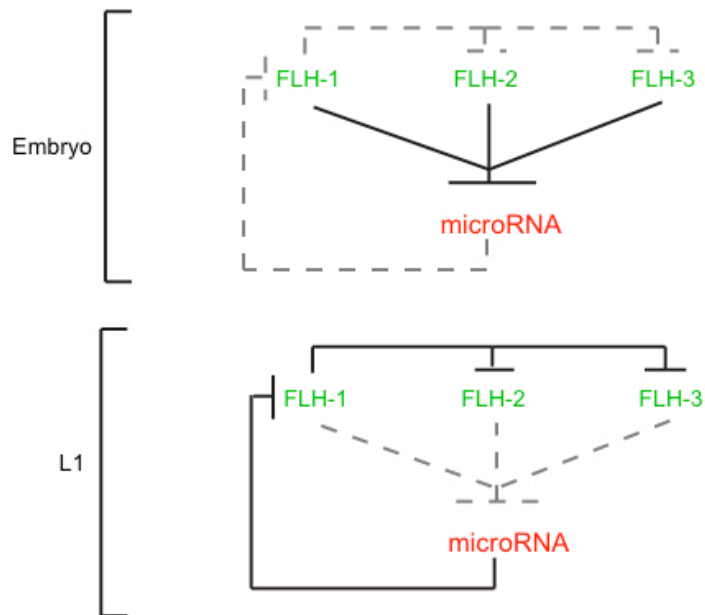
**Figure V-1.** Summary of presence of hub nodes in transcriptional and post-transcriptional networks. Nodes: green circles depict TFs; red rectangles depict miRNAs; grey diamonds represent other protein-coding genes. Edges: black arrows depict transcriptional activation; black blunted arrows depict transcriptional repression; red blunted arrows represent post-transcriptional repression.

Figure V-2



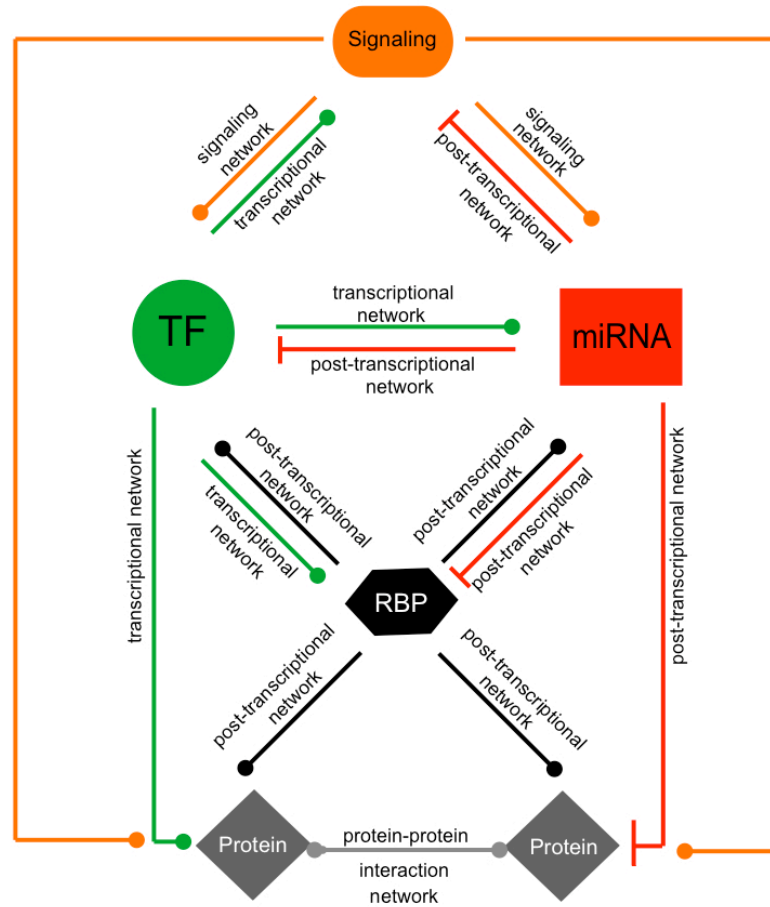
**Figure V-2.** Network circuits allow the spreading of regulatory effects. (A) TF and miRNAs that participate in feedback loops are highly connected and not only regulate each other but also each others' targets. In this example, an upstream signal activates the miRNA, which in turns represses all its direct targets, including the TF in the feedback loop. As a result, all downstream targets of the TF are also repressed (indirect targets of the miRNA). Red blunted arrows – post-transcriptional repression; black dashed arrows – inhibition of transcriptional activation; grey diamonds – protein-coding genes; red rectangle – miRNA; green circle – TF. (B) MiRNAs tend to target hubs in protein-protein interaction networks, hence spreading its regulatory effects to large set of proteins. In this example, an upstream signal activates the miRNA, which in turn represses the hub (direct target). As a result, all protein-protein interactions between the hub and other nodes (indirect targets) are inhibited. Red blunted arrow – post-transcriptional repression; blue dashed lines – inhibition of protein-protein interactions; red rectangle – miRNA; grey diamonds – protein-coding genes.

Figure V-3



**Figure V-3.** A putative FLH-miRNA regulatory circuit. In the embryonic stage, FLH proteins function redundantly to repress microRNA expression. In the late embryonic stage, FLH-1 represses the expression of *flh-2* and *flh-3*, relieving miRNA repression. Increasing levels of miRNAs then target FLH-1 3'UTR for repression upon hatching into the L1 stage.

Figure V-4



**Figure V-4.** *Integration of functional data into “meta network models”.* Metanetworks can be constructed by integrating transcriptional (red), post-transcriptional (black), signaling (orange), and protein-protein interactions networks (grey).

## REFERENCES

- Alon, U. 2007. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**: 450-461.
- Abbott, A. L., Alvarez-Saavedra, E., Miska, E. A., Lau, N. C., Bartel, D. P., Horvitz, H. R., and Ambros, V. 2005. The *let-7* microRNA family members *mir-48*, *mir-84* and *mir-241* function together to regulate developmental timing in *Caenorhabditis elegans*. *Dev Cell* **9**: 403-414.
- Aboobaker, A. A., Tomancak, P., Patel, N. H., Rubin, G. M., and Lai, E. C. 2005. Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proc Natl Acad Sci U S A* **102**: 18017-18022.
- Abrahante, J. E., Miller, E. A., and Rougvie, A. E. 1998. Identification of heterochronic mutants in *Caenorhabditis elegans*. Temporal misexpression of a collagen::green fluorescent protein fusion gene. *Genetics* **149**: 1335-1351.
- Adams, A., Gottschling, D. E., Kaiser, C. A., and Stearns, T. 1997. Methods in yeast genetics. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.
- Adams, M. D. et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185-2195.
- Ambros, V., and Horvitz, H. R. 1984. Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* **226**: 409-416.
- Ambros, V., and Horvitz, H. R. 1987. The *lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific postembryonic developmental events. *Genes Dev.* **1**: 398-414.
- Ambros, V. 2004. The functions of animal microRNAs. *Nature* **431**: 350-355.
- Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T., and Jewell, D. 2003. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol* **13**: 807-818.
- Ambros, V., and Lee, R. C. 2004. Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. *Methods Mol. Biol.* **265**: 131-58.
- Ao, W., Gaudet, J., Kent, W. J., Muttumu, S., and Mango, S. E. 2004. Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. *Science* **305**: 1743-1746.
- Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., and Teichmann, S. A. 2004. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14**: 283-291.
- Babu, M. M., Iyer, L. M., Balaji, S., and Aravind, L. 2006. The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* **34**: 6505-6520.
- Baek, D., Villen, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. 2008. The impact of microRNAs on protein output. *Nature* **455**: 64-71.

- Barabasi, A. L., and Oltvai, Z. N. 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**: 101-113
- Barrasa, M. I., Vaglio, P., Cavasino, F., Jacotot, L., and Walhout, A. J. M. 2007. EDGEDb: a transcription factor-DNA interaction database for the analysis of *C. elegans* differential gene expression. *BMC Genomics* **8**: 21.
- Bartel, D. P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.
- Bartel, D. P., and Chen, C. Z. 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet* **5**: 396-400.
- Baskerville, S., and Bartel, D. P. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**: 241-247.
- Baugh, L. R., and Sternberg, P. W. 2006. DAF-16/FOXO regulates transcription of *cki-1/Cip/Kip* and repression of *lin-4* during *C. elegans* L1 arrest. *Curr Biol* **16**: 780-785.
- Beaster-Jones, L., and Okkema, P. G. 2004. DNA binding and *in vivo* function of *C. elegans* PEB-1 require a conserved FLYWCH motif. *J. Mol. Biol.* **339**: 695-706.
- Becksei, A., and Serrano, L. 2000. Engineering stability in gene networks by autoregulation. *Nature* **405**: 590-593.
- Behlke, M., Dames, S. A., McDonald, W. H., Gould, K. L., Devor, E. J., and Walder, J. A. 2000. Use of high specific activity StarFire oligonucleotide probes to visualize low-abundance pre-mRNA splicing intermediates in *S. pombe*. *Biotechniques* **29**: 892-897.
- Berezikov, E., Bargmann, C. I., and Plasterk, R. H. 2004. Homologous gene targeting in *Caenorhabditis elegans* by biolistic transformation. *Nucleic Acids Res* **32**: e40.
- Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R. H., and Cuppen, E. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120**: 21-24.
- Berger, M. F., Badis, G., Gehrke, A. R., Talukder, S., Philippakis, A. A., Pena-Castillo, L., Alleyne, T. M., *et al.* 2008. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**: 1266-1276.
- Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V., and Hannon, G. J. 2003. Dicer is essential for mouse development. *Nat Genet* **35**: 215-217.
- Borneman, A. R., Leigh-Bell, J. A., Yu, H., Bertone, P., Gerstein, M., and Snyder, M. 2006. Target hub proteins serve as master regulators of development in yeast. *Genes Dev* **20**: 435-48.
- Boutros, M., Kiger, A. A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S. A., Consortium, H. F. A., Paro, R., and Perrimon, N. 2004. Genome-wide

- RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**: 832-835.
- Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**: 947-56.
- Bracht, J., Hunter, S., Eachus, R., Weeks, P. and Pasquinelli, A. E. 2004. Trans-splicing and polyadenylation of *let-7* microRNA primary transcripts. *RNA* **10**: 1586-1594.
- Brenner, S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77**: 71-94.
- Brennecke, J., Stark, A., Russell, R. B., and Cohen, S. M. 2005. Principles of microRNA-target recognition. *PLoS Biol* **3**: e85.
- Buchner, K., Roth, P., Schotta, G., Krauss, V., Saumweber, H., Reuter, G., and Dorn, R. 2000. Genetic and molecular complexity of the position effect variegation modifier *mod(mdg4)* in *Drosophila*. *Genetics* **155**: 141-157.
- Buechner, M. 2002. Tubes and the single *C. elegans* excretory cell. *Trends Cell Biol* **12**: 479-484.
- Burnett, W. V. 1997. Northern blotting of RNA denatured in glyoxal without buffer recirculation. *Biotechn* **22**: 668-671.
- Bushati, N., and Cohen, S. M. 2007. MicroRNA functions. *Annu. Rev. Cell Dev. Biol.* **23**: 175-205.
- Bryne, J. C., Valen, E., Tang, M. H., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., and Sandelin, A. 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res* **36**: D102-106.
- Chang, S., Johnston, R. J., Frokjaer-Jensen, C., Lockery, S., and Hobert, O. 2004. MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* **430**: 785-789.
- Chang, T. C., Yu, D., Lee, Y. S., Wentzel, E. A., Arking, D. E., West, K. M., Dang, C. V., Thomas-Tikhonenko, A., and Mendell, J. T. 2008. Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat Genet.* **40**: 43-50.
- Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., Barbisin, M., Xu, N. L., Mahuvakar, V. R., Andersen, M. R., Lao, K. Q., Livak, K. J., and Guegler, K. J. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* **33**: e179.
- Cobb, B. S., Nesterova, T. B., Thompson, E., Hertweck, A., O'Connor, E., Godwin, J., Wilson, C. B., Brockdorff, N., Fisher, A. G., Smale, S. T., and Merkenschlager, M. 2005. T cell lineage choice and differentiation in the absence of the RNaseIII enzyme Dicer. *J Exp Med* **201**: 1367-1373.
- Collas, P., and Dahl, J. A. 2008. Chop it, ChiP it, check it: the current status of chromatin immunoprecipitation. *Front biosci* **13**: 929-943.

- Coller, H. A., Forman, J. J., and Legesse-Miller, A. 2007. "Myc'ed messages": myc induces transcription of E2F1 while inhibiting its translation via a microRNA polycistron. *PLoS Genet.* **3**: e146.
- Conaco, C., Otto, S., Han, J. J., and Mandel, G. 2006. Reciprocal actions of REST and a microRNA promote neuronal identity. *Proc. Natl. Acad. Sci. USA* **103**: 2422-2427.
- The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799-816.
- Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520-562.
- The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**:2012-2018.
- Cui, Q., Yu, Z., Pan, Y., Purisima, E. O., and Wang, E. 2007. MicroRNAs preferentially target the genes with high transcriptional regulation complexity. *BBRC* **352**: 733-738.
- Deplancke, B., Dupuy, D., Vidal, M., and Walhout, A. J. M. 2004. A Gateway-compatible yeast one-hybrid system. *Genome Res* **14**: 2093-2101.
- Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A. M., Grove, C. A., Martinez, N. J., Sequerra, R., Doucette-Stam, L., Reece-Hoyes, J. S., Hope, .I. A., Tissenbaum, H. A., Mango, S. E., and Walhout, A. J. M. 2006a. A gene-centered *C. elegans* protein-DNA interaction network. *Cell* **125**: 1193-1205.
- Deplancke, B., Vermeirssen, V., Arda, H. E., Martinez, N. J., and Walhout, A. J. M. 2006b. Gateway-compatible yeast one-hybrid screens. *CSH Protocols* doi:10.1101/pdb.prot4590.
- Dorn, R., and Krauss, V. 2003. The modifier of *mdg4* locus in *Drosophila*: functional complexity is resolved by trans splicing. *Genetica* **117**: 165-77.
- Dupuy, D., Li, Q., Deplancke, B., Boxem, M., Hao, T., Lamesch, P., Sequerra, R., Bosak, S., Doucette-Stam, L., Hope, I. A., Hill, D., Walhout, A. J. M., and Vidal M. 2004. A first version of the *Caenorhabditis elegans* promoterome. *Genome Res* **14**: 2169-2175.
- Duursma, A. M., Kedde, M., Schrier, M., le Sage, C., and Agami, R. 2008. miR-148 targets human DNMT3b protein coding region. *RNA* **14**: 872-877.
- Edgley, M., D'Souza, A., Moulder, G., McKay, S., Shen, B., Gilchrist, E., Moerman, D., and Barstead, R. 2002. Improved detection of small deletions in complex pools of DNA. *Nucleic Acids Res* **30**: e52.
- Elnitski, L., Jin, V. X., Farnham, P. J., and Jones, S. J. M. 2006. Locating mammalian transcription factor binding sites: A survey of computational and experimental techniques. *Genome Res* **16**: 1455-1464.
- Elowitz, M. B., and Leibler, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**: 335-338.



- Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. 2003. MicroRNA targets in *Drosophila*. *Genome Biol* **5**: R1.
- Esquela-Kerscher, A., Johnson, S. M., Bai, L., Saito, K., Partridge, J., Reinert, K. L., and Slack, F. J. 2005. Post-embryonic expression of *C. elegans* microRNAs belonging to the *lin-4* and *let-7* families in the hypodermis and reproductive system. *Dev Dyn* **234**: 868-877.
- Farh, K. K., Grimson, A., Jan, C., Lewis, B. P., Johnston, W. K., Lim, L. P., Burge, C. B., and Bartel, D. P. 2005. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* **310**: 1817-21.
- Fazi, F., Rosa, A., Fatica, A., Gelmetti, V., De Marchis, M.L., Nervi, C., and Bozzoni, I. 2005. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis. *Cell* **123**: 819-831.
- Feinbaum, R., and Ambros, V. 1999. The timing of *lin-4* RNA accumulation controls the timing of postembryonic developmental events in *Caenorhabditis elegans*. *Dev Biol* **210**: 87-95.
- Fernandez, A. P., Gibbons, J., and Okkema, P. G. 2004. *C. elegans* *peb-1* mutants exhibit pleiotropic defects in molting, feeding, and morphology. *Dev Biol* **276**: 352-366.
- Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**: 102-114.
- Freeman, G. H., and Halton, J. H. 1951. Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika* **38**: 141-149.
- Gardner, T. S., Cantor, C. R., and Collins, J. J. 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**: 339-342.
- Gerasimova, T. I., Gdula, D. A., Gerasimov, D. V., Simonova, O., and Corces, V. G. 1995. A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell* **82**: 587-597.
- Ghosh, D., Gerasimova, T. I., and Corces, V. G. 2001. Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. *EMBO J* **20**: 2518-2527.
- Giraldez, A. J., Cinalli, R. M., Glasner, M. E., Enright, A. J., Thomson, J. M., Baskerville, S., Hammond, S. M., Bartel, D. P., and Schier, A. F. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**: 833-838.
- Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**: D140-144.
- Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., Degnan, B. M., Rokhsar, D. S., and Bartel, D. P. 2008. Early origins and

- evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**: 1193-1197.
- Grove, C. A., and Walhout, A. J. 2008. Transcription factor functionality and transcription regulatory networks. *Molecular BioSystems* **4**: 309-314.
- Hall, D. A., Zhu, H., Zhu, X., Royce, T., Gerstein, M., and Snyder, M. 2004. Regulation of gene expression by a metabolic enzyme. *Science* **306**: 482-484.
- Hammell M., Long D. Z., Zhang L., Lee A., Carmack C. S., Han M., Ding Y., and Ambros V. 2008. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat Methods* **5**: 813-819.
- Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., *et al.* 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**: 99-104.
- Harfe, B. D., McManus, M. T., Mansfield, J. H., Hornstein, E., and Tabin, C. J. 2005. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. *Proc Natl Acad Sci U S A* **102**: 10898-10903.
- Harris, K. S., Zhang, Z., McManus, M. T., Harfe, B. D., and Sun, X. 2006. Dicer function is essential for lung epithelium morphogenesis. *Proc Natl Acad Sci U S A* **103**: 2208-2213.
- Hertel, J., Lindemeyer, M., Missal, K., Fried, C., Tanzer, A., Flamm, C., Hofacker, I. L., Stadler, P. F., and 2005, S. o. B. C. L. a. 2006. The expansion of the metazoan microRNA repertoire. *BMC Genomics* **15**: 25.
- Hirata, H., Yoshiura, S., Ohtsuka, T., Bessho, Y., Harada, T., Yoshikawa, K., and Kageyama, R. 2002. Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. *Science* **298**: 840-843.
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. 1989. Site directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**: 51-59.
- Hollenhorst, P. C., Shah, A. A., Hopkins, C., and Graves, B. J. 2007. Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the *ETS* gene family. *Genes Dev* **21**: 1882-1894.
- Hornstein, E., and Shomron, N. 2006. Canalization of development by microRNAs. *Nat Genet* **38**: S20-24.
- Hristova, M., Birse, D., Hong, Y., and Ambros, V. 2005. The *Caenorhabditis elegans* LIN-14 is a novel transcription factor that controls the developmental timing of transcription from the insulin/insulin-like growth factor gene *ins-33* by direct DNA binding. *Mol Cell Bio* **25**: 11059-11072.
- Huang, P., Pleasance, E. D., Maydan, J. S., Hunt-Newbury, R., O'Neil, J. N, Mah, A., Baillie, D. L., Marra, M. A., Moerman, D. G., and Jones, S. J. 2007. Identification and analysis of internal promoters in *Caenorhabditis elegans* operons. *Genome Res* **17**: 1478-1485.

- Hunt-Newbury, R., Viveiros, R., Johnsen, R., Mah, A., Anastas, D., Fang, L., Halfnight, E., *et al.* 2007. High-throughput in vivo analysis of gene expression in *Caenorhabditis elegans*. *PLoS Biol* **5**: e237.
- Iitzkovitz, S., Milo, R., Kashtan, N., Ziv, G., and Alon, U. 2003. Subgraphs in random networks. *Phys Rev E* **2**: E68, 026127.
- Jeong, H., Mason, S. P., Barabasi, A. -L., and Oltvai, Z. N. 2001. Lethality and centrality in protein networks. *Nature* **411**: 41-42.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabasi, A.-L. 2000. The large-scale organization of metabolic networks. *Nature* **407**: 651-654.
- Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K. L., Brown, D., and Slack, F. 2005. RAS is regulated by the *let-7* microRNA family. *Cell* **120**: 635-647.
- Johnson, S. M., Lin, S. Y., and Slack, F. J. 2003. The time of appearance of the *C. elegans let-7* microRNA is transcriptionally controlled utilizing a temporal regulatory element in its promoter. *Dev Biol* **259**: 364-379.
- Johnston, R. J., Chang, S., Etchberger, J. F., Ortiz, C. O., and Hobert, O. 2005. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc Natl Acad Sci U S A* **102**: 12449-54.
- Johnston, R. J., and Hobert, O. 2003. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**: 845-849.
- Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., *et al.* 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**: 231-237.
- Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G., and Ahringer, J. 2001. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *C. elegans*. *Genome Biol* **2**: 1-10.
- Kashtan, N., Iitzkovitz, S., Milo, R., and Alon, U. 2004. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* **20**: 1746-58.
- Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigergiou, A. G., and Nishikura, K. 2007. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* **315**: 1137-1140.
- Kedde, M., Strasser, M. J., Boldajipour, B., Oude Vrielink, J. A., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P. M., van Duijse, J., Orom, U. A., *et al.* 2007. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131**: 1273-1286.
- Kim, J., Inoue, K., Ishii, J., Vanti, W. B., Voronov, S. V., Murchison, E., Hannon, G., and Abeliovich, A. 2007. A microRNA feedback circuit in midbrain dopamine neurons. *Science* **317**: 1220-1224.
- Kim, V. N. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* **6**: 376-385.
- Kim, V. N., and Nam, J. W. 2006. Genomics of microRNA. *Trends Genet* **22**: 165-173.

- Krauss, V., and Dorn, R. 2004. Evolution of the *trans*-splicing *Drosophila* locus *mod(mdg4)* in several species of Diptera and Lepidoptera. *Gene* **331**:165-176.
- Krek, A., Grun, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., MacMenamin, P., da Piedade, I., Gunsalus, K. C., Stoffel, M., and Rajewski, N. 2005. Combinatorial microRNA target predictions. *Nat Genet* **37**: 495-500.
- Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K. G., Tuschl, T., Manoharan, M., and Stoffel, M. 2005. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**: 685-689.
- Kummerfeld, S. K., and Teichmann, S. A. 2006. DBD: a transcription factor prediction database. *Nucleic Acids Res* **34**: D74-D81.
- Labouesse, M., Sookhareea, S., and Horvitz, H. R. 1994. The *Caenorhabditis elegans* gene *lin-26* is required to specify the fates of hypodermal cells and encodes a presumptive zinc-finger transcription factor. *Development* **120**: 2359-68.
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853-858.
- Lai, E. C., Tomancak, P., Williams, R. W., and Rubin, G. M. 2003. Computational identification of *Drosophila* microRNA genes. *Genome Biol* **4**: R42.
- Lall, S., Grun, D., Krek, A., Chen, K., Wang, Y. L., Dewey, C. N., Sood, P., Colombo, T., Bray, N., *et al.* 2006. A Genome-wide map of conserved microRNA targets in *C. elegans*. *Curr Biol* **16**: 460-471.
- Landgraf, P., M. Rusu, R. Sheridan, A. Sewer, N. Iovino, A. Aravin, S. Pfeffer, A. Rice, A. O. *et al.* 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**: 1401-1414.
- Landmann, F., Quintin, S., and Labouesse, M. 2004. Multiple regulatory elements with spatially and temporally distinct activities control the expression of the epithelial differentiation gene *lin-26* in *C. elegans*. *Dev Biol* **265**: 478-490.
- Latchman, D. S. 1998. Eukaryotic transcription factors. (San Diego, Academic Press).
- Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858-862.
- Lee, R. C., Feinbaum, R. L., and Ambros, V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843-854.
- Lee, R. C., and Ambros, V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**: 862-864.
- Lee, T. I., and Young, R. A. 2000. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* **34**: 77-137.

- Lee, Y., Kim, M., Han, J., Yeom, K. -H., Lee, S., Baek, S. H., and Kim, V. N. 2004. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**: 4051-4060.
- Levine, M., and Tjian, R. 2003. Transcription regulation and animal diversity. *Nature* **424**: 147-151.
- Lewis, B. P., Burge, C. B., and Bartel, D. P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15-20.
- Li, M., Jones-Rhoades, M. W., Lau, N. C., Bartel, D. P., and Rougvie, A. E. 2005. Regulatory mutations of *mir-48*, a *C. elegans let-7* family microRNA, cause developmental timing defects. *Dev Cell* **9**: 415-422.
- Liang, H., and Li, W. H. 2007. MicroRNA regulation of human protein-protein interaction network. *RNA* **13**: 1402-1408.
- Lillycrop, K. A. and Latchman, D. S. 1992. Alternative splicing of the Oct-2 transcription factor RNA is differentially regulated in neuronal cells and B cells and results in protein isoforms with opposite effects on the activity of octamer/TAATGARAT-containing promoters. *J Biol Chem* **267**: 24960-24965.
- Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta S., Rhoades, M. W., Burge, C. B., and Bartel, D. P. 2003. The microRNAs of *Caenorhabditis elegans*. *Genes Dev* **17**: 991-1008.
- Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S., and Johnson, J. M. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**: 769-773.
- Liu, J., Wilson, T. E., Milbrandt, J., and Johnston, M. 1993. *Methods Enzymol* **6**: 1-13.
- Liu, Z., Kirch, S., and Ambros, V. 1995. The *Caenorhabditis elegans* heterochronic gene pathway controls stage-specific transcription of collagen genes. *Development* **121**: 2471-2478.
- Livak, K. J., and Schmittgen, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2[-Delta Delta C(T)] Method. *Methods* **25**: 402-408.
- Lytle, J. R., Yario, T. A., and Steitz, J. A. 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* **104**: 9667-9672.
- Maduro, M., and Pilgrim, D. 1995. Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. *Genetics* **141**: 977-988.
- Mangan, S., and Alon, U. 2003. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A* **100**: 11980-11985.
- Mango, S. E., Lambie, E. J., and Kimble, J. 1994. The *pha-4* gene is required to generate the pharyngeal primordium of *Caenorhabditis elegans*. *Development* **120**: 3019-3031.

- Maniatis, T., and Tasic, B. 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**: 236-243.
- Martinez, N. J., Ow, M. C., Barrasa, M. I., Hammell, M., Sequerra, R., Doucette-Stamm, L., Roth, F. P., Ambros, V., and Walhout, A. J. M. 2008a. A genome-scale microRNA regulatory network in *C. elegans* reveals composite feedback motifs that provide high information flow. *Genes Dev* **22**: 2535-2549.
- Martinez, N. J., Ow, M. C., Reece-Hoyes, J., Ambros, V., and Walhout, A. J. 2008b. Genome-scale spatiotemporal *Caenorhabditis elegans* microRNA promoter activity. *Genome Res* **18**: 2005-2015.
- Marson, A., Levine, S. S., Cole, M. F., Frampton, G. M., Brambrink, T., Johnstone, S., Guenther, M. G., Johnston, W. K., Wernig, M., Newman, J., et al. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**: 521-533.
- Mattick, J. S. 2004. RNA regulation: a new genetics? *Nat Rev Genet* **5**: 316-323.
- Mazumder, B., Seshadri, V., and Fox, P. L. 2003. Translational control by the 3' UTR: the ends specify the means. *Trends Biochem Sci* **28**: 91-98.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. 2002. Network motifs: simple building blocks of complex networks. *Science* **298**: 824-827.
- Miska, E. A., Alvarez-Saavedra, E., Abbott, A. L., Lau, N. C., Hellman, A. B., McGonagle, S. M., Bartel, D. P., Ambros, V. R., and Horvitz, H. R. 2007. Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet* **3**: e215.
- Muljo, S. A., Ansel, K. M., Kanellopoulou, C., Livingston, D. M., Rao, A., and Rajewski, K. 2005. Aberrant T cell differentiation in the absence of Dicer. *J Exp Med* **202**: 261-269.
- Nagai, T., Iyata, K., Park, E. S., Kubota, M., Mikoshiba, K., and Miyawaki, A. 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* **20**: 87-90.
- Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H., and Wolfe, S. A. 2008. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133**: 1277-1289.
- Obernosterer, G., Leuschner, P. J., Alenius, M., and Martinez, J. 2006. Post-transcriptional regulation of microRNA expression. *RNA* **12**: 1161-1167.
- O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., and Mendell, J. T. 2005. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**: 839-843.
- Ow, M. C., Martinez, N. J., Olsen, P., Silverman, S., Barrasa, M. I., Conradt, B., Walhout, A. J. M., and Ambros, V. R. 2008. The FLYWCH transcription factors FLH-1, FLH-2 and FLH-3 repress embryonic expression of microRNA genes in *C. elegans*. *Genes Dev* **22**: 2520-2534.

- Patterson, G. I., Koweeck, A., Wong, A., Liu, Y., and Ruvkun, G. 1997. The DAF-3 Smad protein antagonizes TGF-B-related receptor signaling in the *Caenorhabditis elegans* dauer pathway. *Genes Dev* **11**: 2679-2690.
- Praitis, V., Casey, E., Collar, D., and Austin, J. 2001. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157**: 1217-1226.
- Praitis, V. 2006. Creation of transgenic lines using microparticle bombardment methods. *Methods Mol. Biol.* **351**: 93-107.
- Rao, P. K., Kumar, R. M., Farkhondeh, M., Baskerville, S., and Lodish, H. F. 2006. Myogenic factors that regulate expression of muscle-specific microRNAs. *Proc. Natl. Acad. Sci. USA* **103**: 8721-8726.
- Rajewski, N. 2006. microRNA target prediction in animals. *Nat Genet* **38**: S8-13.
- Reece-Hoyes, J. S., Deplancke, B., Shingles, J., Grove, C. A., Hope, I. A., and Walhout, A. J. M. 2005. A compendium of *C. elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol* **6**: R110.
- Reece-Hoyes, J. S., Shingles, J., Dupuy, D., Grove, C. A., Walhout, A. J., Vidal, M., and Hope, I. A. 2007. Insight into transcription factor gene duplication from *Caenorhabditis elegans* Promoterome-driven expression patterns. *BMC Genomics* **8**: 27.
- Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. 2004. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**: 1507-17.
- Retelska, D., Iseli, C., Bucher, P., Jongeneel, C. V., and Naef, F. 2006. Similarities and differences of polyadenylation signals in human and fly. *BMC Genomics* **12**: 176.
- Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D. P. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**: 1193-1207.
- Ruby, J. G., Stark, A., Johnston, W. K., Kellis, M., Bartel, D. P., and Lai, E. C. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* **17**: 1850-1864.
- Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901-906.
- Rodriguez, A., Griffith-Jones, S., Ashurst, J. L., and Bradley, A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**: 1902-1910.
- Rodriguez-Gabriel, M. A., Burns, G., McDonald, W. H., Martin, V., Yates, J. R., Bahler, J., and Russell, P. 2003. RNA-binding protein Csx1 mediates global control of gene expression in response to oxidative stress. *Embo J* **22**: 6256-6266.

- Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E. E. 2007. A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev* **21**: 436-49.
- Schier, A. F., and Giraldez, A. J. 2006. MicroRNA function and mechanism: insights from zebra fish. *Cold Spring Harb Symp Quant Biol* **71**: 195-203.
- Schwikowski, B., Uetz, P., and Fields, S. 2000. A network of protein-protein interactions in yeast. *Nat Biotechnol* **18**: 1257-61.
- Selbach, M., Schwanhauser, B., Thierfelder, N., Fang, Z., Khanin, R. and Rajewsky, N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**: 58-63.
- Sethupathy, P., Megraw, M., and Hatzigergiou, A. G. 2006. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat methods* **3**: 881-886.
- Shalgi, R., Lieber, D., Oren, M., and Pilpel, Y. 2007. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Computational Biology* **3**: e131.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498-504.
- Sheffield, P., Garrard, S., and Derewenda, Z. 1999. Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. *Protein Expr Purif* **15**: 34-39.
- Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* **31**: 64-8.
- Simmer, F., Tijsterman, M., Parrish, S., Koushika, S. P., Nonet, M. L., Fire, A., Ahringer, J., and Plasterk, R. H. A. 2002. Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr Biol* **12**: 1317-1319.
- Simon, D. J., Madison, J. M., Conery, A. L., Thompson-Peer, K. L., Soskis, M., Ruvkun, G. B., Kaplan, J. M., and Kim, J. K. 2008. The microRNA miR-1 regulates a MEF-2-dependent retrograde signal at neuromuscular junctions. *Cell* **133**: 891-902.
- Sood, P., Krek, A., Zavolan, M., Macino, G., and Rajewsky, N. 2006. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc Natl Acad Sci U S A* **103**: 2746-51.
- Sokol, N. S., and Ambros, V. 2005. Mesodermally expressed *Drosophila* microRNA-1 is regulated by Twist and is required in muscles during larval growth. *Genes Dev* **19**: 2343-2354.
- Stark, A., Brennecke, J., Bushati, N., Russell, R. B., and Cohen, S. M. 2005. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* **123**: 1133-46.
- Stefani, G. and Slack, F. J. 2008. Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol* **9**: 219-230.



- Sulston, J. E., and Hodgkins, J. 1998. In *The nematode Caenorhabditis elegans*, Methods, ed Wood, W. B., et. al. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). In, pp 587-606.
- Sulston, J. E., and Horvitz, H. R. 1977. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol* **56**: 110-156.
- Sylvestre, Y., De Guire, V., Querido, E., Mukhopadhyay, U. K., Bourdeau, V., Major, F., Ferbeyre, G., and Chartrand, P. 2007. An E2F/miR-20a autoregulatory feedback loop. *J Biol Chem* **282**: 2135-43.
- Tay, Y., Zhang, J., Thomson, A. M., Lim, B., and Rigoutsos, I. 2008. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **445**: 1124-1128.
- Thatcher, J. D., Haun, C., and Okkema, P. G. 1999. The DAF-3 Smad binds DNA and represses gene expression in the *Caenorhabditis elegans* pharynx. *Development* **126**: 97-107.
- Thatcher, J. D., Fernandez, A. P., Beaster-Jones, L., Haun, C., and Okkema P. G. 2001. The *Caenorhabditis elegans* *peb-1* gene encodes a novel DNA-binding protein involved in morphogenesis of the pharynx, vulva, and hindgut. *Dev Biol* **229**: 480-493.
- Thomson, J. M., M. Newman, J. S. Parker, E. M. Morin-Kensicki, T. Wright, and S. M. Hammond. 2006. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev* **20**: 2202-2207.
- Tsang, J., Zhu, J., and van Oudenaarden, A. 2007. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol Cell* **26**: 753-767.
- Varghese, J., and Cohen, S. M. 2007. microRNA miR-14 acts to modulate a positive autoregulatory loop controlling steroid hormone signaling in *Drosophila*. *Genes Dev* **21**: 2277-2282.
- Vermeirssen, V., Barrasa, M. I., Hidalgo, C., Babon, J. A. B., Sequerra, R., Doucette-Stam, L., Barabasi, A. L., and Walhout, A. J. M. 2007a. Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network. *Genome Res* **17**: 1061-1071.
- Vermeirssen, V., Deplancke, B., Barrasa, M. I., Reece-Hoyes, J. S., Arda, H. E., Grove, C.A., Martinez, N. J., Sequerra, R., Doucette-Stamm, L., Brent, M., and Walhout, A. J. M. 2007b. Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping. *Nat Methods* **4**: 659-664.
- Viswanathan, S. R., Daley, G. Q., and Gregory, R. I. 2008. Selective blockade of microRNA processing by Lin28. *Science* **320**: 97-100.
- Vowels, J. J., and Thomas, J. H. 1994. Multiple chemosensory defects in *daf-11* and *daf-21* mutants of *Caenorhabditis elegans*. *Genetics* **138**: 303-16.
- Walhout, A. J. M. 2006. Unraveling Transcription Regulatory Networks by Protein-DNA and Protein-Protein Interaction Mapping. *Genome Res* **16**: 1445-1454.

- Walhout, A. J. M., Sordella, R., Lu, X., Hartley, J. L., Temple, G. F., Brasch, M. A., Thierry-Mieg, N., and Vidal, M. 2000a. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**: 116-122.
- Walhout, A. J. M., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S., and Vidal, M. 2000b. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods in enzymology: "Chimeric genes and proteins"* **328**: 575-592.
- Walhout, A. J., Reboul, J., Shtanko, O., Bertin, N., Vaglio, P., Ge, H., Lee, H., Doucette-Stamm, L., Gunsalus, K. C., Schetter, A. J., Morton, D. G., Kempfues, K. J., Reinke, V., Kim, S. K., Piano, F., and Vidal, M. 2002. Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr Biol* **12**: 1952-1958.
- Wang, J. and Kim, S. K. 2003. Global analysis of dauer gene expression in *C. elegans*. *Development* **130**: 1621-1634.
- Wang, X., Gu, J., Zhang, M. Q., and Li, Y. 2008. Identification of phylogenetically conserved microRNA cis-regulatory elements across 12 Drosophila species. *Bioinformatics* **24**: 165-171.
- Wienholds, E., Kloosterman, W. P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., Horvitz, H. R., Kauppinen, S., and Plasterk, R. H. 2005. MicroRNA expression in zebrafish embryonic development. *Science* **309**: 310-311.
- Wienholds, E., Koudijs, M. J., van Eeden, F. J., Cuppen, E., and Plasterk, R. H. 2003. The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat Genet* **35**: 217-218.
- Wightman, B., Ha, I., and Ruvkun, G. 1993. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855-862.
- Williams, B. D., Schrank, B., Huynh, C., Shownkeen, R., and Waterston, R. H. 1992. A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. *Genetics* **131**: 609-624.
- Wilson, D., Charoensawan, V., Kummerfeld, S. K., and Teichmann, S. A. 2008. DBD-taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res* **36**: D88-92.
- Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Lieblich, I., Krull, M., Matys, V., Michael, H., Ohnhauser, R., et al. 2001. The TRANSFAC system on gene expression regulation. *Nucleic Acids Res* **29**: 281-283.
- Wolfe, S. A., Neklodova, L., and Pabo, C. O. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* **29**: 183-212.
- Wood, W. B. 1988. *The nematode Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor.

- Woods, K., Thomson, J. M., and Hammond, S. M. 2007. Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors. *J Biol Chem* **282**: 2130-2134.
- Wulczyn, F. G., Smirnova, L., Rybak, A., Brandt, C., Kwidzinski, E., Ninnemann, O., Strehle, M., Seiler, A., Schumacher, S., and Nitsch, R. 2007. Post-transcriptional regulation of the *let-7* microRNA during neural cell specification. *FASEB J* **21**: 415-426.
- Xu, H., Wang, X., Du, X., and Li, N. 2006. Identification of microRNAs from different tissues of chicken embryo and adult chicken. *FEBS Lett* **580**: 3610-3616.
- Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R. Y., Alon, U., and Margalit, H. 2004. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc Natl Acad Sci U S A* **101**: 5934-9.
- Yoo, A. S. and Greenwald. I. 2005. LIN-12/Notch activation leads to microRNA-mediated downregulation of *Vav* in *C. elegans*. *Science* **310**: 1330-1333.
- Zhao, Y., Ransom, J. F., Li, A., Vedantham, V., von Drehle, M., Muth A. N., Tsuchihashi, T., McManus, M. T., Schwartz, R. J., and Srivastava, D. 2007. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* **129**: 303-317.