Universidade de Lisboa
Faculdade de Medicina de Lisboa

# The RNA 5' Phosphatase PIR-1 Cooperates with Dicer to Produce Endogenous Small RNAs and Suppress Viral Replication in *C. elegans*

Daniel Marques de Almeida de Melo Chaves

Doutoramento em Ciências Biomédicas

Especialidade em Ciências Biopatológicas

2015

Universidade de Lisboa
Faculdade de Medicina de Lisboa

# The RNA 5' Phosphatase PIR-1 Cooperates with Dicer to Produce Endogenous Small RNAs and Suppress Viral Replication in *C. elegans*

Daniel Marques de Almeida de Melo Chaves

Orientadores:

- Professora Doutora Maria do Carmo Salazar Velez Roque da Fonseca;
- Professor Doutor Craig Cameron Mello.

Tese especialmente elaborada para obtenção do grau de Doutor em Ciências Biomédicas, especialidade em Ciências Biopatológicas.

Júri:

Presidente:

- Doutor José Augusto Gamito Melo Cristino, Professor Catedrático e Presidente do Conselho Científico da Faculdade de Medicina da Universidade de Lisboa

Vogais:

- Doutor Craig Cameron Mello, Distinguished Professor da University of Massachusetts Medical School;
- Doutor Rui Gonçalo Viegas Russo da Conceição Martinho, Professor Auxiliar da Universidade do Algarve;
- Doutora Cecília Maria Pais de Faria de Andrade Arraiano, Investigadora Coordenadora do Instituto de Tecnologia Química e Biológica da Universidade Nova de Lisboa;
- Doutora Maria do Carmo Salazar Velez Roque da Fonseca, Professora Catedrática da Faculdade de Medicina da Universidade de Lisboa;
- Doutor Bruno Miguel de Carvalho e Silva Santos, Professor Associado com Agregação da Faculdade de Medicina da Universidade de Lisboa;
- Doutor João António Augusto Ferreira, Professor Associado da Faculdade de Medicina da Universidade de Lisboa.

Este doutoramento foi financiado pela Fundação para a Ciência e a Tecnologia do Ministério da Ciência, Tecnologia e Ensino Superior do Governo de Portugal (Bolsa de Doutoramento com a referência SFRH/BD/17629/2004/H6BM), e por fundos de investigação atribuídos ao Professor Doutor Craig Mello.

2015

Todas as afirmações efectuadas no presente documento são da exclusiva responsabilidade do seu autor, não cabendo qualquer responsabilidade à Faculdade de Medicina de Lisboa pelos conteúdos nele apresentados.

Dedico este trabalho

à minha mãe Noémia, ao meu pai José e à minha irmã Susana.

# ACKNOWLEDGEMENTS

Throughout the (undeniably) long years that took me to reach the completion of this work, I have had the opportunity and pleasure to cross paths with many wonderful scientists, colleagues and friends. To those whose names I will not list, rest assured that your contribution to my development as both a scientist and a human being will never be forgotten!

I would like to start by thanking my supervisor Craig Mello at the University of Massachusetts Medical School for making me a part of his team and giving me the freedom and resources to carry out this project. I have learned immensely from working in his laboratory.

To my supervisor Maria Carmo-Fonseca at the Faculty of Medicine from the University of Lisbon I also extend my gratitude. My scientific life began in her laboratory, and my desire to keep doing research is due, in no small part, to that first, very positive contact with the scientific world. Additionally, Carmo played an active and decisive role in ensuring that I found a laboratory that suited my scientific interests.

I would also like to thank other scientists who accompanied my work as part of my PhD committee at the IMM (Instituto de Medicina Molecular) in Lisbon. They are António Jacinto, Luís Moita and Francisco Enguita.

I thank Victor Ambros for valuable feedback and encouragement at various stages of my work both inside and outside of the monthly joint meetings between his laboratory and ours.

I am indebted to several members of the laboratory – past and present – for their help, friendship and support. I have learned something from them all.

To Weifeng Gu, now heading his own laboratory at the University of California Riverside, I am especially indebted for very generously joining forces with me to tackle PIR-1. Without his contribution the project would not have reached its current state. From him I also learned (as did the rest of the team) the tricky art of small RNA cloning and beyond.

Thomas Duchaine, with whom I initiated my work, introduced me to the ABC of *C. elegans* and to many experimental techniques that every molecular biologist should master. To Thomas, who initiated the study of *C. elegans* PIR-1, I also owe the privilege of working with this incredibly interesting protein, with which I maintained an intense love/hate relationship over the years.

Yanxia Bei, who also worked with PIR-1, Erbay Yigit and Chun-Chieh Chen taught me the basics of *C. elegans* genetics and small RNA techniques and were my favorite discussion "buddies" during the initial phase of my work.

Hsin-Yue Tsai, who came along a bit later, has been a constant and invaluable source of help, discussion and support. Hsin-Yue and I also performed several experiments together.

Other members of the laboratory with whom I have always been able to count on for productive discussions were Pedro Batista, Colin Conine, Elaine Youngman, Sandra Vergara and Wen Tang. Another big thank you goes out to Jean O'Connor, our administrator, for her invaluable support in all things bureaucratic.

My collaborator James Moresco at the Scripps Research Institute in California, who performed all the proteomics work, provided key results to our study. Thank you for the great work!

I thank my collaborators at UMass who have started work on orthologs of PIR-1: Keith Boundy from the laboratory of Phillip Zamore for his work on *Drosophila*, and Chris MacKay from the laboratory of Robert Finberg for his work with PIR1 knockout mice.

Regarding the writing of this thesis I would like to thank Sandra Vergara, Gina Caldas and Elena Merino-Rodríguez for critical reading and comments on some of the sections. To James Birtley, who has patiently read and commented on the whole document until the very last minute, I grant a special thesis reviewing merit award. You have all greatly helped to improve these pages!

Various experiments described here would not have been possible without the kind sharing of strains and reagents from various sources. I thank Shohei Mitani, Scott Kennedy, Valerie Reinke, David Wang, Victor Ambros and the *Caenorhabditis* Genetics Center for *C. elegans* strains. For reagents, I am grateful to Witold Filipowicz, Thomas Duchaine, Victor Ambros and Stephen Buratowski.

This work would not have been possible without the financial support from an FCT pre-doctoral fellowship for the first four years of my research (Fundação para a Ciência e a Tecnologia, the scientific grant agency of the Portuguese Government; fellowship reference SFRH/BD/17629/2004/H6BM). During the post-fellowship years I was generously supported by Craig Mello.

I want to extend my deep gratitude to all my friends, old and new, in and out of the lab, who have helped me to reach this stage in one piece. Fortunately, they are too many to mention!

Living in this area of the United States also afforded me the great opportunity to enjoy the company of several of my numerous relatives, who moved here from the Azores during the 1970s in search of better life prospects. A very special thank you goes to my aunt Jorgeana Boavida, for the support that I knew I could always rely on.

To Mário Ramos, who accompanied me from afar for several years during this endeavor, I also want to express my warmest gratitude for his loving support in difficult times.

Without James Birtley, the last phase of my PhD would have been nearly impossible to complete. With his energy, enthusiasm, support and love, he gave me the balance I needed to persevere and keep moving in the right direction. I am forever indebted to him and look forward to retribute in the best way possible.

Lastly, I am who I am today because of my fantastic parents. From them I have always received nothing but unconditional love and support. They have given me the freedom to be myself. They have given me everything they could to help me meet my aspirations. My great little sister has also always been there for me. I hope to some day compensate her for the long years we have spent apart. I love you all very much!

Obrigado!!!

# ABSTRACT

Most organisms utilize small RNAs (sRNA) to control diverse aspects of development, reproduction and physiology by regulating gene expression at the transcriptional and post-transcriptional levels. The essential ribonuclease Dicer is a key enzyme in the production of several types of sRNAs. Prior work on the nematode *Caenorhabditis elegans* (*C. elegans*) has uncovered PIR-1 – a small protein conserved in all metazoans – as an interacting partner of Dicer *in vivo*. The human ortholog of PIR-1 has RNA 5' tri- and diphosphatase activities *in vitro*, but its biological role remains unclear. With the intent of finding its function, we characterized various aspects of *C. elegans* PIR-1. We found this enzyme to be essential for general growth and development, germline proliferation, and sperm maturation. We confirmed that PIR-1 associates with Dicer *in vivo* and expanded its repertoire of known interactions. Profiling of sRNAs from *pir-1* loss-of-function animals by high-throughput sequencing revealed that PIR-1 is required for the production of 26G-RNAs during spermatogenesis, a class of Dicer-dependent sRNAs. 26G-RNAs are essential to promote appropriate sperm development, in agreement with the *pir-1* mutant sperm defect. Additionally, we discovered a second, 26G-RNA-independent role for PIR-1, in which it cooperates with Dicer and other canonical RNA interference (RNAi) pathway components to suppress the replication of the *C. elegans* Orsay RNA virus. By demonstrating that PIR-1 functions as its human counterpart *in vitro*, and that a *pir-1* transgene with a mutated phosphatase active site cannot rescue any of the mutant defects, we concluded that PIR-1 acts as an RNA phosphatase *in vivo*. This is the first study in which concrete biological functions are assigned to this enzyme. Given its high degree of conservation, these results provide a solid basis for studies on the multiple functions of PIR-1 in more complex animals.

**Keywords:** *C. elegans*; small RNA; Dicer; RNA 5' phosphatase; RNAi; virus.

# RESUMO

A maioria dos organismos utiliza pequenos RNAs (sRNA) para controlar diversos aspectos do seu desenvolvimento, reprodução e fisiologia, através da regulação da expressão génica ao nível transcricional e pós-transcricional. A ribonuclease essencial Dicer desempenha um papel central na produção de vários tipos de sRNAs. Um estudo anterior realizado no nemátode *Caenorhabditis elegans* (*C. elegans*) revelou que PIR-1 – uma pequena proteína conservada em todos os metazoários – se associa à Dicer *in vivo*. A proteína ortóloga humana PIR1 funciona *in vitro* como uma tri- e di-fosfatase 5' de RNA. A sua função biológica, porém, não é clara. Com o objectivo de encontrar a sua função, procedemos à caracterização da PIR-1 de *C. elegans*. Esta enzima é essencial para o desenvolvimento geral e crescimento, proliferação da linha germinal e para a maturação de espermatozóides. Neste estudo comprovámos a interacção entre PIR-1 e Dicer *in vivo*, e expandimos o repertório de proteínas com as quais a PIR-1 se associa. Sequenciação "high-throughput" dos sRNAs de um mutante de *pir-1*, revelou que esta proteína é necessária, juntamente com a Dicer, para a produção de 26G-RNAs, que promovem a espermatogénese. Adicionalmente, descobrimos que em cooperação com a Dicer e outros componentes da via da interferência por RNA (RNAi), a PIR-1 suprime a replicação do vírus Orsay, que infecta especificamente *C. elegans*. Ao demonstrar que a PIR-1 possui a mesma actividade que a proteína humana *in vitro*, e que um transgene de *pir-1* com o centro catalítico inactivo não permite a supressão de nenhum dos defeitos do mutante, concluímos que a PIR-1 funciona como uma fosfatase de RNA *in vivo*. Este é o primeiro estudo a atribuir funções biológicas a esta enzima. Dada a sua conservação, estes resultados formam uma base sólida para estudos futuros sobre as múltiplas funções da fosfatase PIR-1 em animais de maior complexidade.

**Palavras chave:** *C. elegans*; small RNA; Dicer; RNA 5' phosphatase; RNAi; virus.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **aa** | amino acid |
| **AGO** | Argonaute |
| **ALG** | Argonaute (plant)-like gene |
| **ATG** | start codon |
| **ATP** | adenosine triphosphate |
| **Avr3x** | ivermectin-resistant genetic triple mutant |
| **BCP** | 1-bromo-3-chloropropane |
| **BLAST** | basic local alignment search tool |
| **bp** | base pair |
| **Cap-seq** | cloning and deep-sequencing of capped RNA fragments |
| **CARD** | caspase activation and recruitment domain |
| **cDNA** | complementary DNA |
| **ChIP** | chromatin immunoprecipitation |
| **CIP** | calf intestinal phosphatase |
| **CLIP** | cross-linking and immunoprecipitation |
| **CLIP-seq** | CLIP followed by cloning and deep-sequencing of RNAs |
| **csRNA** | capped short RNA |
| **CTD** | C-terminal domain of Pol II |
| **DAPI** | 4',6-diamidino-2-phenylindole |
| **DIC** | differential interference contrast microscopy |
| **DMSO** | dimethyl sulfoxide |
| **DNA** | deoxyribonucleic acid |
| **DNase** | deoxyribonuclease |
| **dNTP** | deoxynucleotide triphosphate |
| **DRB** | 5,6-dichloro-1-β-D-ribobenzimidazole |
| **DRH** | Dicer-related helicase |
| **dsDNA/RNA** | double-stranded DNA/RNA |
| **DTC** | distal tip cell |
| **DTT** | dithiothreitol |
| **DUSP** | dual-specificity protein phosphatase |
| **EDTA** | (ethylenedinitrilo)tetra-acetic acid |
| **EGTA** | ethylene-bis(oxyethylenenitrilo)tetra-acetic acid |
| **Eri** | enhancer of RNAi |
| **FHV** | Flock House virus |
| **GAPDH** | glyceraldehyde 3-phosphate dehydrogenase |
| **GFP** | green fluorescent protein |
| **H3K9** | lysine at position 9 of histone 3 |
| **HEPES** | N-(2-hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid) |
| **hnRNP** | heterogenous ribonucleoprotein |
| **HRDE** | heritable RNAi-defective |
| **HRP** | horseradish peroxidase |
| **IFN** | interferon |
| **IgG** | immunoglobulin G |
| **IP** | immunoprecipitation |
| **iPAR-CLIP** | *in vivo* photoactivatable ribonucleoside enhanced CLIP |
| **kDa** | kilo-Dalton |
| **LB** | Luria-Bertani |
| **mAb** | monoclonal antibody |
| **miRISC** | microRNA-induced silencing complex |
| **miRNA** | microRNA |
| **MosSCI** | *Mos*1-mediated single copy insertion |
| **mRNA** | messenger RNA |
| **MudPIT** | multidimensional protein identification technology |
| **MW** | molecular weight |
| **ncRNA** | non-coding RNA |
| **NGM** | nematode growth medium |
| **NRDE** | nuclear RNAi-defective |

| | |
|---|---|
| **nt** | nucleotide |
| **OH** | hydroxyl group |
| **ORF** | open reading frame |
| **P** | monophosphate group |
| **PAZ** | Piwi Argonaute and Zwille |
| **PCR** | polymerase chain reaction |
| **piRNA** | Piwi Argonaute-interacting RNA |
| **PIWI** | P-element induced wimpy testis |
| **PMSF** | phenylmethylsulfonyl fluoride |
| **PNK** | polynucleotide kinase |
| **Pol I/II/III** | DNA-dependent RNA polymerase I/II/III |
| **Pol II LS** | large subunit of RNA polymerase II |
| **PPP** | triphosphate group |
| **PTP** | protein tyrosine phosphatase |
| **qRT-PCR** | quantitative real-time PCR |
| **R** | purine base |
| **RACE** | rapid amplification of cDNA ends |
| **RDE** | RNAi-defective |
| **RdRP** | RNA-dependent RNA polymerase |
| **RFP** | red fluorescent protein |
| **RLR** | RIG-I-like receptor |
| **RNA** | ribonucleic acid |
| **RNAi** | RNA interference |
| **RNase** | ribonuclease |
| **RNA-seq** | cloning, reverse transcription and deep-sequencing of RNA |
| **RNP** | ribonucleoprotein |
| **rpm** | reads per million |
| **RRF** | RdRP family |
| **rRNA** | ribosomal RNA |
| **scRNA** | small cytoplasmic RNA |
| **SDS-PAGE** | sodium dodecyl-sulphate-polyacrylamide gel electrophoresis |
| **SEM** | standard error of the mean |
| **sRNA** | small RNA |
| **siRNA** | small interfering RNA |
| **snoRNA** | small nucleolar RNA |
| **SNP** | single-nucleotide polymorphism |
| **snRNA** | small nuclear RNA |
| **SSC** | saline-sodium-citrate |
| **ssRNA** | single-stranded RNA |
| **TAE** | Tris-acetate-EDTA |
| **TAP** | tobacco acid pyrophosphatase |
| **TB** | terrific broth |
| **TBE** | Tris-borate-EDTA |
| **TE** | Tris-EDTA |
| **Tris** | Tris hydromethyl aminomethane |
| **tRNA** | transfer RNA |
| **TSS** | transcription start site |
| **UTR** | untranslated region |
| **UV** | ultra violet |
| **WAGO** | worm-specific Argonaute |
| **Y** | pyrimidine base |
| **YAC** | yeast artificial chromosome |

# FOREWORD

This thesis is organized into five chapters and a global conclusion. Chapter I consists of an introduction intended to provide the reader with a foundation to understand and critically evaluate the results described in the following chapters. It presents a thorough, yet non-exhaustive, overview of the main *C. elegans* small RNA pathways; a summary of the current state of knowledge for orthologs of the protein PIR-1 (the focus of our study); and states the context from which this project was born. Results Chapters II, III, and IV contain very brief introductions describing the context, goals and main findings for the respective set of experiments. Results Chapter V includes a more extensive introduction of RNAi-based antiviral immunity in *C. elegans*, which, for simplicity, we found appropriate to exclude from Chapter I. In the final conclusion, we reiterate the main findings, state their significance and propose ideas for the continued study of PIR-1 in *C. elegans* and beyond.

As a note to the reader not accustomed to *C. elegans* nomenclature, all genes, RNA transcripts, and mutants are written with lower case italicized letters (*e.g.*, *pir-1*); proteins are written with upper case letters (*e.g.*, PIR-1); and transcriptional or translational fusions are indicated by two colons (*e.g.*, $P_{sur\text{-}5}$*::gfp (promoter::ORF), pir-1::gfp*, PIR-1::GFP). Specific mutations designate the laboratory, the organization or the consortium that generated the mutant: for example, *rde-1(ne300)* is the mutant number 300 generated in the laboratory of Craig Mello (designated by the acronym 'ne') affecting the *rde-1* gene. The central online repository for all *C. elegans* genetic and molecular information and data is WormBase.org and provides a useful reference to the reader about specific genes mentioned in this thesis. Lastly, we have chosen to refer to all Argonaute binding RNAs as small RNAs (sRNAs), including the more familiar siRNAs and miRNAs.

# CHAPTER I

Introduction

## SMALL RNA PATHWAYS IN *C. ELEGANS*

The discovery of microRNAs (miRNAs; Lee *et al.*, 1993) and subsequently of sRNA-mediated RNA interference (RNAi; Fire *et al.*, 1998) in the nematode *C. elegans* not only opened a new field of research that has been constantly expanding our view of how gene regulation is achieved in all organisms, but also revolutionized the way experimental biology is conducted. In the last quarter of a century we have witnessed an exponential increase in the number of processes regulated by antisense ~15-30 nucleotide-long sRNAs across the entire evolutionary spectrum, from bacteria to humans. This is due, in no small part, to the recent advent of powerful high-throughput sequencing technologies that allowed the identification and characterization of endogenous sRNAs at an unprecedented scale. Nonetheless, akin to the years that followed the full sequencing of the human genome, we are still far from fully understanding how sRNA networks work together to orchestrate biological processes, and how disruption of those networks can lead to diseased states.

*C. elegans* is a particularly well-suited model for surveying sRNA networks in animals. It combines a simple and thoroughly characterized anatomy, straightforward genetic manipulation, and ease with which experiments can be performed at distinct developmental stages of the whole animal. In contrast to more complex animals, these properties greatly simplify the difficult task of dissecting multiple sRNA pathways. Furthermore, despite its morphological simplicity, the *C. elegans* genome carries ~20,000 protein-coding genes, approximately the same number as the most recent estimates for the number of genes in the human genome (~19,000; Ezkurdia *et al.*, 2014). Since the need for a tight orchestration of the expression of this vast set of genes is just as essential in humans as it is in *C. elegans*, the study of sRNAs in this nematode is extraordinarily relevant. Indeed, this model organism already holds an impressive track record for providing generally applicable principles of sRNA-based gene regulation.

At least two major categories of effector endogenous sRNAs coexist in *C. elegans*: miRNAs and 22G-RNAs. The *C. elegans* genome encodes ~250 miRNAs, which act in various combinations to recognize and silence an unknown number of transcripts (likely in the thousands). 22G-RNAs are generally produced from the transcripts that they target, which comprise ~50% of all protein-coding genes. This number is likely an underestimation, as existing studies are far from having reached saturation. While it was initially thought that 22G-RNAs functioned primarily by post-transcriptional silencing of

their target RNAs, it is now clear that they can either repress or activate gene expression at the transcriptional level, depending on which particular pathway they enter. A variety of recent studies have shown that sRNA-induced chromatin states can be inherited across generations. Evidence suggests that this occurs through the passage of sRNAs from parents to progeny, thus providing an epigenetic memory of parental gene expression programs to be recapitulated in the next generation.

The synthesis of 22G-RNAs is itself controlled by highly complex populations of distinct sRNA species, namely 21U-RNAs and 26G-RNAs. The functional outcome of each type of sRNA is dictated by the association with specific members of the conserved Argonaute family of sRNA-binding proteins. *C. elegans* encodes 26 different Argonautes, a remarkably high number if compared to the eight Argonaute proteins encoded in the human genome. As it will become apparent in the following sections, these pathways are highly branched, interconnected and multi-functional. This complexity has only recently started to be unveiled and thus the number of solved problems still dwarfs the immensity of existing questions. It is therefore challenging to provide a concise integrated overview of these pathways, as models keep shifting at dizzying speeds. Rather, the goals of this part of the introduction are to (1) define each type of sRNA and introduce the main biogenesis steps and key enzymes required for their production, and (2) to illustrate the combined regulatory potential of these different sRNAs, which we currently, and somewhat artificially, separate into distinct pathways. Where identified, brief descriptions of biological functions will be given. In the interest of simplicity and space, we integrated relevant comparisons to other organisms in the Results and Discussion sections of each chapter. Table 1.1 on the next page summarizes the main sRNAs pathways occurring in *C. elegans*, and is intended to help the reader to navigate the descriptions in the sections that follow. Additionally, before diving into the complex world of *C. elegans* sRNAs, we begin with a short overview of *C. elegans* anatomy and germline development in order to provide a spatio-temporal context to sRNA pathways.

**Table 1.1.** Summary of *C. elegans* Small RNA Pathways

| sRNA pathway | sRNA type | key enzymes | 5'-end | 3'-end | Argonaute | primary accumulation site | primary subcellular compartment | primary targets | regulatory outcome |
|---|---|---|---|---|---|---|---|---|---|
| **miRNA** | miRNAs | Pol II, DRSH-1, DCR-1 | pU | OH | ALG-1, ALG-2 | soma | cytoplasm | diverse genes | translational repression or degradation |
| **RDE-1** | primary siRNAs | DCR-1 | pN | OH | RDE-1 | soma and germline | cytoplasm | exogenously targeted genes or endogenous sources of dsRNA | no direct effect |
| | secondary siRNAs | RRF-1, EGO-1 | pppG | OH | WAGOs | soma and germline | nucleus (NRDE-3, HRDE-1) or cytoplasm (other WAGOs) | RDE-1 targets | silencing (TGS and PTGS) |
| **21U-RNA/ piRNA** | 21U-RNAs (primary) | Pol II | pU | 2'-*O*-methyl | PRG-1 | germline | P granules | transposons, pseudogenes, foreign sequences (*e.g.*, transgenes) | no direct effect (?) |
| | WAGO 22G-RNAs (secondary) | RRF-1, EGO-1 | pppG | OH | WAGOs | germline | nucleus (HRDE-1), P granules (other WAGOs) | PRG-1 targets | silencing (TGS and PTGS) |
| **ERGO-1 26G-RNA** | 26G-RNAs (primary) | RRF-3, DCR-1 | pG | 2'-*O*-methyl | ERGO-1 | germline, embryos | unknown | pseudogenes, duplicated genes, unknown number of sequences targeted in *trans*. | no direct effect (?) |
| | WAGO 22G-RNAs (secondary) | RRF-1, EGO-1 | pppG | OH | WAGOs | somatic (embryos and early larvae) | Nucleus (NRDE-3) and cytoplasm (other WAGOs) | ERGO-1 targets | silencing (TGS and PTGS) |
| **ALG-3/4 26G-RNA** | 26G-RNAs (primary) | RRF-3, DCR-1 | pG | OH | ALG-3, ALG-4 | germline (spermatogenesis) | P granules | spermatogenesis related genes | no direct effect (?) |
| | WAGO 22G-RNAs (secondary) | RRF-1, EGO-1 | pppG | OH | WAGOs | germline (spermatogenesis) | P granules | subset of negatively regulated ALG-3/4 targets | silencing (PTGS) |
| | CSR-1 22G-RNAs (secondary) | EGO-1 | pppG | OH | CSR-1 | germline (spermatogenesis, spermiogenesis) | P granules, nucleus | subset of positively regulated ALG-3/4 targets | licencing,transcriptional activation |
| **CSR-1** | CSR-1 22G-RNAs | EGO-1 | pppG | OH | CSR-1 | germline | P granules, nucleus | germline-expressed genes | licencing,transcriptional activation |

This table is adapted and modified from Youngman and Claycomb, 2014. 'OH', hydroxyl; 'p', monophosphate; 'ppp', triphosphate; 'N', any nucleotide; 'TGS', transcriptional gene silencing; 'PTGS', post-transcriptional gene silencing.

## A Brief Primer on *C. elegans* Germline Development

Due to its genetic tractability, short life cycle and relative simplicity of both somatic and germline tissues, *C. elegans* has been invaluable in dissecting the variety of sRNA pathways and functions that can coexist in an entire animal. The expression of sRNAs varies substantially during development, particularly in the germline where they are most abundant and diverse, with the notable exception of miRNAs, which are more abundantly expressed in the soma (Ruby *et al.*, 2006; Gu *et al.*, 2009; Claycomb *et al.*, 2009). In *C. elegans*, endogenous 22G-RNA-based gene regulation is particularly important in the germline, where stem cell proliferation and gamete differentiation processes occur simultaneously and continuously. Not only do 22G-RNAs control aspects of these processes, but also provide protection of germ cell genome integrity against internal or external agents (*e.g.*, transposable elements or UV radiation, repectively). In order to facilitate the understanding of sRNA biogenesis and function in *C. elegans* it is therefore important to discuss key aspects of its anatomy, life-cycle and germline development (an extensive and up-to-date description of these and other aspects of *C. elegans* biology can be found at WormAtlas.org, Altun *et al.*, 2002-2015).

*C. elegans* has a diploid genome distributed across five autosomes (I-V) and one sex chromosome (X). It reproduces predominately by self-fertilization in its hermaphrodite form, dictated by the presence of two X chromosomes. Males, which only have one X chromosome (XO), normally exist at a frequency of ~0.1% and arise as the result of spontaneous non-disjunction of the X chromosome during meiosis in the hermaphrodite. Upon mating, male sperm out-compete stored hermaphrodite sperm and progeny can reach a male frequency of up to 50%. An adult hermaphrodite is composed of exactly 959 somatic cells and ~1000-2000 germ cells, while an adult male has 1,031 somatic cells and ~1,000 germ cells. Main organ-like systems include a large two-arm tubular germline (only one arm in males), a pharynx to process food and to help circulate fluid throughout the body cavity, and a large intestine that runs across the entire anterior-posterior axis of the animal. Under standard conditions of food and temperature (~20˚C), an adult hermaphrodite can lay around 300 eggs and live up to three weeks. Embryonic development is initiated inside the adult, but most of it occurs after egg-laying. Progeny hatch as fully autonomous larvae and go through four molting events, separating four larval stages (L1 to L4) before reaching the adult, fertile stage. At 20˚C the development from a single-celled embryo to an egg-laying adult takes only 60-70 hours (Fig. 1.1A).

Hermaphrodite germline development begins at the L1 stage through the division of pluripotent germ precursor cells that originate during embryogenesis, Z2 and Z3 (Fig. 1.1B). From the L1 to L3 stages, germline growth occurs exclusively through mitotic divisions of germline cells. The distal tip cells (DTCs), which are of somatic origin, function at the growing ends of gonadal arms to continuously signal mitotic cell proliferation. Germ cells line the inner surface of the tubular structure of the germline and are only partially enveloped by a plasma membrane, sharing a common cytoplasm contained in the rachis (or the hollow center of the germline), effectively constituting a syncytium (Figure 1.1B). Meiosis does not begin until the L4 stage at the proximal end of the germline, near the section that unites the two gonadal arms, where two spermathecae, a uterus, and a vulva develop. As the germline grows, a higher proportion of the cells become meiotic. During the late L4 stage and the beginning of adulthood, meiosis is dedicated to hermaphrodite sperm differentiation. This starts with the formation of cellularized primary spermatocytes that detach from the germline wall. Completion of meiosis I divides the 4N primary spermatocyte into two, 2N, secondary spermatocytes. After meiosis II is completed, four haploid spermatids are generated. The sequence of events that go from a primary spermatocyte to a spermatid is termed spermatogenesis. These cells are stored in the spermatheca of the adult until the passage of oocytes leads to their activation in a process termed spermiogenesis, in which they become highly motile, fertilization-competent spermatozoa. During adulthood, meiotic cells start differentiating exclusively into oocytes which arrest at the diakinesis stage of meiosis I. Upon passage through the spermatheca, fertilization occurs, oocyte meiosis is resumed, and embryonic development is initiated. Embryos are stored in the uterus until they reach the gastrula stage (~30 cells) at which time they are expelled through the vulva.

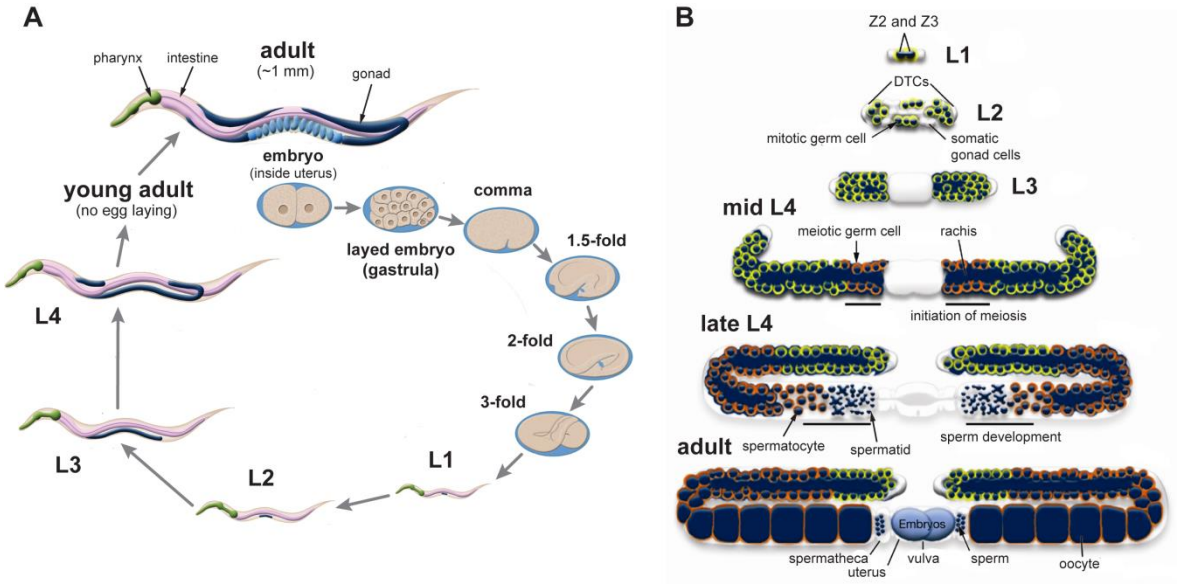



**Figure 1.1. Life cycle of *C. elegans* and germline development. (A)** Life cycle of a self-fertilizing hermaphrodite. Major stages of embryogenesis are shown. Every larval stage is separated by molting to replace the cuticle. The last molting event occurs in the L4 to adult transition. Wild-type *C.elegans* grows over a range of temperatures (typically 15-25˚C in the laboratory), influencing brood size and the duration of each developmental stage. **(B)** General architecture and development of the hermaphrodite germline. In the L1 stage, the two pluripotent germ precursor cells Z2 and Z3 begin to divide mitotically. These mitotic divisions continue into adulthood, and are stimulated by signaling from the DTCs, one per each gonad arm. The DTCs, the cells that envelope the germline, and the cells composing the spermathecae, uterus and vulva are all part of the so called somatic germline. The core of each gonadal arm is termed the rachis and constitutes a common cytoplasm to the partially cellularized nuclei that line the inner surface of the germline. Meiosis initiates in the proximal part of the germline during the L4 stage, during which they undergo differentiation into spermatids that are stored in the spermatheca of the developing L4 larva and adult. The sperm to oocyte switch occurs during adulthood. Oocytes are fertilized upon passage through the spermatheca, and embryos accumulate in the uterus for some time before being laid. Figures were adapted from WormAtlas.org (Altun *et al.*, 2002-2015).

## The MicroRNA Pathway

MicroRNAs represent the most abundant and well-studied class of sRNAs. *C. elegans* expresses ~253 distinct miRNAs (WormBase release WS248; WormBase.org, Harris *et al*., 2014). Mature miRNAs are ~23-nt long, with 5'-monophosphate and 3'-hydroxyl termini. They act by binding to partially complementary sites in the 3' untranslated regions (UTR) of mRNAs, to direct their post-transcriptional silencing (Jonas and Izaurralde, 2015). mRNAs frequently have more than one miRNA binding site, such that one target may be subjected to simultaneous regulation by several different miRNAs. The rules that dictate the pairing of miRNAs with their targets, primarily based on studies of the most abundant and conserved miRNAs, are fairly consensual. However, as new exceptions to these rules continue to be discovered, the prediction that the majority of

protein-coding RNAs are susceptible to some degree of miRNA-mediated regulation is gaining ground (Bartel, 2009).

Each miRNA is genomically encoded and transcribed by RNA polymerase II (Pol II) into a long, capped and polyadenylated RNA termed primary miRNA (pri-miRNA; Fig. 1.2; Bracht *et al.*, 2004). The miRNA sequence is embedded in the dsRNA portion of a stem-loop structure within the pri-miRNA, which is released by the joint activity of the endoribonuclease Drosha and the dsRNA binding protein Pasha/DGCR8 (DRSH-1 and PASH-1 in *C. elegans*, respectively; Denli *et al.*, 2004). This stem-loop of ~70 nt, with a 3' 2-nt overhanging terminus, termed pre-miRNA, is recognized by an ortholog of the human nuclear export protein exportin 5 (possibly XPO-1/IMB-4 in *C. elegans*) that transports it to the cytoplasm (Parry *et al.*, 2004).

The next processing steps require Dicer (DCR-1 in *C. elegans*) and two redundant Argonaute proteins, ALG-1 and ALG-2 (*A*rgonaute-*L*ike *G*ene 1 and 2); Grishok *et al.*, 2001; Ketting *et al.*, 2001; Hutvagner *et al.*, 2001). Dicer is a large (~220 kDa in *C. elegans*), highly conserved endoribonuclease belonging to the RNase III family of dsRNA cleaving enzymes, to which Drosha also belongs. Dicer is composed of a helicase domain, a PAZ (*P*iwi *A*rgonaute and *Z*wille) domain that binds the substrate 3' end, two RNase III domains and a dsRNA-binding domain (Bernstein *et al.*, 2001). As in humans, *C. elegans* only expresses one Dicer enzyme which is central for the synthesis of not only miRNAs but also for small-interfering RNAs (siRNAs; discussed in the next section) and certain endogenous sRNAs (endo-sRNAs). Argonautes are evolutionarily conserved sRNA-binding proteins consisting of a highly variable N-terminal domain, a PAZ domain for 3'-end nucleotide binding, a MID (*mid*dle) domain, which makes contact with the 5'-phosphates of sRNAs, and a C-terminal PIWI (*P*-element *i*nduced *w*impy *t*estis) domain with an RNase H-like fold to promote RNA cleavage, an activity termed ‘slicing’ (reviewed in Hutvagner and Simard, 2008).

DCR-1 binds the pre-mRNA using its PAZ domain to anchor the 2-nt 3' overhang generated by DRSH-1, and cleaves the dsRNA stem 22-23 nt away from the 3' end (Zhang et al., 2002; Macrae et al., 2007), resulting in a 22-23-nt long duplex RNA bearing 5'-monophosphates and 2-nt 3'-overhanging termini. The duplex is loaded into ALG-1 or -2 and one of the strands is cleaved (the passenger strand) by the slicer activity of the Argonaute, leaving the guide strand free to interact with its target (Bouasker and Simard, 2012). In *C. elegans* the majority of mature miRNAs carry a 5' uridine, which is thought to promote the specific interaction of miRNAs with ALG-1/2 (Ruby *et al.*, 2006).

MiRNA-loaded ALG-1/2 Argonautes interact with a variety of proteins to bind the partially complementary 3' UTR sites on target mRNAs. Collectively, these sets of proteins are termed miRNA-induced silencing complexes (or miRISCs). A variety of miRISCs exist, based on the composition of factors associating with the core Argonaute-miRNA complex. One type of miRISC includes the micrococcal nuclease-related TSN-1 (*T*udor *S*taphylococcal *N*uclease ortholog) in association with the dsRNA-binding factor VIG-1 (ortholog of the *Drosophila V*asa *I*ntronic *G*ene) (Caudy *et al.*, 2003). Another type is defined by the ortholog of human GW182, AIN-1 (*A*LG-1 *In*teracting Protein 1), which facilitates both miRNA loading and the interaction of miRNAs with their mRNA targets in association not only with ALG-1/2, but also with DCR-1 (Ding *et al.*, 2005). AIN-2 acts redundantly with AIN-1 by also binding DCR-1 and ALG-1/2, but can additionally recruit TSN-1 and the mRNA-binding translational repressing protein GLD-1 (Defective in *G*erm*l*ine *D*evelopment 1). Moreover, AIN-2 has been shown to interact with components of the translation initiation machinery (Parry *et al.*, 2007; Zhang *et al.*, 2007). Finally, NHL-2 (carrying a TRIM-NHL domain, which defines a large class of metazoan proteins; Wulczyn *et al.*, 2010) in association with the conserved DEAD-box (Asp-Glu-Ala-Asp motif) RNA helicase CGH-1 (*C*onserved *G*ermline *H*elicase 1), is required to increase the strength of binding of miRISCs to their target mRNAs (Hammell *et al.*, 2009). The DEAD-box motif is present in numerous enzymes acting in sRNA pathways (including Dicer). Proteins carrying this motif form a subgroup within a large family of proteins present in all organisms defined more generally by the DExD/H motif, and are involved in a wide range of processes pertaining mostly to RNA metabolism (reviewed in Fuller-Pace, 2006 and Linder and Jankowsky, 2011). CGH-1 orthologs have been demonstrated to be required for miRISC activity, and are involved in several other aspects of mRNA post-transcriptional regulation and turnover (Chu and Rana, 2006; Eulalio *et al.*, 2007; Rajyaguru and Parker, 2009).

All of the aforementioned proteins are required for effective silencing of miRNA targets, whether it occurs by translational repression, mRNA degradation, or both (Wightman *et al.*, 1993; Bagga *et al.*, 2005; reviewed in Jonas and Izaurralde, 2015). Notably, both AIN-1 and NHL-2/CGH-1 complexes localize to cytoplasmic domains called processing bodies (P-bodies), which contain a variety of ribonucleoprotein (RNP) complexes dedicated to processes such as nonsense-mediated decay, mRNA decay upon decapping and translational repression (Parker and Sheth, 2007). Through these molecular activities, miRNAs control various aspects of somatic development and physiology,

namely in aging, nervous system patterning and function, and general regulation of cell fate and developmental timing (Ambros, 2003; Abbott, 2011; Inukai and Slack, 2013). Interestingly, while individually deleting most miRNAs does not result in perceivable deleterious phenotypes (Miska *et al.*, 2005), the collective deletion of some of the 23 miRNA families found in *C. elegans* can have very dramatic developmental consequences (Alvarez-Saavedra and Horvitz, 2010). The best example is the *mir-35-41* family, which is essential to enable proper embryonic development over a wide range of temperatures, and to ensure full reproductive capacity in adult animals (McJunkin and Ambros, 2014). Studies of miRNA function in *C. elegans* continue to support the increasingly consensual idea that miRNAs have evolved to confer robustness upon sensitive gene expression programs in the face of challenging environmental conditions (Burke *et al.*, 2015; Ren and Ambros., 2015).

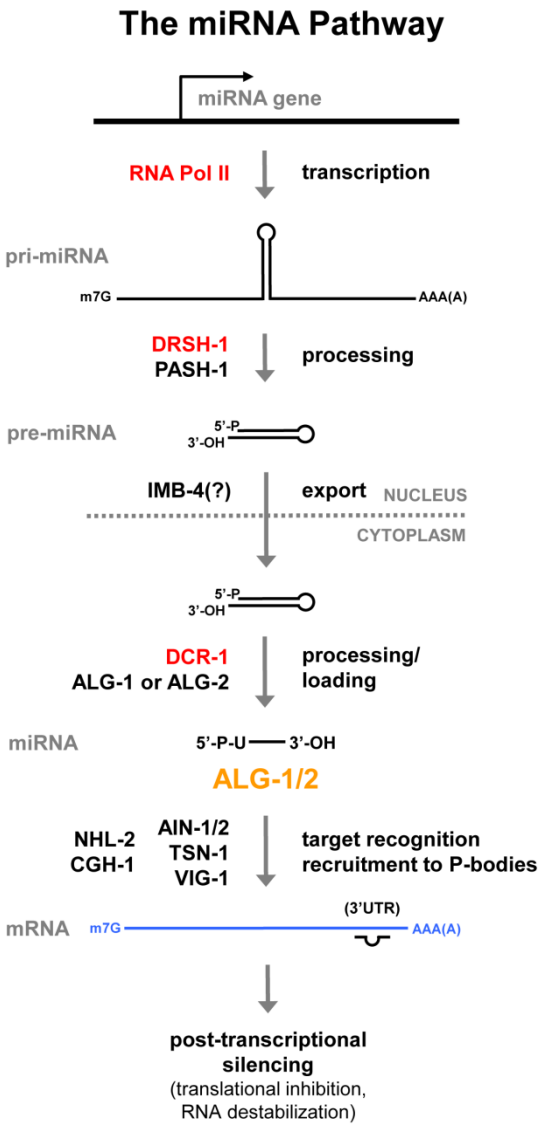


**Figure 1.2. The microRNA pathway.** Processing of miRNAs takes place in the nucleus and cytoplasm. Loading into ALG-1 or ALG-2 is thought to be concomitant with Dicer cleavage, as the two proteins stably interact *in vivo*. Loaded ALG-1/2 complexes then bind additional factors that recruit them to sites of RNA regulation, such as cytoplasmic P-bodies, where thay associate with target mRNAs, predominately via imperfect base pairing within 3' UTRs. This can lead to translational impairment and/or to direct degradation, upon recruitment of destabilizing factors such as RNA deadenylases.

## The RDE-1 Pathway: Canonical RNA Interference and Beyond

Experimentally, RNAi can be triggered in *C. elegans* by exposing cells to double-stranded RNA (dsRNA) directly by injection or through feeding of dsRNA-expressing bacteria, in a process termed exogenous RNAi or exo-RNAi (Fig. 1.3 and Table 1.1). The process is strictly dependent on the Argonaute RDE-1, which is loaded with siRNAs resulting from the cleavage of dsRNA by Dicer. RNAi is both specific, leading to the silencing of sequences identical to the original dsRNA, and systemic, in that it is able to spread to most cells via the SID-1 (*S*ystemic RNA*i* *D*efective 1) dsRNA specific transmembrane channel (Winston *et al.*, 2002; Shih and Hunter, 2011). In association with the dsRNA-binding protein RDE-4 (*R*NAi-*de*ficient 4; Tabara *et al.*, 2002), DCR-1 cleaves the dsRNA to generate 23-nt duplex RNAs with 2-nt 3'-hydroxyl overhangs and 5'-monophosphate termini (Grishok *et al.*, 2001; Ketting *et al.*, 2001; Knight and Bass, 2001). The small interfering RNA (siRNA) duplexes are loaded into the Argonaute protein RDE-1, which also stably interacts with DCR-1 (Tabara *et al.*, 1999). Similarly to ALG-1/2, after binding the siRNA duplex, RDE-1 cleaves one of the strands (the passenger strand) using its RNase H slicer activity, to retain a single strand that will serve as the guide strand to recognize and bind RNA target molecules (Steiner *et al.*, 2009). Target RNA is cleaved near the RDE-1 binding site by the endonuclease RDE-8, followed by the addition of a 3' poly-uridine tract synthesized by the polynucleotidyl transferase RDE-3 (Chen *et al.*, 2005; Tsai *et al.*, 2015). This event is thought to signal the recruitment of the RNA-dependent RNA polymerases (RdRP) RRF-1 (*R*d*R*P *f*amily 1), expressed in the soma and germline (Sijen *et al.*, 2001), and EGO-1 (*E*nhancer of *G*lp-*O*ne 1), expressed exclusively in the germline (Smardon *et al.*, 2001).

RdRPs use the RNA template to catalyze the primer-independent synthesis of antisense 22-23 nt siRNAs bearing a 5'-triphosphorylated guanosine, termed 22G-RNAs (Sijen *et al.*, 2001; Aoki *et al.*, 2007; Pak and Fire, 2007; Sijen *et al.*, 2007). These siRNAs spread across the template RNA, 5' of the RDE-1 binding site, usually within a range of

100-180 nt (Alder *et al.*, 2003). Both RRF-1 and EGO-1 must physically associate with two co-factors to produce 22G-RNAs: EKL-1 (*E*nhancer of *K*SR-1 *L*ethality 1), a Tudor domain protein (Kim *et al.*, 2005; Robert *et al.*, 2005; Rocheleau *et al.*, 2008; Claycomb *et al.*, 2009), and the helicase DRH-3 (*D*icer-*R*elated *H*elicase 3; Duchaine *et al.*, 2006; Aoki *et al.*, 2007; Nakamura *et al.*, 2007; Claycomb *et al.*, 2009; Gu *et al.*, 2009). Loss-of-function mutants for either of these factors are completely sterile and unable to produce 22G-RNAs, including those that are triggered endogenously, as discussed ahead (Claycomb *et al.*, 2009; Gu *et al.*, 2009). 22G-RNAs are loaded into Argonautes belonging to a group of 12 semi-redundant worm-specific Argonautes (or WAGOs) to target complementary RNAs for silencing (Yigit *et al.*, 2006; Gu *et al.*, 2009; reviewed in Buck and Blaxter, 2013). *C. elegans* specifically requires four WAGOs for exo-RNAi silencing, as their combined deletion leads to complete RNAi resistance in somatic and germline cells (Yigit *et al.*, 2006). The four Argonautes are WAGO-4, WAGO-7(PPW-1), WAGO-6(SAGO-2), and WAGO-8(SAGO-1). PPW-1 (*P*AZ/*P*I*W*I Domain-Containing 1) is required predominately for germline RNAi, while the remaining Argonautes are employed in somatic RNAi (Tijsterman *et al.*, 2002; Yigit *et al.*, 2006).

The RNAi process is thus divided into two main phases: the primary phase, in which low abundance siRNAs are directly derived from Dicer-mediated cleavage of the dsRNA trigger; and the secondary (or amplification) phase, in which primary siRNA-loaded RDE-1 initiates the abundant RdRP synthesis of antisense 22G-RNAs (Fig. 1.3 and Table 1.1). The exo-RNAi response has been shown to be self-contained, in that (1) secondary 22G-RNAs are unable to initiate new rounds of RdRP-mediated amplification (Pak *et al.*, 2012), and (2) RDE-1 cannot engage secondary 22G-RNAs or silence targets in the absence of 22G-RNA synthesis (Sijen *et al.*, 2001; Yigit *et al.*, 2006). Since the most critical Argonautes for exo-RNAi are assumed to be predominately cytoplasmic, the bulk of silencing is thought to occur post-transcriptionally. However, exogenously triggered 22G-RNAs can also be loaded into WAGO proteins that translocate into the nucleus to elicit transcriptional silencing, a process termed nuclear RNAi (discussed ahead).

RNAi can be triggered naturally by ingestion and uptake of dsRNA from the environment through the intestinal SID-2 transmembrane channel (Winston *et al.*, 2007) or through infection by RNA viruses (Felix *et al.*, 2011; introduced in Chapter V), to cite two known examples. Cells also make use of this pathway from endogenous sources of dsRNA arising, for instance, from transposons or highly-repetitive transgene arrays (Ketting *et al*., 1999; Sijen and Plasterk, 2003; Vastenhouw *et al*., 2003; Grishok *et al*., 2005; Robert *et*

*al*., 2005). Additionally, the formation of regions of dsRNA within *C. elegans* transcripts has recently been shown to be widespread (Whipple *et al.*, 2015). Several of these sequences overlap with protein coding genes that are known targets of endogenous 22G-RNAs. In combination with other factors, Dicer is proposed to use these dsRNA regions as substrates to produce primary siRNAs that likely enter the RDE-1 pathway to elicit the silencing of endogenous RNAs.

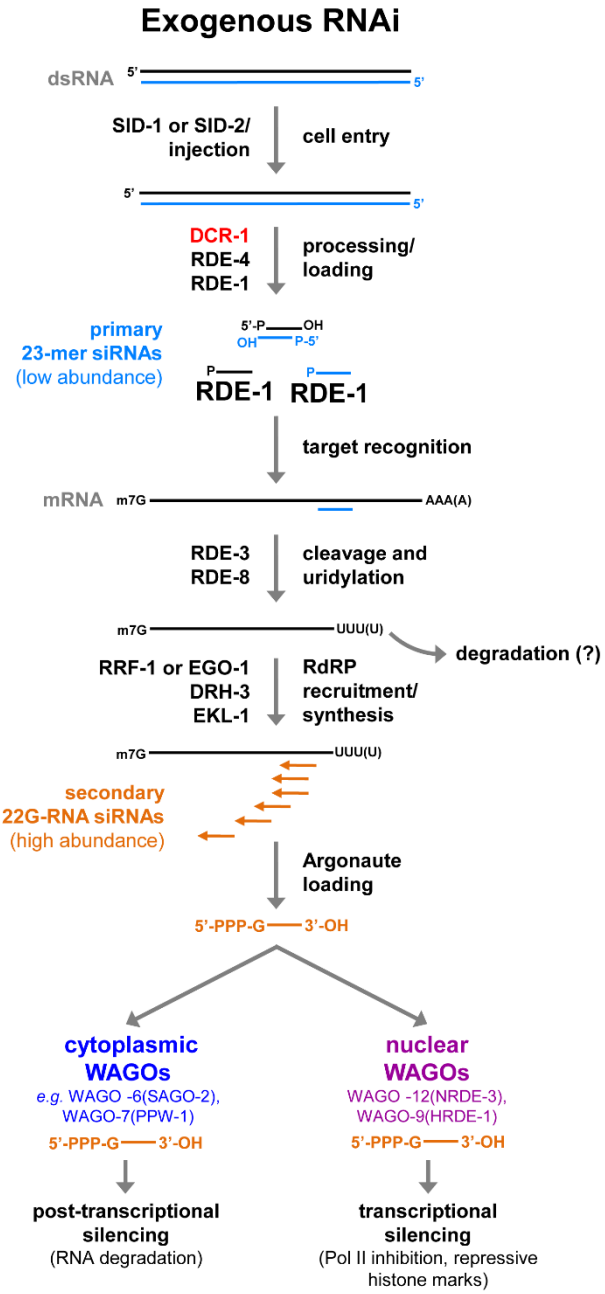



**Figure 1.3. The exogenous RNAi pathway.** Exo-RNAi is triggered by the uptake of dsRNA from the environment via the intestinal dsRNA transmembrane channel SID-2, or experimentally by

direct injection of dsRNA. There are two types of siRNAs: the primary siRNAs, which are direct products of Dicer cleavage and are only as abundant as the dsRNA source; and the secondary siRNAs, the result of RdRP synthesis triggered by RDE-1-bound primary siRNAs in somatic cells (by RRF-1), and in the germline (by both RRF-1 and EGO-1). The diagram is a simplification of our current knowledge of the process, as many of the factors required for 22G-RNA accumulation have been omitted. The RDE-8/RDE-3 step is presumed, but not proven, to recruit RdRPs to the target RNA. Downstream silencing can occur through post-transcriptional mechanisms in germline P granules or somatic P-body-like granules, and through transcriptional mechanisms inside the nucleus, depending on which WAGOs secondary siRNAs associate with. 'm7G', 5' methylguanosine cap.

## The 21U-RNA/piRNA Pathway

*C. elegans* expresses over 30,000 distinct, genomically encoded 21-nt RNAs carrying 5'-monophosphorylated uridines and 2'-*O*-methyl modified 3' termini, referred to as 21U-RNAs (Fig. 1.4A and Table 1.1; Ruby *et al.*, 2006; Batista *et al.*, 2008; Gu *et al.*, 2012). Owing to their predominant germline expression, 5' and 3' molecular features, and their Dicer-independent biogenesis, 21U-RNAs are considered the *C. elegans* equivalents of animal piRNAs (*P*IWI-*i*nteracting *RNA*s). This class of endogenous sRNAs has been demonstrated to promote proper germline development and gamete differentiation while protecting the integrity of germline genomes from the mutagenic effects of active transposing elements, from *Drosophila* to mice (reviewed in Luteijn and Ketting, 2013). In contrast to all other sRNAs that interact with either Argonaute-clade (RDE-1, ALG-1/2, ALG-3/4; conserved from plants to animals) or with *C. elegans*-specific WAGO-clade Argonautes, 21U-RNAs interact with a conserved PIWI-clade Argonaute named PRG-1 (*P*iwi *R*elated *G*ene 1; Batista *et al.*, 2008), similarly to piRNAs in other species.

21U-RNAs are transcribed by Pol II, and are subdivided into two groups depending on the elements that govern their expression (Fig. 1.4A; Ruby *et al.*, 2006; Cecere *et al.*, 2012; Gu *et al.*, 2012). Type-I 21U-RNAs are encoded primarily in intergenic clusters spread throughout chromosome IV or from sequences overlapping with other Pol II genes. Each Type-I locus is preceded by a two-part motif 25-60 nt upstream of the locus, which is recognized by specific Forkhead-family transcription factors (Ruby *et al.*, 2006; Batista *et al.*, 2008; Das *et al.*, 2008; Cecere *et al.*, 2012; Billi *et al.*, 2013). Type-II 21U-RNAs arise from a subset of abortive Pol II transcription products at transcription start sites (TSS) of active promoters. These short (10-40 nt), capped products termed capped small RNAs (csRNAs), are often transcribed bi-directionally around the TSS (Gu *et al.*, 2012). As part of their upstream motif, both Type-I and Type-II 21U-RNAs have a short consensus YRNT sequence (where Y is a pyrimidine, R is a purine, N is any nucleotide and T

encodes the first nucleotide of the mature 21U-RNA), that determines processing of a ~26-nt long precursor capped at the R position of the motif. Type-II 21U-RNAs originate whenever a csRNA happens to contain the YRNT motif that determines the synthesis of a capped ~26 nt precursor (Gu *et al.*, 2012). After decapping and cleavage by unknown factors, the 3' end is methylated by the conserved methyltransferase HENN-1 (*HEN*1 of *N*ematode; Billi *et al.*, 2012; Kamminga *et al.*, 2012; Montgomery *et al.*, 2012). Recent screens have identified various new factors required for 21U-RNA processing and accumulation, the activities of which are still not clear (de Albuquerque *et al.*, 2014; Goh *et al.*, 2014).

The loss of PRG-1 leads to the complete disappearance of 21U-RNAs and has severe phenotypic consequences. *prg-1* null-mutant animals exhibit lower brood sizes at 20˚C and complete sterility at 25˚C due to a failure to produce enough mature spermatids (Yigit *et al.*, 2006; Batista *et al.*, 2008; Das *et al.*, 2008; Wang and Reinke, 2008). The proliferation of germ cells is also heavily affected, regardless of temperature (Batista *et al.*, 2008). PRG-1 is present in adults and embryos, localizing primarily to germline and embryonic P granules. In the germline, P granules are found on the cytoplasmic side of the nuclear membrane, surrounding nuclear pores (Pitt *et al.*, 2000). These domains are functionally analogous to P-bodies and stress granules (Gallo *et al.*, 2008). Intriguingly, the loss of PRG-1 does not lead to dramatic changes in the overall gene expression profile (Batista *et al.*, 2008). The main exception concerns the Tc3 transposon, the upregulation of which resulted in a 100-fold increase in transposition rate. Tc3 silencing was found to depend both on PRG-1 and on WAGOs, revealing the first link between the 21U-RNA pathway and 22G-RNA effectors (Das *et al.*, 2008).

Exceptions aside, and unlike piRNAs in other species, most 21U-RNAs lack perfect complementarity to transposons or any other gene sequences (Siomi *et al.*, 2011). Through the use of GFP reporter transgenes carrying sites complementary to particular 21U-RNAs, it was demonstrated that PRG-1-21U-RNA complexes result in GFP silencing. This activity was shown not to require the slicer activity of PRG-1 and to tolerate up to four mismatches between the 21U-RNA and the target sequence, suggesting that 21U-RNAs have the potential to interact with almost any sequence they encounter (Bagijn *et al.*, 2012; Lee *et al.*, 2012). Furthermore, these studies demonstrated that 21U-RNAs trigger the production of secondary WAGO-associated 22G-RNAs, and that these 22G-RNAs are passed from parents to progeny to maintain transgene silencing via the establishment of

repressive chromatin marks (Ashe *et al.*, 2012; Bagijn *et al.*, 2012; Lee *et al.*, 2012; Luteijn *et al.*, 2012; Shirayama *et al.*, 2012).

Together, these findings have led to the hypothesis that the 21U-RNA pathway has evolved as a nucleic acid-based adaptive immune system primed to recognize and silence foreign, potentially deleterious sequences. Additionally, since *prg-1* mutant animals become progressively less fertile after each generation, 21U-RNAs are thought to safeguard the integrity of germline genomes, thus maintaining the immortality of germline stem cells. This likely occurs via the control of the Tc3 transposon, but possibly through the control of additional destabilizing elements (Batista *et al.*, 2008; Das *et al.*, 2008). Finally, 21U-RNAs have been shown to regulate coding-gene expression by specifically promoting the accumulation of spermatogenesis transcripts in males to produce wild-type sperm (Wang and Reinke, 2008), and by directing the synthesis of distinct populations of 22G-RNAs from male or female germline-specific transcripts (Billi *et al.*, 2013; Shi *et al.*, 2013). The great diversity of regulatory outcomes imposed by the 21U-RNA pathway raises numerous questions. Perhaps the most important ones pertain to which factors and mechanisms are employed to ensure that 21U-RNAs are sorted into distinct functional groups.

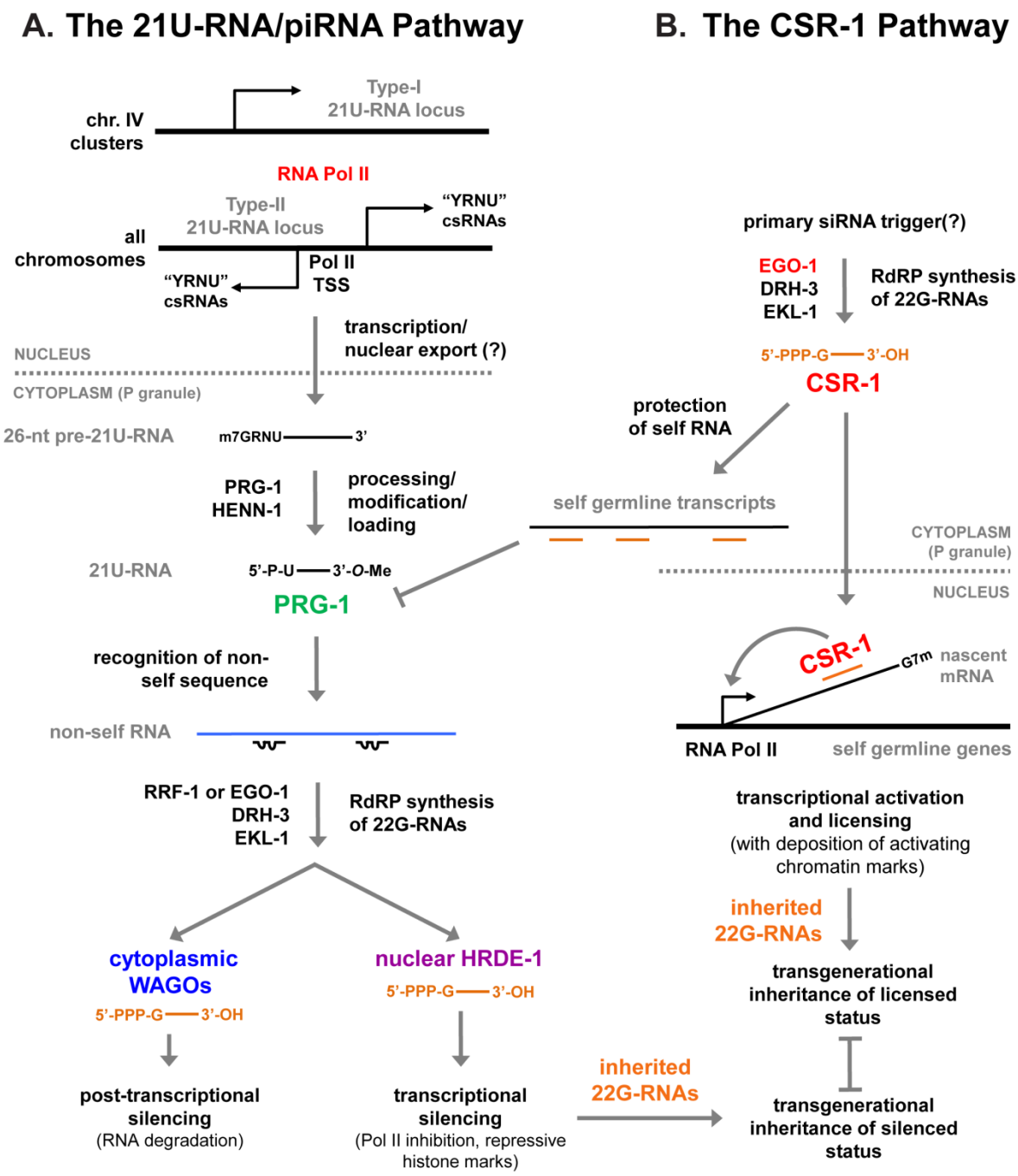



**Figure 1.4. The 21U-RNA/piRNA pathway and the CSR-1 22G-RNA pathway of the germline. (A)** 21U-RNA precursors are synthesized by Pol II from two types of loci. Type-I loci occur predominately in clusters located on chromosome IV where each 21U-RNA is expressed from its own promoter element. Type-II loci occur at promoter proximal regions associated with Pol II transcriptional units. At these loci, Pol II pausing and early termination in both directions produces csRNAs that are processed into 21U-RNAs, provided they are sufficiently long and contain the 5' YRNU motif. The stage at which the precursor is exported out of the nucleus into P granules and factors required for the 3' and 5' processing of the capped 26-nt precursor are still largely unknown. 21U-RNAs associate with PRG-1 to recognize their targets. Most 21U-RNAs do not exhibit perfect complementarity to endogenous transcripts, but are still able to elicit 22G-RNA synthesis by secondary RdRPs by an unknown mechanism. This property allows 21U-RNAs to scan the transcriptome for potentially deleterious sequences considered as non-self (transposons, recently acquired genes, pseudogenes, transgenes, etc…) and elicit their silencing by post-

transcriptional and transcriptional mechanisms. Targeted genes likely remain stably silenced by repressive chromatin marks until germ cell transcription starts in the next generation. 22G-RNAs antisense to these sequences are passed on in gametes to reestablish repressive chromatin domains at the same sites in the germ cells of the progeny. **(B)** CSR-1 22G-RNAs (discussed in the last section dedicated to 22G-RNAs, page 51) target all expressed germline transcripts. For the most part, primary siRNA triggers are not known, suggesting that EGO-1, the only RdRP that produces CSR-1 22G-RNAs, may be recruited to its substrates by alternative mechanisms, namely in the nucleus, where it also localizes. CSR-1 does not silence its targets. Rather, it associates with nascent mRNAs in the nucleus to promote their transcription by stimulating the deposition of activating chromatin marks, and with mature mRNAs in cytoplasmic P granules to protect them from 21U-RNA targeting. Together, these activities led to the hypothesis that CSR-1 is the central agent of a protective "licensing" mechanism that ensures continuous germline gene expression, permanently counteracting the inhibitory effects of the 21U-RNA pathway. CSR-1 22G-RNAs are also thought to be inherited to the following generation to perpetuate active chromatin states, guaranteeing the fidelity of germline gene expression programs. '3'-*O*-Me', 3'-*O*-methyl modification.

## 26G-RNAs and the ERI Complex

The discovery of 21U-RNAs was accompanied by the unveiling of an additional distinct group of endogenous sRNAs named 26G-RNAs, owing to their characteristic length of 26 nt and 5'-monophosphorylated guanosine (Ruby *et al.*, 2006; Asikainen *et al.*, 2008; Han *et al.*, 2009; Stoeckius *et al.*, 2009; Gent *et al.*, 2010). 26G-RNAs were found to constitute endogenous primary siRNAs that trigger the amplification of 22G-RNAs to regulate the expression of target genes at particular developmental stages (Table 1.1).

26G-RNAs biogenesis depends on the activity of a stable ~850 kDa multi-protein complex known as the ERI complex. It consists of a core RdRP module containing RRF-3, DRH-3 and ERI-5, with the latter protein connecting it to a second module composed of DCR-1, RDE-4, ERI-1b and ERI-3 (Duchaine *et al.*, 2006; Thivierge *et al.*, 2012). Loss of RRF-3, ERI-5, ERI-1b or ERI-3 results in a temperature-sensitive sperm-related defect that renders animals sterile at the higher limit temperature of 25˚C, as well as in the enhancement of RNAi (Eri) phenotype. As the name implies, the Eri phenotype occurs whenever RNAi of genes that are normally insensitive or display mild phenotypes upon silencing, results in perceivable or stronger phenotypes (Simmer *et al.*, 2002; Duchaine *et al.*, 2006). Loss of DCR-1, DRH-3, and RDE-4 disables RNAi, and loss of the first two also leads to unconditional sterility and other defects due to pleiotropy (Tabara *et al.*, 1999; Grishok *et al.*, 2001; Ketting *et al.*, 2001; Knight and Bass, 2001; Duchaine *et al.*, 2006; Gu *et al.*, 2009). Although the triggering and processing events that lead to a mature 26G-RNA are still unclear, the properties of some of the ERI-complex proteins explain how some of the key steps may occur.

The RdRP RRF-3 is assumed to transcribe 26G-RNAs, since it is required for their accumulation (Simmer *et al.*, 2002; Gent *et al.*, 2009; Pavelec *et al.*, 2009). Similarly to its paralogs RRF-1 and EGO-1, its products indicate that it prefers to initiate transcription with GTP (Aoki *et al.*, 2007). Moreover, based on the activity of the other RdRPs, RRF-3 products are expected to be short, despite the fact that 26G-RNAs, unlike 22G-RNAs, require processing by DCR-1 (Han *et al.*, 2009). Synthesis appears not to occur processively, as 26G-RNAs do not exhibit regular phasing along their template RNAs and often overlap with other 26G-RNAs (Ruby *et al.*, 2006; Vasale *et al.*, 2010; Fischer *et al.*, 2011). Notably, as Dicer enzymes are not known to have dephosphorylation activity, and RRF-3 presumably generates 5'-triphosphorylated ends, the process through which 26G-RNAs become 5'-monophosphorylated is unknown.

DCR-1 was found to be required for 26G-RNA accumulation through the study of a mutant that impairs the activity of its conserved N-terminal helicase domain. *dcr-1(mg375)* helicase mutant animals exhibit both temperature-sensitive sterility and Eri phenotypes. Consistently, 26G-RNAs were found not to accumulate in this mutant, while the miRNA population was intact (Pavelec *et al.*, 2009; Welker *et al.*, 2010). When different RNA substrates were treated with cell-free extracts containing either intact or helicase mutant DCR-1, the cleavage of dsRNA with 2-nt 3' overhangs gave rise to 21-23 nt products regardless of the helicase domain. Since this substrate simulates the dsRNA end of a pre-miRNA, it explains why *dcr-1(mg375)* mutant animals still express miRNAs. This property is conserved in *Drosophila*, where miRNA-processing Dicer-1 naturally lacks a functional helicase domain (flies express two Dicer enzymes, unlike humans and worms which express only one; Lee *et al.*, 2004; Jiang *et al.*, 2005). However, when the provided substrate was blunt-ended dsRNA, wild-type DCR-1 was able to generate a 5'-3' 26-nt product with an accompanying 22-23-nt 3'-5' product, while helicase-deficient DCR-1 was not (Welker *et al.*, 2011). In agreement with this, *Drosophila* Dicer-2, which possesses a functional helicase domain and is dedicated to processing exo- and endo-siRNAs from long dsRNA, was also shown to require a functional helicase domain to cleave dsRNA, and it did so much more efficiently on blunt-ended dsRNA (Welker *et al.*, 2011; Sinha *et al.*, 2015). Therefore, the helicase domain of *C. elegans* Dicer allows the enzyme to adopt an altered mode of cleavage to produce 26-nt RNAs. This is currently the strongest piece of evidence to explain how DCR-1 is capable of generating 26G-RNAs from presumably longer RRF-3 dsRNA precursors. It also implies that a blunt end must be produced by

processing of the sense template strand in a way that results in the alignment of its 3' end with the 5'G of the antisense strand.

The function of the helicase DRH-3 and of the dsRNA-binding protein RDE-4 in 26G-RNA biogenesis is less clear. DRH-3 is found in association with all RdRPs in *C. elegans* (Duchaine *et al.*, 2006; Gu *et al.*, 2009; Claycomb *et al.*, 2009; Thivierge *et al.*, 2012). Accordingly, it is essential for the accumulation of both 22G- and 26G-RNAs (Gu *et al.*, 2009; Vasale *et al.*, 2010). Current thought hypothesizes that due to an accumulation of a vestigial amount of 22G-RNAs at the 3' ends of transcripts in *drh-3* mutants, this putative helicase may either promote loading of 22G-RNAs into Argonautes by unwinding the RdRP product from the template, or facilitate translocation of the RdRP complex along the template by weakening regions of intra-molecular base pairing (Gu *et al.*, 2009). Regarding RDE-4, while it is required for accumulation of 26G-RNAs, the desilencing of their target RNAs in *rde-4* mutant animals is less pronounced than in other Eri mutants, suggesting that a basal amount of 26G-RNAs may still be produced (Vasale *et al.*, 2010; Gent *et al.*, 2010; Zhang *et al.*, 2011). Additionally, *rde-4* mutant animals exhibit a milder temperature-sensitive defect, laying mostly inviable embryos at 25˚C (Blanchard *et al.*, 2011). This is an indication that RDE-4 may not be essential to generate 26G-RNAs, but may rather confer efficiency to the process as a co-factor of DCR-1.

ERI-1b is strictly required for 26G-RNA accumulation, as null mutants show a complete loss of 26G-RNAs accompanied by the characteristic phenotypes (Kennedy *et al.*, 2004; Duchaine *et al.*, 2006; Han *et al.*, 2009; Conine *et al.*, 2010; Gent *et al.*, 2010). ERI-1b contains the conserved DEDDh-like 3'-5' exonuclease domain, which is able to catalyze the removal of overhanging nucleotides from duplex siRNAs with 2-nt 3' overhangs *in vitro*, but does not degrade single-stranded RNA (ssRNA) or the ends of a large RNA hybridized to a small RNA (Kennedy *et al.*, 2004). Its exonuclease activity was shown to be required for 26G-RNA accumulation (Gabel and Ruvkun, 2008). The mammalian ortholog of ERI-1b, Eri1, has been shown to play a role in the metabolism of histone mRNAs to promote their degradation at the end of the S phase of the cell cycle (Hoefig *et al.*, 2013). Eri1 achieves this by binding to the terminal stem-loop structure in histone mRNAs, followed by the removal of two unpaired nucleotides at the 3' end. The subsequent addition of a poly-uridine tract to the 3' end prompts the recruitment of a complex that interacts with Eri1, which in turn is able to degrade the stem-loop nucleotide by nucleotide, making the transcript vulnerable to general RNA degradation mechanisms (namely via the exosome from the 3' end and through decapping followed by Xrn1-

homolog exonucleolytic degradation from the 5' end). Although not yet explored, it is plausible that *C. elegans* ERI-1b may be recruited to 26G-RNA template regions of secondary structure and that, after RRF-3 synthesis, it helps to create the blunt end that DCR-1 requires for cleavage of 26-nt RNAs.

ERI-3 is a nematode-specific protein without sequence homology to any known domains. However, a null mutant shows the characteristic phenotypes associated with loss of 26G-RNAs (Duchaine *et al.*, 2006). Since it directly interacts with DCR-1, but not with the RdRP module of the ERI complex, it is thought to mediate the interaction between ERI-1b and DCR-1 (Thivierge *et al.*, 2012). ERI-5 has been shown to also bind DCR-1 directly and to bridge the RdRP module to the DCR-1 module of the ERI complex (Thivierge *et al.*, 2012). Loss of ERI-5 leads to a dramatic, but not complete, loss of 26G-RNAs, because the protein EKL-1 (a paralog of ERI-5) can partially compensate for its absence by interacting with DRH-3 bound to RRF-3. In this context the DCR-1 module no longer interacts with the rest of the complex, but can apparently still transiently engage the RRF-3/EKL-1/DRH-3 module (Thivierge *et al.*, 2012).

The 26G-RNA biogenesis pathway is still poorly-characterized. The biochemical activity of most components has either not been demonstrated or it is not known at all. How is the ERI complex recruited to its RNA targets? Which enzyme is responsible for cleavage and trimming of the target? How long and how stable are RRF-3 dsRNA precursors? These and other questions are worthy of further research, and should provide insight into how RdRPs and Dicer function together in other organisms.

### 26G-RNAs Occur as Two Distinct Populations Defined by the ERGO-1 and ALG-3/4 Argonautes

26G-RNAs associate with dedicated Argonautes that define two functionally distinct pathways: The ERGO-1 pathway and the pathway defined by the redundant ALG-3 and ALG-4 Argonautes (Fig. 1.5 and Table 1.1). These pathways comprise two phases: a primary phase characterized by the production of primary 26G-RNAs, and a secondary, amplification phase during which targets are bound by 26G-RNAs to trigger the synthesis of effector 22G-RNAs (Gent *et al.*, 2009; Han *et al.*, 2009; Conine *et al.*, 2010; Gent *et al.*, 2010; Vasale *et al.*, 2010).

ERGO-1 (*e*ndogenous *R*NAi-deficient Ar*go*naute 1), a PIWI-clade Argonaute, is predominately expressed in the cytoplasm of cells during oogenesis and embryogenesis, and residually expressed during the L3 and L4 stages (Vasale *et al.*, 2010; Billi *et al.*,

2012). The *ergo-1* null mutant manifests the Eri phenotype, but not the sperm-related sterility at 25˚C (Han *et al.*, 2009; Pavelec *et al.*, 2009). Molecularly, ERGO-1 26G-RNAs are different than ALG-3/4 26G-RNAs in that they carry a 3'-end 2'-*O*-methyl modification imparted by HENN-1. This feature was shown to confer extra stability to the ERGO-1 26G-RNAs, as a *henn-1* loss-of-function mutant expressed fewer 26G-RNAs and a higher frequency of untemplated 3' uridines on the remaining 26G-RNAs, which presumably promote their degradation (Kamminga *et al.*, 2012; uridylation effects reviewed in Lee *et al.*, 2014). Curiously, while the global stability of 21U-RNAs is only subtly affected in the *henn-1* mutant, 21U-RNA-induced silencing is compromised (Kamminga *et al.*, 2012). This suggests that the 2'-*O*-methyl modification may have a role beyond sRNA stabilization, perhaps as a specificity determinant for loading into the correct Argonaute.

In addition to the ERI complex, ERGO-1 26G-RNA accumulation is dependent on a set of extra proteins which are dispensable for ALG-3/4 26G-RNA accumulation. ERI-9 has been recovered in immunoprecipitates of DCR-1 and ERI-1 (Thivierge *et al.*, 2012), and its loss leads to the same phenotypes exhibited by *ergo-1* mutant animals (Pavelec *et al.*, 2009). ERI-9 is not conserved beyond nematodes and its function is currently uncharacterized. ERI-6/7 is a cytoplasmic helicase which has never been identified in studies of the ERI complex (Fischer *et al.*, 2008; Fischer *et al.*, 2011). Based on homology with human Mov10, which associates with Argonautes (Meister *et al.*, 2005), and *Drosophila* Armitage, which is required for the formation of RNA-induced silencing complexes (Tomari *et al.*, 2004), ERI-6/7 is proposed to associate with ERGO-1 to perhaps facilitate its loading or activity.

Finally, ERGO-1 26G-RNAs strictly require some proteins belonging to the Mutator group (Zhang *et al.*, 2011). Mutants from this class of proteins were found in screens in which the rate of mutation induced by transposition of sequences of the Tc1/*mariner* family in the germline was increased (Ketting *et al.*, 1999; Vastenhouw *et al.*, 2003). Some of these mutants are also RNAi-defective, highlighting the convergence of different sources of RNA into the central 22G-RNA pathway. The Mutator proteins required for ERGO-1 26G-RNAs are MUT-2 (or RDE-3), a nucleotidyltransferase mentioned earlier, MUT-7, a 3'-5' exonuclease, MUT-15, and MUT-16, both with unknown biochemical activities (Ketting *et al.*, 1999; Vastenhouw *et al.*, 2003; Chen *et al.*, 2005). MUT-16 is responsible for the aggregation of these proteins in cytoplasmic, perinuclear regions of germline cells, termed Mutator foci (Phillips *et al.*, 2012). The dependence of ERGO-1 26G-RNAs on these proteins therefore suggests that these foci are

the primary sites for their biogenesis. Mutator foci normally locate adjacently to P granules, but they assemble independently of P granules (Phillips *et al.*, 2012). In the germline, the majority of exported mRNAs have to pass through P granules (Sheth *et al.*, 2010). These highly dynamic sites contain numerous RNA-related proteins functioning in diverse processes such as RNAi (DRH-3, EGO-1, CSR-1, PRG-1 and WAGO-1) and regulation of mRNA stability and translation, among others (reviewed in Updike and Strome, 2009). It is therefore widely accepted that in the germline, P granules constitute the domain where most sRNAs first meet their targets, as they emerge from nuclear pores.

The majority of 26G-RNAs associate with ERGO-1 to generate WAGO-loaded 22G-RNAs, that specifically silence a restricted number of genes in *cis* (78 in embryos and 75 in adults, with a 60-gene overlap). These include protein coding genes, pseudogenes and unannotated transcription units (Vasale *et al.*, 2010; Fischer *et al.*, 2011). Curiously, two-thirds of these genes share regions of high sequence identity, suggesting that this pathway evolved to downregulate the expression of recently duplicated non-conserved genes (Fischer *et al.*, 2011). Examples of ERGO-1 acting in *trans* (*i.e.*, on RNAs other than their template RNAs) have also been identified, in which loci from which no 26G-RNAs are produced, show an ERGO-1-dependent accumulation of 22G-RNAs (Vasale *et al.*, 2010; Montgomery *et al.*, 2012). As *trans*-acting 26G-RNAs do not exhibit perfect complementarity with their targets, this implies that the ERGO-1 pathway may influence the expression of a much more vast set of genes (Montgomery *et al.*, 2012).

The biological relevance of the ERGO pathway is unknown, as *ergo-1* mutant animals do not exhibit any noticeable developmental or physiological abnormalities, at least under laboratory growth conditions. Despite targeting very few genes, ERGO-1 26G-RNAs (and the resulting 22G-RNAs) are very abundant during embryogenesis and early larval development (Gent *et al.*, 2010). The Eri phenotype is solely manifested in mutants that are essential for ERGO-1 26G-RNA biogenesis and accumulation. This effect is thought to arise from the availability of somatic RNAi machinery that is normally engaged in ERGO-1 target silencing. This permits the processing of a larger amount of exogenous dsRNA which in turn elicits more potent silencing of cognate targets. Several groups have observed this competition phenomenon between endogenous and exogenous silencing pathways due to the limited quantity of RNAi machinery (Kennedy *et al.*, 2004; Duchaine *et al.*, 2006; Lee *et al.*, 2006; Yigit *et al.*, 2006; Guang *et al.*, 2008; Gu *et al.*, 2009; Sarkies *et al.*, 2013).

The ALG-3 and ALG-4 Argonautes (belonging to the Argonaute-clade) and their associated 26G-RNAs are expressed exclusively during spermatogenesis (Han *et al*., 2009; Colin *et al.*, 2010). In hermaphrodites, spermatogenesis occurs from the late L4 stage until early adulthood, and in males from the L4 stage onwards. ALG-3 is enriched in the P granules of developing spermatocytes, as well as in the cytoplasm surrounding them, but absent from mature spermatids (Colin *et al.*, 2010, Colin *et al.*, 2013). *alg-3; alg-4* double mutants (but not single mutants) exhibit the fully penetrant temperature-sensitive sterility phenotype that characterizes many of the Eri mutants, while not manifesting the ERGO-1-dependent Eri phenotype (Han *et al*., 2009; Colin *et al*., 2010). The sterility is due to a variety of sperm defects, such as a decreased production of sperm, defects in nuclear shape of developing spermatocytes (frequently arrested as multinucleated cells or with chromatin bridges that fail to resolve during budding of spermatids), and the inability to be activated into the motile stage that allows fertilization to occur (Gent *et al.*, 2009; Pavelec *et al.*, 2009; Conine *et al.*, 2010).

In males, 26G-RNAs target as many as ~1,400 genes, 63% of which exhibit sperm-specific expression (Conine *et al.*, 2013). While ~200 of the mRNAs for these genes are downregulated in *alg-3; alg-4* mutant animals, just as many mRNAs are upregulated. In wild-type males grown at 25˚C the expression of most positively regulated genes increased, whereas the expression of most negatively regulated genes decreased. The same tendency was verified at the protein level. Genes that are positively regulated by ALG-3/4 26G-RNAs are mostly involved in sperm activation, motility and general spermatogenesis (Conine *et al.*, 2013). This outcome explains why ALG-3/4 26G-RNAs are required for proper sperm function at elevated temperatures, as mutants for some of the upregulated genes exhibit the temperature-sensitive sterility phenotype. On the other hand, negatively regulated genes are not enriched for sperm-related functions. As their decrease was found to depend on the expression of WAGO Argonautes, it is assumed that their silencing occurs via 22G-RNAs (Conine *et al.*, 2010; Conine *et al.*, 2013). Interestingly, regarding positively regulated genes, it was found that the 22G-RNAs produced against them were loaded into the CSR-1 Argonaute (described in the next section), which correlated with the increased deposition of activating chromatin marks in spermatocytes developing at 25˚C (Conine *et al.*, 2013).

The ERGO-1 and ALG-3/4 pathways are illustrative of how the differential expression of Argonautes at the cellular, tissue and developmental levels can serve dramatically distinct regulatory purposes. The mechanisms by which ERGO-1 and ALG-

3/4 in complex with 26G-RNAs target their transcripts and recruit secondary RdRP machinery are currently uncharacterized. In the future it will be very interesting to know whether this process differs for each pathway, considering that ERGO-1 is phylogenetically closer to the 21U-RNA PRG-1 PIWI Argonaute, and that ALG-3 and ALG-4 are closer to the miRNA-specific ALG-1 and ALG-2 Argonautes (Youngman and Claycomb, 2014).

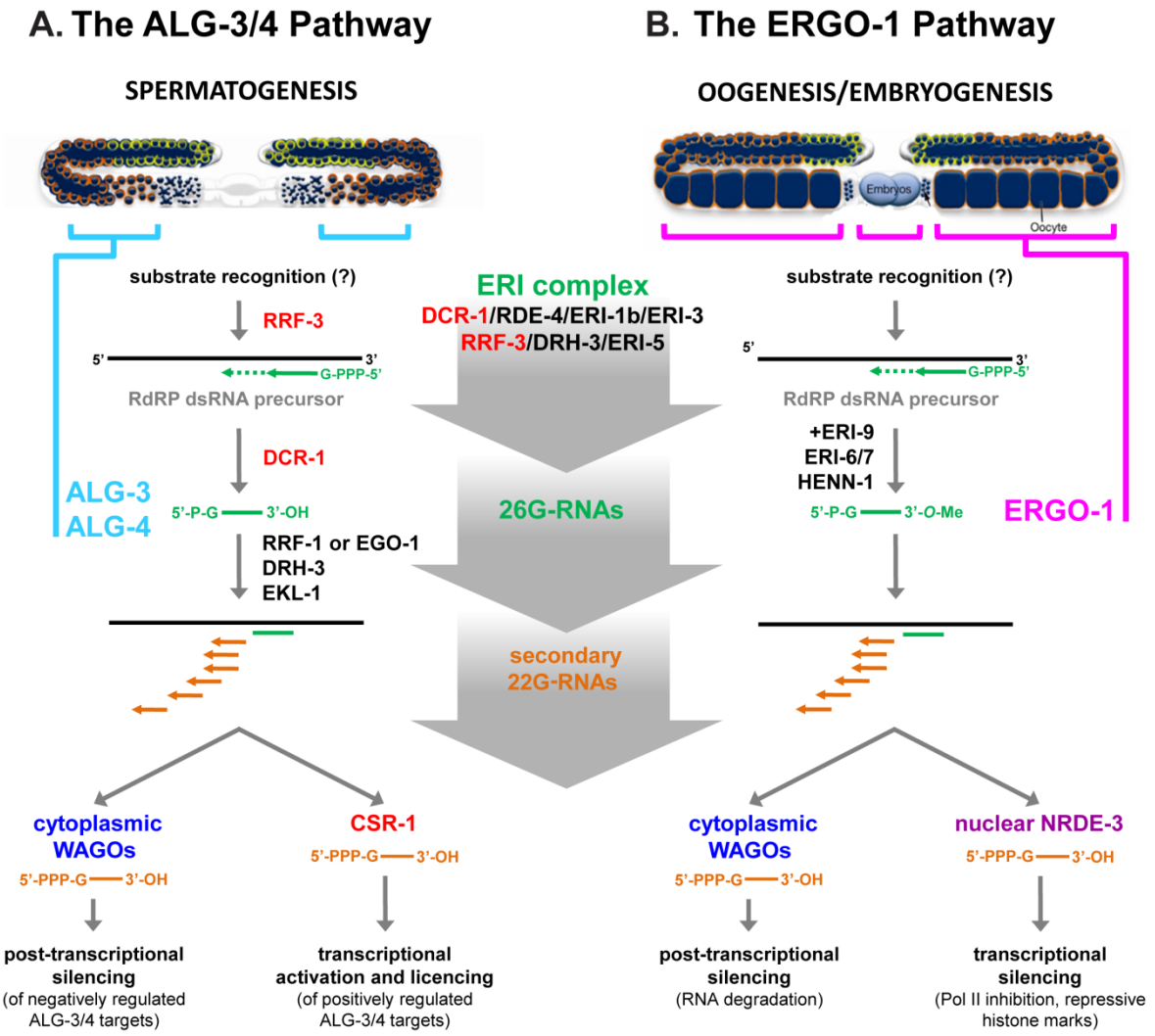



**Figure 1.5. The ALG-3/4 and the ERGO-1 26G-RNA pathways.** Both pathways generate primary 26G-RNA primary siRNAs through the concerted activity of proteins grouped in the ERI complex. It is not known what characteristics of the substrates make them recognizable to the ERI complex. The RdRP RRF-3 synthesizes a complementary strand of unknown length (dashed green line), starting with a presumably triphosphorylated 5'G. Through as of yet uncharacterized processing steps, it is predicted that a blunt ended dsRNA at the RRF-3 initiation site is produced. This provides a substrate for cleavage by Dicer, to generate a non-canonical product of 26 nt. These 26G-RNAs are loaded into specialized Argonautes that define two distinct pathways. **(A)** The

ALG-3 and ALG-4 Argonautes are expressed during spermatogenesis of hermaphrodites and males. They guide the production of secondary 22G-RNAs that can have a post-transcriptional silencing effect on targets, or activate the expression of a subset of targets that are specifically related to the development and activation of sperm cells. These engage the CSR-1 Argonaute pathway, which reinforces the expression of the corresponding genes and permits the paternal inheritance of gene expression information via the incorporation of loaded CSR-1 complexes in the mature sperm of males. **(B)** ERGO-1 protein expression predominately takes place in oocytes and embryos. In addition to the ERI complex, ERGO-1 26G-RNA accumulation requires additional factors that are dispensable by ALG-3/4 26G-RNAs. These include the RNA methyltransferase HENN-1, which modifies the 3' ends of the 26G-RNAs in order to promote their stability. The resulting 22G-RNAs predominately silence a small group of ~80 highly expressed genes in somatic cells of embryos and post-embryonic larvae, encoding mostly pseudogenes and duplicated sequences with no apparent biological functions. There are currently no clues suggesting the mechanism whereby 26G-RNAs recruit RdRPs to their targets.

## 22G-RNAs Regulate Gene Expression by Remarkably Diverse Mechanisms: WAGO *versus* CSR-1 Pathways

All sRNA pathways, whether they are exogenously or endogenously triggered, converge on the production of effector, RdRP-synthesized antisense 22G-RNAs. The regulatory outcome of 22G-RNAs depends largely on which downstream WAGO-clade Argonautes they interact with (Table 1.1). There are two main 22G-RNA pathways determined by whether they interact with any of the 12 semi-redundant WAGO-clade Argonautes, or with the CSR-1 Argonaute (also a member of the WAGO clade). For simplicity we refer to these two pathways as the WAGO pathway and the CSR-1 pathway. The simultaneous deletion of the 12 WAGOs leads to a loss of the majority of 22G-RNAs, except for a low amount of germline 22G-RNAs associating specifically with CSR-1 (Gu *et al.*, 2009; Claycomb *et al.*, 2009). Globally, ~50% (~9,000) of all protein-coding genes in *C. elegans* are targeted by 22G-RNAs from both pathways and quantitatively, the majority of endogenous 22G-RNAs are produced in the germline (Gu *et al.*, 2009). The WAGO pathway employs transcriptional and post-transcriptional mechanisms to silence protein-coding genes, transposable elements, repetitive sequences, pseudogenes and other unannotated expressed intergenic sequences (Gu *et al.*, 2009). The CSR-1 pathway targets essentially all germline-expressed protein coding genes (~4,000), not to silence them, but rather to promote their expression and to protect them from the silencing effects of the PRG-1/21U-RNA pathway, in a process termed "licensing" (Fig. 1.4B, page 42 and Table 1.1, page 29; Claycomb *et al.*, 2009; Gu *et al.*, 2009; Seth *et al.*, 2013; Wedeles *et al.*, 2013; Cecere *et al.*, 2014).

WAGO pathway 22G-RNAs are triggered by the sRNAs associated with RDE-1, PRG-1, ERGO-1 and ALG-3/4, all of which can be considered primary sRNAs. Since PRG-1 and ALG-3 have been shown to concentrate in nuclear pore-associated P granules, these are thought to be the main sites where primary Argonaute complexes associate with their targets in the germline (Batista *et al.*, 2008; Conine *et al.*, 2010). WAGO 22G-RNAs are then amplified in adjacent Mutator foci, which harbor EGO-1 and RRF-1 RdRP complexes, as well as a variety of additional factors required for the robust accumulation of 22G-RNAs (Phillips *et al.*, 2012, Yang *et al.*, 2012; Zhang *et al.*, 2012). In contrast, the production of CSR-1 22G-RNAs does not require Mutator foci (Zhang *et al.*, 2011). Rather, it depends mostly on EGO-1-mediated synthesis in P granules, where both proteins co-localize (Claycomb *et al.*, 2009; Gu *et al.*, 2009). In somatic cells, which lack perinuclear P granules, the cellular location of these events has not been elucidated. However, based on the observation that the proteins that aggregate in Mutator foci are equally required for RNAi in both germline and somatic cells, 22G-RNA synthesis and accumulation is presumed to take place in cytoplasmic P-body-like domains (Phillips *et al.*, 2012).

The mechanism of 22G-RNA-mediated post-transcriptional silencing is still poorly understood. A reason for this knowledge gap stems from the fact that WAGO-pathway Argonautes lack the slicer-activity conserved residues that would allow cleavage of target RNAs (Yigit *et al.*, 2006). The recent discovery of additional factors necessary for exo-RNAi provided new ideas to explain how post-transcriptional silencing may occur in *C. elegans*. A complex composed of nematode-specific proteins RDE-10 and RDE-11 with potential RNA binding and nuclease properties, respectively, was shown to be required for RDE-1- and ERGO-1-driven 22G-RNA amplification in both germline Mutator foci and the cytoplasm of somatic cells (Yang *et al.*, 2012, Zhang *et al.*, 2012). Notably, RDE-10 was shown to bind exo-RNAi targeted mRNAs, promoting their deadenylation and subsequent RDE-11-dependent degradation (Yang *et al.*, 2012). In another study from our laboratory, the endonuclease RDE-8 (related to the human Zc3h12a RNase) was shown to exhibit a similar pattern of localization as RDE-10/RDE-11 and to generally promote the accumulation of both endogenous 22G-RNAs and of exo-RNAi triggered 22G-RNAs, both in the germline and soma (Tsai *et al.*, 2015). In association with the nucleotidyltransferase RDE-3, target RNAs are cleaved by RDE-8 and untemplated uridines are added to the 3' end of the 5'-cleavage fragment. This event may trigger RdRP recruitment for 22G-RNA synthesis but also lead to the subsequent degradation of the mRNA fragment, as 3'

uridylation promotes RNA decay (Lee *et al.*, 2014). The degradation activity of these complexes, together with the potential recruitment of mRNAs by loaded WAGOs to sites where canonical mRNA degradation pathways operate (P granules, cytoplasmic P-bodies, etc…) may collectively contribute to the post-transcriptional silencing of exo- and endo-RNAi targets. Further work will be required to understand these processes in detail.

At the transcriptional level, silencing can occur via two WAGO Argonautes capable of shuttling between the nucleus and the cytoplasm: NRDE-3(WAGO-12) and HRDE-1(WAGO-9). NRDE-3 (*N*uclear *R*NAi-*de*fective) mediates nuclear RNAi in somatic cells, while HRDE-1 (*H*eritable *R*NAi-*de*fective) acts in germline nuclei (Guang *et al.*, 2008; Buckley *et al.*, 2012). The association of these Argonautes with exogenously and endogenously triggered 22G-RNAs, causes their translocation into the nucleus. In cooperation with other factors, namely NRDE-1, NRDE-2 and NRDE-4, they are able to simultaneously interact with chromatin and target nascent transcripts to inhibit Pol II elongation, and subsequently to establish a silent chromatin state through the deposition of repressive histone H3K9 trimethylation marks (Guang *et al.*, 2008; Guang *et al.*, 2010; Burkhart *et al.*, 2011; Burton *et al.*, 2011; Ashe *et al.*, 2012; Buckley *et al.*, 2012; Gu, S.G., *et al.*, 2012; Luteijn *et al.*, 2012; Shirayama *et al.*, 2012).

NRDE-3 binds endogenous 22G-RNAs triggered by ERGO-1 26G-RNAs, evidenced by the fact that when the ERI complex is disabled, NRDE-3 becomes predominately cytoplasmic (Guang *et al.*, 2008; Pavelec *et al.*, 2009; Fisher *et al.*, 2011). In this context, following exposure to dsRNA targeting any somatically expressed gene, NRDE-3 reacquires a nuclear localization, showing that it can also act in response to exogenous triggers (Guang *et al.*, 2008). NRDE-3 may therefore cooperate with other secondary WAGO Argonautes to robustly silence genes via simultaneous transcriptional and post-transcriptional mechanisms. Consistent with an effector role in the ERGO-1 pathway, *nrde-3* loss-of-function mutants appear to be wild-type. In contrast, mutants of the nuclear counterpart *hrde-1*, exhibit a mortal germline phenotype (Mrt) at 25˚C in which the ability of the germline to produce oocytes and sperm decreases gradually over a few generations, ultimately resulting in animals that are completely sterile (Buckley *et al.*, 2012). In agreement with this phenotype, HRDE-1 was demonstrated to direct transgenerational silencing of germline-expressed transcripts targeted by exo-RNAi by maintaining H3K9 trimethylation on the corresponding genes (Ashe *et al.*, 2012; Buckley *et al.*, 2012; Gu, S.G. *et al.*, 2012). Cloning of HRDE-1-associated endogenous 22G-RNAs revealed a large overlap with the 22G-RNAs associated with the P-granule WAGO-1

Argonaute, confirming that 22G-RNAs targeting the same gene can simultaneously execute transcriptional and post-transcriptional gene silencing (Shirayama *et al.*, 2012). Since WAGO-1 shows a preference for 22G-RNAs targeting transposons and repetitive elements (Gu *et al.*, 2009), the Mrt phenotype of the *hrde-1* mutant is thought to arise from the inability to maintain repressive epigenetic marks on these elements, which in turn leads to the accumulation of mutations incompatible with stem cell maintenance.

The Argonaute CSR-1 (*C*hromosome *S*egregation and *R*NAi-defective) is highly expressed in the germline, oocytes and embryos (Claycomb *et al.*, 2009). As its name indicates, *csr-1* mutants display lethal chromosome segregation defects during mitotic divisions in embryos, exhibit partial resistance to exo-RNAi in the germline, as well as changes in the morphology and proliferation of germline nuclei (Yigit *et al*, 2006; Claycomb *et al.*, 2009). CSR-1 concentrates throughout the germline in P granules, localizes to meiotic chromosomes of oocytes and later to mitotic chromosomes of developing embryos (Claycomb *et al.*, 2009). CSR-1 22G-RNAs depend strictly on the RdRP EGO-1, DRH-3 and EKL-1, mutants of which exhibit several overlapping phenotypes with *csr-1* (Smardon *et al.*, 2000; Vought *et al.*, 2005; Duchaine *et al.*, 2006; Nakamura *et al.*, 2007; Gu *et al.*, 2009; She *et al.*, 2009). As mentioned earlier, CSR-1 22G-RNAs target germline protein-coding mRNAs without leading to their downregulation, in contrast to the silencing effects of the 12 semi-redundant Argonautes composing the WAGO pathway (Claycomb *et al.*, 2009). Consistent with this, the levels of CSR-1 22G-RNAs are much lower than those of the WAGO pathway (Gu *et al.*, 2009). One identified reason for their lower abundance is the 3' poly-uridylation of a large portion of CSR-1 22G-RNAs by the conserved nucleotidyltransferase CDE-1 (*C*osuppression *De*fective 1; also known as CID-1, *C*affeine *I*nduced *D*eath Homolog of *S. pombe* 1, or PUP-1, *P*oly(*U*) *P*olymerase 1), which leads to their destabilization (van Wolfswinkel *et al.*, 2009).

The CSR-1 pathway has recently been shown to promote Pol II transcription of the germline expressed genes it targets (Cecere *et al.*, 2014). This is consistent with prior observations that CSR-1 regulates the distribution of repressive H3K9 dimethyl chromatin marks (Maine *et al.*, 2005; She *et al.*, 2009), and that on embryonic chromatin, regions bound by CSR-1 and regions bound by the centromere-defining protein CENP-A (*Cen*tromere *P*rotein A) are mutually exclusive (Claycomb *et al.*, 2009; Gassmann *et al.*, 2012). These findings suggest that CSR-1 22G-RNAs may maintain zones of euchromatin among centromeres (which in *C. elegans* are found along the entire length of

chromosomes) to perhaps allow the maintenance of gene expression required for successful cell division and proper development of early embryos. Furthermore, CSR-1 was shown to counteract the silencing effects of the PRG-1/21U-RNA pathway in the germline (Lee *et al.*, 2012; Seth *et al.*, 2013; Wedeles *et al.*, 2013). These studies demonstrated that both transgenes and endogenous genes targeted by CSR-1 are completely immune to recognition and silencing by PRG-1. This immediately suggested the hypothesis that the CSR-1 and the PRG-1 pathways act simultaneously and in opposition of each other, as part of a system that discriminates "self" from "non-self" gene expression (Fig. 1.4., page 42). This system would ensure proper germline function while protecting its genetic material from the deleterious effects of invading or existing foreign sequences (*e.g.*, retroviruses and transposable elements).

Lastly, very little is known regarding the events that trigger CSR-1 22G-RNA synthesis. The first clue that some CSR-1 22G-RNAs may also be triggered by primary sRNAs came from a recent study of CSR-1 in males (Conine *et al.*, 2013). CSR-1 was shown to interact with 22G-RNAs derived from ALG-3/4 target mRNAs, and to promote the transcription of those mRNAs in developing spermatocytes. Accordingly, *csr-1* mutant males were shown to have the same temperature sterility phenotype of *alg-3; alg-4* mutant males. When *csr-1* homozygous males were mated to heterozygous hermaphrodites for successive generations, their fertility decreased gradually until the sixth generation, which was completely sterile. Since 22G-RNA-loaded CSR-1 was found in mature sperm, it was proposed that CSR-1 passes along 22G-RNAs that ensure proper expression of ALG-3/4-class genes in the next generation (Conine *et al.*, 2013).

The aforementioned examples illustrate an elevated degree of interconnectedness between sRNA pathways in *C. elegans*, which we have just recently begun to realize. Researchers still hold many loose pieces of this incredibly large and intricate puzzle. The discoveries made thus far in *C. elegans*, particularly those that implicated sRNAs as inheritable agents that propagate epigenetic programs from one generation to the next, have propelled an enthusiastic search for similar mechanisms in vertebrates.

# OVERVIEW OF DICER-ASSOCIATED *C. ELEGANS* PIR-1 AND ITS RNA 5' PHOSPHATASE ORTHOLOGS

## *C. elegans* PIR-1

In 2006 our laboratory published a study aiming to complement previous sRNA pathway screens in *C. elegans* through a mass spectrometry-based proteomic survey of DCR-1 immunoprecipitates (Duchaine *et al.*, 2006). At the time, the role of Dicer in the RNAi and miRNA pathways was already well characterized. The study aimed not only to expand the list of proteins necessary for these pathways, but also to discover new functions for Dicer. The chosen experimental approach had the advantage of allowing the identification of proteins also required for viability, which tend to be missed in traditional forward and reverse genetic screens. The results confirmed previously described Dicer interactors and provided the basis for the discovery of the ERI complex and its functions. Additionally, two essential proteins were uncovered: the helicase DRH-3, implicated in RNAi and fertility, and PIR-1, named after the human ortholog PIR1 (*P*hosphatase that *I*nteracts with *R*NA/RNP complex 1; Yuan *et al.*, 1998), which acts as an RNA phosphatase *in vitro*.

The characterization of a *C. elegans pir-1* null deletion mutant in the aforementioned study showed that PIR-1 is essential, as the animals exhibited a fully penetrant larval arrest before reaching reproductive age. The study also found that *pir-1* mutant animals were unable to silence genes via RNAi by not producing siRNAs. It was therefore proposed, based on its putative RNA 5' phosphatase activity, that PIR-1 could dephosphorylate the termini of long dsRNA products generated by the RdRPs RRF-1 and EGO-1 to allow Dicer-mediated cleavage of the dsRNA into effector siRNAs. This model has since been abandoned, however, in light of subsequent work from other laboratories showing that RRF-1 and EGO-1 directly generate effector 22G secondary siRNAs from template RNAs without the need for additional processing steps (Aoki *et al.*, 2007; Pak and Fire, 2007; Sijen *et al.*, 2007).

As the initial characterization of a *pir-1* loss-of-function mutant in Duchaine *et al.* (2006) did not exhaustively address the involvement of PIR-1 in sRNA biogenesis and development, numerous important questions were left unanswered. For instance, at which step of the RNAi process is PIR-1 required for the accumulation of 22G-RNAs? Since miRNAs, 26G-RNAs and 21U-RNAs all bear a 5'-monophosphate, is the putative RNA phosphatase activity of PIR-1 required to achieve this phosphorylation state? Does PIR-1

function as an RNA 5' phosphatase *in vivo*? Is PIR-1 associated with every Dicer molecule or just a subset? These and other questions motivated the continued study of PIR-1. From the onset of this project, the pioneering studies of the baculoviral and human orthologs of PIR-1 provided an invaluable basis for our investigation. For this reason, in the following two sections, we provide a brief review about the current state of knowledge concerning these proteins.

## Baculovirus Phosphatase

Interest in human PIR1 was prompted by earlier studies of Baculovirus Phosphatase, or BVP, with which it shares 58% similarity at the amino acid level (Fig. 1.6). This protein is encoded by the circular dsDNA baculovirus *Autographa californica* nuclear polyhedrosis virus, which carries over 100 genes and replicates in the nuclei of lepidopteran insect cells. BVP was identified based on its homology to eukaryotic protein tyrosine phosphatases or PTPs (Sheng and Charbonneau, 1993). The human genome encodes more than 100 PTPs required for the regulation of various processes, such as transcription, mRNA processing, cell proliferation, and signaling pathways, to ensure, for example, faithful development and appropriate immune responses (reviewed in Alonso *et al.*, 2004). PTPs exhibit enormous diversity in size and structural organization, occurring as receptor-like transmembrane forms or smaller non-receptor forms. They all share a highly conserved active site motif – HC*XXXXX*R(S/T) – termed the P-loop (Phosphatase loop), in which the cysteine acts as the nucleophile that disrupts the phosphomonoester bond to release the protein phosphate (Zhang , 2003).

BVP is one of the smallest known PTPs (~19 kDa), and has been shown to possess tyrosine phosphatase and serine/threonine phosphatase activities *in vitro* (Kim and Weaver, 1993; Sheng and Charbonneau, 1993). Studies of infection in insect SF-21 cells established that BVP is expressed late in infection, when DNA replication occurs, localizing to the nucleus, cytoplasm and viral particles (Kim and Weaver, 1993; Li and Miller, 1995a). The infection of cells with a BVP-deletion mutant virus led to overall reduced viral titers (>50%) and heterogeneity among cells regarding the production of occluded viral particles, produced very late in the infective cycle to mediate transmission from one host to another (Li and Miller, 1995a; Li and Miller, 1995b). Intriguingly, when insect larvae were infected with either mutant or wild-type occluded viruses, the concentration of virus and the time required for animals to die did not differ between the two conditions (Li and Miller, 1995b). Curiously, the BVP ortholog of a silkworm baculovirus was demonstrated

to be essential for virally induced light-activated locomotory behavior, which leads the animals to seek the tops of vegetation thereby facilitating virus dispersion (Kamita *et al.*, 2005). This observation suggests that BVP may be required to modulate aspects of host physiology (perhaps by altering host gene expression), rather than to promote viral replication.

BVP (as well as PIR-1) exhibits more extensive sequence homology with the triphosphatase domain of metazoan mRNA capping enzymes, than to PTPs (Fig. 1.6; Takagi *et al.*, 1997). Studies of the *C. elegans* capping enzyme CEL-1 (mRNA *C*apping *E*nzyme-*L*ike 1) demonstrated that its RNA triphosphatase domain removes the γ-phosphate from the 5'-ends of RNA *in vitro*, but that it lacks phosphotyrosine or phosphoserine activity (Takagi *et al.*, 1997; Takagi *et al.*, 2003). Based on these findings, the *in vitro* activity of BVP was further scrutinized, leading to the discovery that it sequentially removes the γ- and β-phosphates from the triphosphorylated 5'-end of RNA molecules, resulting in a 5'-monophosphorylated molecule (Takagi *et al.*, 1998). In contrast, dephosphorylation activity of BVP on phosphoprotein substrates was found to be two to three orders of magnitude lower than that of the budding yeast PTP Cdc14 (*C*ell *d*ivision *c*ycle 14). Finally, the substitution of the active site cysteine with a serine resulted in the complete loss of both activities.

As the mechanism of RNA dephosphorylation by metazoan RNA triphosphatases had not yet been elucidated, BVP was used as a model to solve this problem. It was shown that, similarly to the mechanism of protein dephosphorylation, BVP also forms a covalent protein-phospho-cysteinyl intermediate as a result of RNA 5'-phosphate hydrolysis (Martins and Schuman, 2000). The active site cysteine thiolate acts as the nucleophile that attacks the bond between two phosphates. The covalent phospho-cysteinyl intermediate is then hydrolyzed to release the inorganic phosphate. Further insights came from extensive mutational analyses of BVP, which led to the identification of several active-site and structural residues required for its triphosphatase activity (Martins and Shuman, 2002a; Martins and Shuman, 2002b). Ultimately, the analysis of the crystal structure of BVP in complex with phosphate helped to rationalize the influence of these mutations on protein activity (Changela *et al.*, 2005). Moreover, the study revealed that there is extensive structural conservation between BVP and the mammalian capping enzyme triphosphatase domain, reinforcing its identity as a *bona fide* RNA 5' phosphatase. Finally, this comparison allowed the identification of the P-loop asparagine residue – essential for

formation of the phospho-cysteinyl intermediate – as the major structural signature of the active site of RNA 5' triphosphatases.

Despite extensive biochemical and structural characterization, the biological significance of BVP remains mysterious. Since it is expressed late in infection, and because BVP was shown to be able to functionally replace the RNA triphosphatase of *Saccharomyces cerevisiae* to allow mRNA capping *in vivo* (Martins and Shuman, 2002a), its influence in the capping of late viral transcripts was studied (Li and Guarino, 2008). The authors demonstrated that BVP is not required for capping of late transcripts. They also showed that LEF-4 (*L*ate Gene *E*xpression *F*actor 4), a putative viral capping enzyme carrying an N-terminal guanylyltransferase and C-terminal RNA triphosphatase domain, is not required for capping of late viral transcripts. This was surprising, given that LEF-4 is essential for viral replication and that the virus encodes no other capping enzymes. It was therefore proposed that the capping of late baculovirus transcripts is performed by a host enzyme, leaving ample room for speculation concerning the biological relevance of both BVP and LEF-4.

Baculoviruses are unique among DNA viruses in that they use the host Pol II for transcription of early genes, and their own RNA polymerase for late transcription (reviewed in Hasnain *et al*., 1997). In the late phase of infection the levels of host transcripts drop dramatically, but viral transcription continues. Considering this, one hypothesis is that BVP disables host transcription by dephosphorylating the ends of nascent mRNA before they can be capped by the host capping enzyme, perhaps ensuring a higher allocation of cell resources for viral replication (Takagi *et al.*, 1998). Another attractive possibility is that BVP could facilitate the replication of viral DNA, by dephosphorylating the 5'-ends of RNA replication primers to promote their removal and replacement with DNA (Takagi *et al.*, 1998). However, it is currently not known whether the phosphorylation state of replication primers influences their removal during replication.

### Human PIR1

PIR1 is a ~39 kDa protein, also commonly known as DUSP11 (*du*al-specificity protein *p*hosphatase 11). DUSPs constitute a subgroup within the PTP superfamily, unified by their ability to dephosphorylate both phosphotyrosine and phosphoserine/ phosphothreonine residues (reviewed in Patterson *et al.*, 2009). Similarly to BVP, PIR1 possesses the defining RNA 5' phosphatase catalytic domain HCTHG*X*NRT, that distinguishes it from its protein dephosphorylating family members (Changela *et al.*,

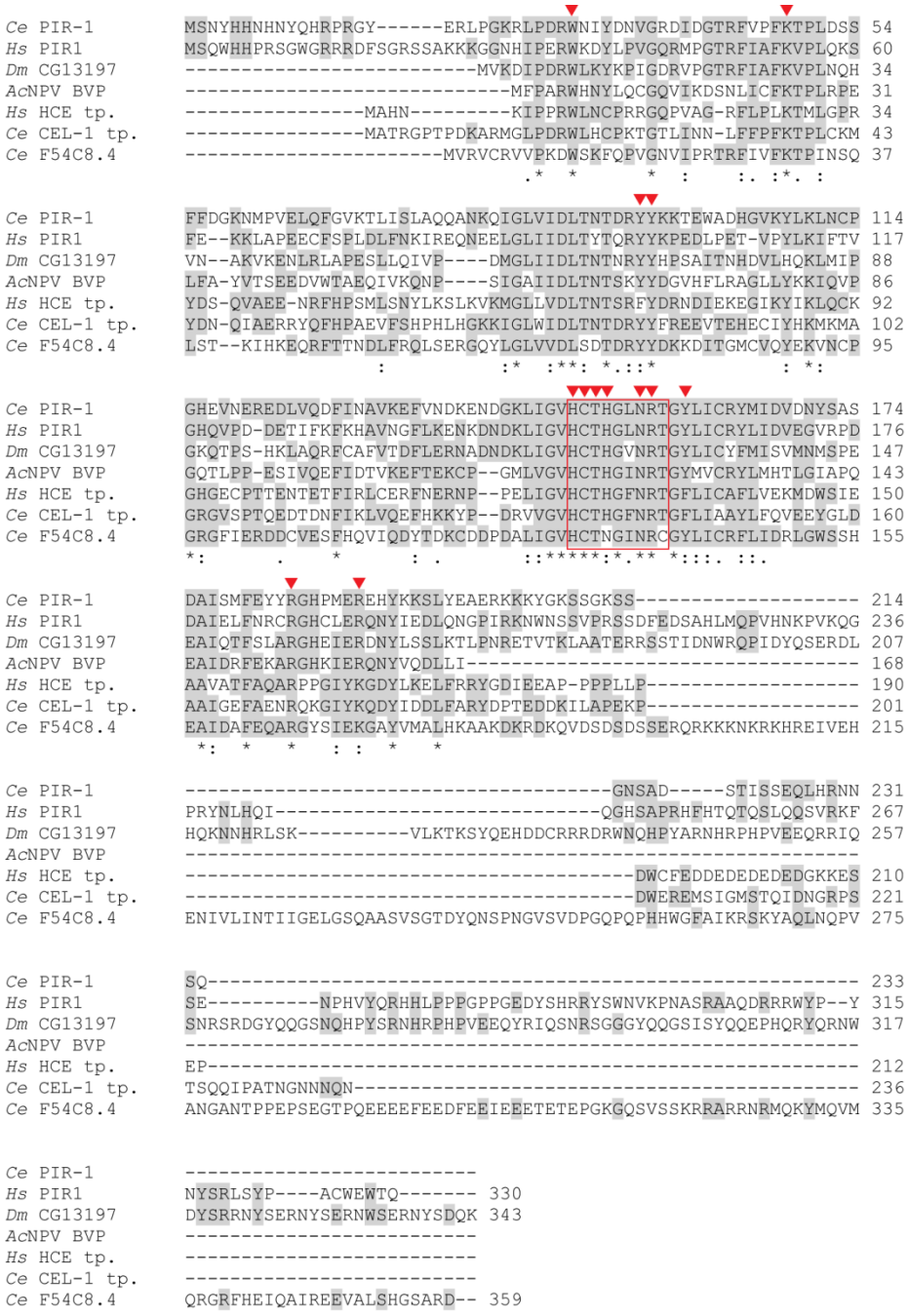
2005). PIR1 orthologs are found among all metazoans for which sequence information exists, but are conspicuously absent from plants, fungi, unicellular eukaryotes and prokaryotes. Beyond the P-loop, this enzyme shares no significant sequence similarity with other DUSPs and lacks any other recognizable functional motifs, apart from N-terminal and C-terminal arginine rich regions (Yuan *et al.*, 1998). It also shares significant sequence similarity with the triphosphatase domain of metazoan mRNA capping enzymes (Fig. 1.6; Deshpande *et al.*, 1999).

The establishment that PIR1 has a strong binding preference for RNA was followed by the demonstration that it recapitulates the *in vitro* biochemical activity of BVP on 5'-triphosphorylated RNA (Yuan *et al.*, 1998; Deshpande *et al.*, 1999). Again, the P-loop cysteine residue was strictly required for both the tri- and diphosphatase activities of PIR1, and the activity on phospho-protein substrates was even lower than the one measured for BVP, making it unlikely that it is involved in endogenous protein dephosphorylation. Recently, the crystal structure of PIR1 was solved, demonstrating a high degree of structural conservation with BVP and the human capping enzyme (Sankhala *et al.*, 2014). Examples of important structural features include (1) a wide, deep cleft at the bottom of which the P-loop is located, lined by positively charged residues (likely to neutralize the negatively charged RNA backbone and provide a good docking site for a di- or triphosphate), (2) two residues in the P-loop (the central histidine and the asparagine) that make close side contact with the incoming phosphate, and are thought to orient the attacking group (thiolate) and leaving group (phospho-RNA), and (3) the presence of an anion stabilizing the position of the basic arginine residue of the P-loop and making contact with the threonine next to the catalytic cysteine (Changela *et al.*, 2005; Sankhala *et al.*, 2014). Further work will be necessary to address important questions regarding the mechanisms of RNA 5' triphosphatases. As an example, it would be important to know which structural features determine that capping enzymes can only hydrolyze the γ-phosphate, while PIR1 and BVP can hydrolyze both γ- and β-phosphates.

Lastly, existing work provides some clues as to what the biological roles of PIR1 in humans may be. PIR1 was found to associate with three splicing factors through yeast two-hybrid screens in two studies, although only one interaction was validated *in vivo* (Yuan *et al.*,1998; Caprara *et al.*, 2009). These interactions are consistent with an experiment in HeLa cells, which revealed that overexpressed PIR1 accumulated primarily in nuclei and co-localized with the splicing factor SC35 (SRSF2, Splicing Factor, arginine/serine-rich 2) in nuclear speckles (Yuan *et al.*, 1998). These regions concentrate multiple pre-mRNA

processing factors and often overlap with domains of active transcription (reviewed in Spector and Lamond, 2011). Moreover, also in the context of cultured cells, expression of endogenous PIR1 was found to be under the control of the master regulator, tumor suppressor p53 protein, essential for the DNA damage response and cell cycle progression (Caprara *et al.*, 2009). In the same study, the overexpression or RNAi knockdown of *PIR1* led to marked underproliferation and overproliferation of cell populations, respectively. Three further studies found correlations between specific diseases and the expression of *PIR1* mRNA. *PIR1* was found to be upregulated in HT-29 colon carcinoma cells grown in culture as multicellular tumor spheroids, in comparison to HT-29 cells growing in monolayer (Dardousis *et al.*, 2007). In a study comparing the expression of 149 splicing factors in the mucosal tissue of patients with inflammatory bowel disease, *PIR1* mRNA was found to be significantly downregulated, compared to its expression in healthy tissue (Hasler *et al*., 2011). A more recent study aiming to identify host factors required for intracellular growth of the bacterium *Salmonella thyphimurium* by siRNA screening, concluded that *PIR1* is involved in the control of intracellular infection by this pathogen, as silencing of *PIR1* mRNA led to the inhibition of *Salmonella* growth in infected human cells (Harald *et al*., 2014). These results suggest that PIR1 may play important roles in RNA metabolism and cell division/proliferation, and that its misexpression may give rise or sensitize humans to various pathologies. Therefore, more detailed studies of this highly conserved protein in model organisms are warranted, and should provide us with knowledge of biological and medical relevance.

**A**




**B**

| | PIR-1 | hPIR1 | CG13197 | BVP | HCE | CEL-1 |
|---|---|---|---|---|---|---|
| **F54C8.4** | 58% | 59% | 59% | 59% | 58% | 51% |
| **CEL-1** | 52% | 50% | 49% | 55% | 64% | |
| **HCE** | 54% | 54% | 53% | 51% | | |
| **BVP** | 58% | 57% | 57% | | | |
| **CG13197** | 62% | 53% | | | | |
| **hPIR1** | 62% | | | | | |

**Figure 1.6. Sequence alignments of *C. elegans* PIR-1 to homologous proteins. (A)** Alignments were produced using ClustalW2 software (Larkin *et al.*, 2007; available at at the EMBL-EBI

website ebi.ac.uk/Tools). The following sequences were used: *C. elegans* PIR-1 (233 aa), human PIR1 (330 aa), *Drosophila melanogaster* CG13197 (343 aa), *Autographa californica* nucleopolyhedrosis virus BVP (168 aa), the triphosphatase domain (tp.) of the human capping enzyme HCE (residues 1-212), the triphosphatase domain of the *C. elegans* capping enzyme CEL-1 (residues 1-236), and the *C. elegans* PIR-1 paralog F54C8.4 (359 aa). An asterisk (*) indicates a fully conserved residue, a colon (:) indicates conservation of residues with strongly similar properties, and a period (.) indicates conservation of residues with weakly similar properties. Gray highlighting indicates that at least one other aligned sequence has an identical residue. The red square delimits the P-loop sequence and red arrowheads indicate residues that are essential for BVP triphosphatase activity as defined by alanine scanning in Martins and Schuman, 2002b. **(B)** Local pairwise alignments between the sequences represented as a matrix of the percentage similarity scores as determined by the Emboss Matcher program (Rice *et al.*, 2000; found at the EMBL-EBI website ebi.ac.uk/Tools).

## RATIONALE FOR THIS PROJECT

The discovery in our laboratory that PIR-1 associates with Dicer provided an exciting angle from which to dissect the biology of this relatively understudied RNA phosphatase. Importantly, it also opened a new door to learn more about Dicer-dependent processes, which are far from being completely understood. By utilizing *C. elegans,* we had the advantage of being able to probe for functions of PIR-1 that could potentially only be perceived in the context of a whole animal (a dimension that prior studies of PIR-1 orthologs lacked). The central objectives of this project were to understand how PIR-1 functions with Dicer to execute RNAi and whether it is involved in additional Dicer-dependent sRNA pathways. With the intent of gathering further functional clues, we also set out to conduct a deeper study of the phenotypes associated with *pir-1* mutant animals, and to characterize the PIR-1 protein at the biochemical, cellular and developmental levels. More broadly, we were interested in finding the function(s) of PIR-1 determining its essential role in *C. elegans* development, and whether such functions could be evolutionarily conserved. Below we provide a summary of key results and briefly state how they may influence the course of future PIR-1 studies.

In Chapter II of this thesis, we show that PIR-1 is required for the proliferation of germ cells and for proper sperm differentiation. We also demonstrate that PIR-1 is expressed ubiquitously throughout *C. elegans* development with a preferential nuclear localization, and that a catalytically-inactive transgenic PIR-1 is unable to rescue the mutant phenotypes. In Chapter III we prove that PIR-1 interacts with the Dicer-containing ERI complex, and that this association extends to the nuclear compartment. In Chapter IV our results reveal that *pir-1* mutant animals fail to accumulate spermatogenesis-associated 5'-monophosphorylated ALG-3/4 26G-RNAs, explaining both the sperm defect of the *pir-1* mutant and the association of PIR-1 with the ERI complex. Importantly, we show that, similarly to its orthologs, PIR-1 is able to convert 5'-triphosphorylated RNA ends into monophosphates *in vitro*. Curiously, we also provide results implying that, unlike other sRNA species, ALG-3/4 26G-RNAs are enriched in cell nuclei, in agreement with the primary localization of PIR-1. Finally, in Chapter V we find that PIR-1 functions in the same pathway as Dicer to suppress the replication of an RNA virus by promoting the accumulation of virus-derived siRNAs.

This work is the first to attribute concrete biological functions to this conserved phosphatase. Our discovery that PIR-1 is a *bona fide* RNA 5' phosphatase required for the

production of Dicer-dependent sRNAs is of high significance in that it adds an unexpected step in sRNA biogenesis. However, as with any biological inquiry, our findings multiplied the number of our questions. Is the RNA 5' phosphatase activity of PIR-1 required to promote the processing of Dicer substrates? Does PIR-1 function in equivalent steps of the 26G-RNA and the antiviral siRNA pathways? Is the nucleus a site of PIR-1-dependent sRNA production in association with Dicer and the ERI complex? Considering that the functions we describe are not strictly essential for *C. elegans* viability, what other processes is PIR-1 involved in? Does PIR-1 participate in sRNA biogenesis in vertebrates? Throughout this thesis we discuss our results in light of these questions and formulate working models intended to guide future experiments.

# CHAPTER II

Characterization and Rescue of a

Novel *pir-1* Loss-of-Function Allele

## INTRODUCTION

In our initial approach to PIR-1, we wanted to expand the phenotypic characterizations associated with a *pir-1* null mutant that were initiated in Duchaine *et al.* (2006), as well as to ask whether these phenotypes derive exclusively from PIR-1 loss. Additionally, we aimed to characterize aspects of *pir-1* expression at the RNA and protein levels in a wild-type context in order to start formulating working hypotheses about PIR-1 function.

In this chapter we describe the phenotypes associated with the novel *pir-1(tm3198)* deletion mutant, specifically obtained for this study. We implicate the lack of PIR-1 as the cause of the severe developmental defects manifested in the new mutant by transformation with wild-type PIR-1 transgenes. Using fully rescued GFP-tagged PIR-1 transgenic lines we characterize the dynamics of PIR-1 expression during development at the tissue and cellular levels. Additionally, by taking advantage of the knowledge gathered in studies of Baculovirus Phosphatase, we study some key conserved amino acids in PIR-1 to explore functional analogies between the two enzymes. This initial characterization of the mutant phenotypes and rescued lines provided valuable clues regarding potential PIR-1 functions and where its association with Dicer may be necessary.

## RESULTS

### A New Deletion Allele of *pir-1* Exhibits Phenotypes that Are Different from Those of the Original *pir-1(tm1496)* Deletion Allele

The original phenotypic characterization of *pir-1* described by Duchaine and colleagues (2006) was made using the *tm1496* deletion allele. This deletion comprises not only a section of the *pir-1* protein coding sequence, but also the entire upstream 246-base pair (bp) intergenic sequence and the first 48 bp of the first exon of the *sec-5* gene (Fig. 2.1A). The potential disruption of *sec-5* could be masking the phenotype associated with the deletion of *pir-1,* and precluded us from performing essential transgene rescue experiments. We therefore requested a new, *pir-1*-restricted deletion mutant allele from our collaborator Shohei Mitani at Women's Medical University School of Medicine in Tokyo. Since this process took a few months, our initial experiments to address the requirement of PIR-1 for sRNAs were performed with the existing *pir-1(tm1496)* mutant (described in Chapter IV). We thought that results obtained with the original allele could prove useful for future comparisons with a novel allele. However, this demanded a careful characterization of the *pir-1(tm1496)* mutant regarding the expression of *sec-5* and *pir-1* mRNAs.

*sec-5* is the first of a three-gene operon transcribed opposite of *pir-1* (CEOP2461). The SEC-5 (Yeast *SEC* Homolog 5) protein has been shown to be required for proper endocytosis and normal intestinal development (Fares and Greewald, 2001). By dsRNA injection, it is described to be embryonic lethal, with escapers arresting at late larval stages (Soennichsen *et al.*, 2005). For confirmation, we synthesized dsRNA of *sec-5* and injected it into N2 Bristol animals (wild-type reference strain), resulting in progeny that die as embryos. We also obtained an available 385 bp deletion allele – *tm1443* – which lacks the first two exons of *sec-5*. *tm1443* homozygotes exhibit a developmental arrest mostly at a larval L4-like stage, identical to that of *tm1496* homozygotes. These results suggested that the arrest phenotype observed in *pir-1* mutant animals could be a consequence of *sec-5* knockout or of any of the genes composing the operon, and not of *pir-1,* as originally proposed. RNAi knockdown of *pir-1* by feeding of *E. coli* expressing *pir-1* dsRNA, or more potently by direct injection of *pir-1* dsRNA, failed to produce any discernible phenotypes.

The new *pir-1* deletion mutant – *tm3198* – deleted a 407-bp segment confined to the *pir-1* gene, removing almost all of the first intron and second exon (Fig. 2.1A). All

potential splicing variants of the remaining sequence give rise to out-of-frame products with premature stop codons. This allele was outcrossed six times against the N2 background to eliminate other potential background mutations. Since it also gave rise to developmentally arrested sterile animals, the mutation was genetically balanced with *mnC1* (the same balancer used for *tm1496*) containing chromosomal rearrangements which suppress crossovers along the right arm of chromosome II, where the *pir-1* locus is situated. This ensured that the allele could be stably maintained, as *pir-1/mnC1* heterozygotes are fully fertile, and *mnC1* homozygotes are sterile. The *tm3198* allele therefore allowed us to conclude that *pir-1* may be essential for *C. elegans* development. Most importantly, while *tm1496* animals never grow beyond the L4 stage, the majority of *tm3198* animals develop into early adults with a larger germline. This phenotypic difference proved to be critical to understand *pir-1* function, as will become apparent from the results we describe. This indicated that, indeed, the *tm1496* deletion masked phenotypes associated with exclusive loss of PIR-1.

Since RNAi of *pir-1* failed to reveal any phenotypes, we measured the downregulation of *pir-1* mRNA by quantitative real-time PCR (qRT-PCR). Following RNAi by feeding the *pir-1* mRNA decreased by only ~50%, showing that silencing was specific but not effective (see Materials and Methods). Since injection of dsRNA typically leads to increased silencing relative to the feeding method, but still did not yield any phenotypes, we concluded that *pir-1* mRNA is partially refractory to silencing by RNAi and/or that the existing PIR-1 protein pool is very stable. We tried to circumvent this problem by feeding *pir-1* dsRNA bacteria to wild-type or Eri mutant animals with heightened sensitivity to RNAi (*eri-1* and *rrf-3*) for several generations, or by injecting Eri mutants with *pir-1* dsRNA. None of these strategies produced any obvious effects on growth and reproduction. Thus our experiments were confined to the developmental stage at which *pir-1(tm3198)* mutant animals arrested.

*pir-1* mutant homozygotes arise from heterozygous mothers, which could transmit maternal *pir-1* material – mRNA or protein – to the progeny. We thus assessed expression by qRT-PCR of *pir-1* mRNA in both deletion mutants at different time points. In order to generate large quantities of arrested animals we placed the balanced mutations in a triple mutant genetic background (designated throughout this work as ‘avr3x’) that confers resistance to ivermectin, a drug that kills nematodes. Together with a gene expressed in the genetic balancer, this combination allowed counter-selection of deletion homozygote animals in the presence of the drug (see Materials and Methods for more details). We

counter-selected animals for five and seven days at 20˚C to compare depletion levels of both *pir-1* and *sec-5* transcripts. While the *tm1496* deletion did abrogate *pir-1* and *sec-5* mRNA expression, *tm3198* interfered exclusively with *pir-1* (Fig. 2.1B). *pir-1* mRNA was nearly undetectable after five and seven days of counter-selection in *tm3198* animals, whereas *sec-5* mRNA accumulated at these time points in the *tm3198* background. From this we concluded that both alleles are effectively null for *pir-1* expression and that potentially maternally loaded *pir-1* mRNA is almost completely cleared within at least five days in *pir-1* homozygotes.

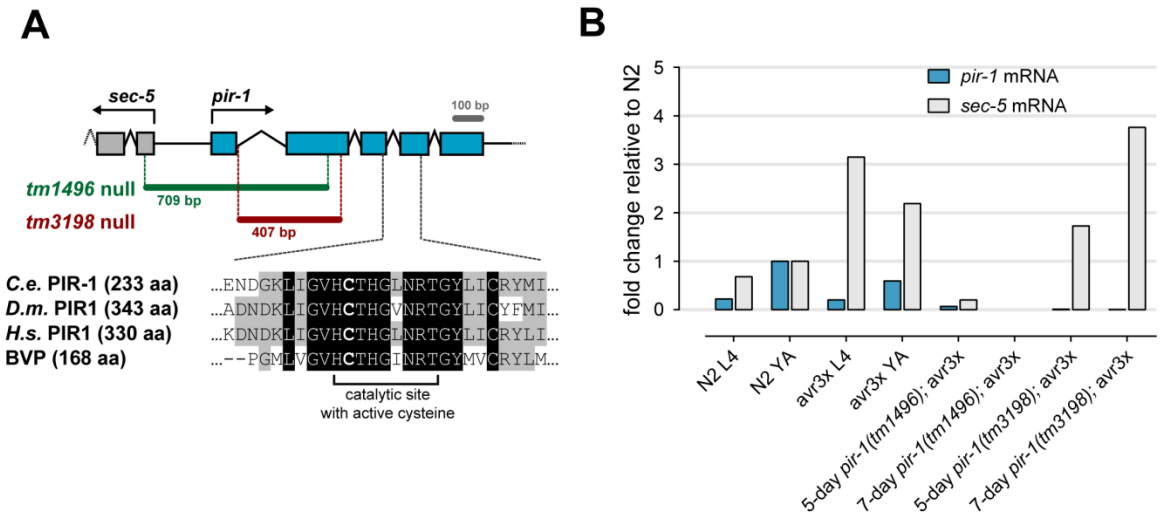



**Figure 2.1. Both *tm1496* and *tm3198* deletions constitute null alleles of *pir-1*. (A)** Diagram of the *pir-1* genomic locus on chromosome II and the regions deleted by the *tm1496* and *tm3198* mutations. Boxes represent protein-coding exons and bent lines indicate introns. The gene position which encodes the catalytic site with the active cysteine is also highlighted, with protein alignments for some orthologs. Amino acid identity among at least two sequences is highlighted in gray, and between all sequences is highlighted in black. **(B)** qRT-PCR of *pir-1* and *sec-5* mRNAs normalized to *gpd-2* (*G*lyceraldehyde 3-*P*hosphate *D*ehydrogenase 2) mRNA levels and relative to levels in N2 wild-type young adults (YA). Mutants were counter-selected on ivermectin for five or seven days before RNA extraction.

## *pir-1(tm3198)* Mutant Animals Exhibit Somatic and Germline Defects

When comparing a normally developing and fertile young adult *pir-1/mnC1* heterozygote hermaphrodite to a *pir-1* homozygote of the same age and approximate size (~3.5 days at 20˚C; Fig. 2.2A) it becomes obvious that the latter are sterile. While heterozygotes have a fully developed U-shaped germline (highlighted in yellow) with a germ cell syncytium, oocytes and developing embryos, *pir-1* homozygotes have a much smaller germline, apparently containing only mitotic/meiotic syncytial germline nuclei.

After seven days, *pir-1(tm3198)* animals exhibit a fully penetrant developmental arrest within a range of stages that vary between L3-like larvae to sterile young adults (Fig. 2.2B and C). Vulvae are frequently protruding and malformed, in some cases leading to bursting of the animals at this position (Fig. 2.2B). Signs of oogenesis can be seen in about 21% of adult animals, although oocytes exhibit serious malformations and never give rise to fertilized eggs (Fig. 2.2B and C). Further confirming that larger animals are adults, they possess a wild-type number of seam cells (16 on each side of the animal), as detected by expression of a nuclear GFP reporter under the control of a seam cell promoter in the *pir-1* background (see Fig. 4.3, page 168). Seam cells (or lateral hypodermal cells) produce alae, antero-posterior lines of thickened cuticle on both sides of the animal after the L4/adult molt. These typically adult structures are also normally formed in larger arrested *pir-1* animals (not shown). Most of the animals are very active and continuously feed, exhibiting a wild-type life-span of 16-18 days at 20˚C. A small fraction (~15%) of the animals is either sick or dead after seven days.

The vulval phenotype is a strong indicator of abnormal somatic development and is also observed in Dicer *dcr-1(ok147)* deletion mutants, which, like *pir-1*, have to be maintained as genetically balanced heterozygotes (Knight and Bass, 2001; Welker, *et al.* 2010). Dicer mutant homozygotes display a range of defects, the severity of which is thought to vary according to how much mRNA or protein each oocyte inherits and gets passed on to embryos. The majority have enough maternally loaded material to allow development into normal-sized adults with defective vulvae that often burst, and with gonads exhibiting relatively normal morphology at the distal end but with malformed and unfertilized oocytes at the proximal end (Knight and Bass, 2001). Similarly, the range of phenotypes in *pir-1(tm3198)* homozygotes is indicative of a variable maternal deposition of mRNA or protein in embryos. Additionally, when we allowed starved mutants to live in the dauer stage for a prolonged time period (2-3 weeks), *pir-1* heterozygotes quickly resumed normal development to become fertile adults upon feeding, while *pir-1* homozygotes exited the dauer stage but never reached adulthood. The dauer stage is an alternative stress-resistant larval form that skips the L3 stage upon stresses such as overcrowding and prolonged starvation, and can live up to three months (reviewed in Hu, 2007). We have seen an identical response with recovered *dcr-1(ok147)* homozygote dauer larvae. This observation suggests that the maternally deposited material is depleted during the starvation period, not allowing development to proceed as it does when feeding is uninterrupted.

We made multiple attempts at raising antibodies against *C. elegans* PIR-1 in rabbits by immunization with affinity purified recombinant fragments of PIR-1 (the N-terminal or the C-terminal half of the protein) and with short peptides of predicted high antigenicity. Unfortunately, we never obtained effective antibodies, even after affinity purification against the used antigens. Although this did not allow us to determine whether the onset of the *pir-1* phenotypes coincided with depletion of maternally derived PIR-1 protein in mutants, we characterized the expression of *pir-1* mRNA along development in N2 animals by qRT-PCR (Fig. 2.2D). We found that in contrast to *unc-22* (*Unc*oordinated 22), a muscle-specific gene whose expression is highest during larval growth, *pir-1* mRNA expression is high in adults but even higher in embryos. This expression profile correlates with germline development and potentially with maternal deposition of *pir-1* mRNA in embryos. However, because these embryo preparations also included late embryos, which, unlike early embryos, undergo active transcription, we cannot exclude the possibility that a large proportion of *pir-1* mRNA is made during embryogenesis.

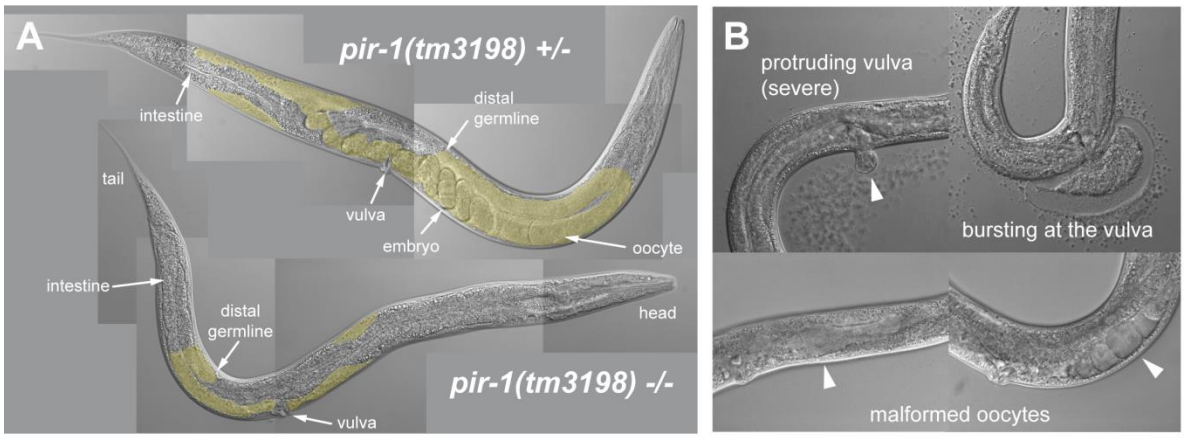
Analysis of dissected gonads from adult *pir-1* mutant arrested animals shows that they are much smaller than those of a wild-type animal of the same size (Fig. 2.3A). This appears to be a consequence of an under-proliferation of syncytial germ nuclei, as the normal distal-to-proximal architecture of the germline is maintained but the density of nuclei is much lower than expected for the apparent developmental stage of the *pir-1* mutant. A wild-type late L4 germline typically consists of (1) a mitotic proliferative region; (2) a transition zone that marks entry into meiosis (with characteristic “crescent”-shaped chromatin) and comprises the leptotene and zygotene stages of meiotic prophase I, when homologous chromosomes pair, undergo synapsis, and start recombining; (3) a large region of nuclei in the pachytene stage of meiosis, when chromosomes complete synapsis and recombination events are finished; and (4) a spermatogenesis zone that includes nuclei undergoing diplotene, during which the synaptonemal complexes are broken down to reveal chiasmata, followed by diakinesis, when chromosomes condense in preparation for division (Lui and Colaiacovo, 2013). Considering their dimension and number of nuclei, *pir-1* germlines roughly resemble the proliferative state of L3 larvae. However, unlike L3 larvae, arrested *pir-1* gonads exhibit signs of spermatogenesis, including the presence of mature spermatids which are visible as very small round DAPI stained foci at the proximal tip of the germline (Fig. 2.3A). In *C. elegans* hermaphrodites, meiotic nuclei first give rise to sperm during the L4 stage, and only during adulthood does meiosis shift to producing oocytes. Spermatids are stored in a spermatheca in each of the two gonad arms, through

which oocytes pass to be fertilized. In the small fraction (~21%) of adult *pir-1(tm3198)* animals that showed signs of oogenesis, it was clear that these cells were very defective (reminiscent of *dcr-1(0k247)*) and rarely reached more than one oocyte per gonadal arm (identified by the presence of a large oocytic nucleus undergoing diakinesis, in preparation for meiotic division). Even though *pir-1(tm1496)* arrested animals were the size of L4 larvae, their germlines were more precocious, never reaching spermatogenesis (not shown).

In addition to a severely under-proliferated mitotic zone, another feature suggests that the germline may arrest due to a defect in mitotic proliferation. In undifferentiated germ cells, all nuclei are surrounded by foci termed P granules. These foci aggregate mRNAs and a variety of proteins with RNA processing and regulatory roles, including components of the RNAi machinery, and play important roles in coordinating post-transcriptional germline gene expression and maintaining overall germ cell identity and integrity (reviewed in Updike and Strome, 2010). We crossed an integrated *pgl-1::rfp* transgene into the *pir-1* background in order to use PGL-1 (*P-G*ranu*l*e Abnormality 1; a major component of P granules) as a marker of normal, undifferentiated germ cells. P granules appear to have a normal morphology and distribution in arrested *pir-1* germlines when compared to a wild-type background (Fig. 2.3B). This is good evidence that those cells retain their germline identity, and that the arrest does not result from premature differentiation into other cell types.

During spermatogenesis, pachytene nuclei condense and cellularize before undergoing the two sequential meiotic divisions that will give rise to haploid spermatids (Ellis and Stanfield, 2014). These spermatid precursors are termed spermatocytes (Fig. 2.3A). While the diplotene condensation phase seems normal, arrested *pir-1* germlines exhibit DNA bridging between dividing primary spermatocytes, implying that meiotic chromosome segregation is impaired (Fig. 2.3C). This defect is also observed in Eri mutants during sperm development, namely *rrf-3*, *eri-1*, *eri-3*, *eri-5* and the *dcr-1(mg375)* allele which possesses an inactivating point mutation in the helicase domain (Gent *et al.*, 2009; Pavelec *et al.* 2009). These mutants are completely sterile when spermatogenesis occurs at the upper-limit temperature of 25˚C. Spermatids from Eri mutant hermaphrodite and male animals at 25˚C exhibit morphological defects and are unable to be activated into a motile stage required for proper fertilization, and, in the case of male sperm, for migration into the hermaphrodite spermatheca after mating. As a result, only unfertilized oocytes are laid. At lower temperatures (15˚C-20˚C) brood sizes are 30-50% lower than in

wild type (Pavelec *et al.*, 2009). We also observed identical morphological abnormalities in a fraction of *pir-1* mutant spermatids (not shown). Additionally, at 25˚C their germlines never reach spermatogenesis, perhaps as a consequence of maternally deposited material being consumed more quickly due to accelerated development. Conversely, at 15˚C they undergo spermatogenesis and DNA bridging between dividing spermatocytes is widespread. We attempted to mate *pir-1* homozygote males raised at 15˚C and 20˚C with wild-type hermaphrodites but never obtained cross progeny. *pir-1* males are smaller (by ~25%) than wild type and often exhibit an immature tail structure, which may not allow insemination to occur.




**C**

***pir-1(tm3198)* phenotypes after 7 days at 20ºC (n=133)**

| | |
|---|---|
| normal to mild protruding vulva | 45% |
| protruding vulva (severe) | 19% |
| pvul with signs of oogenesis | 21% |
| smaller/without vulva (L3/L4-like) | 21% |
| sick/dead | 15% |
| sick/dead that burst | 15% |

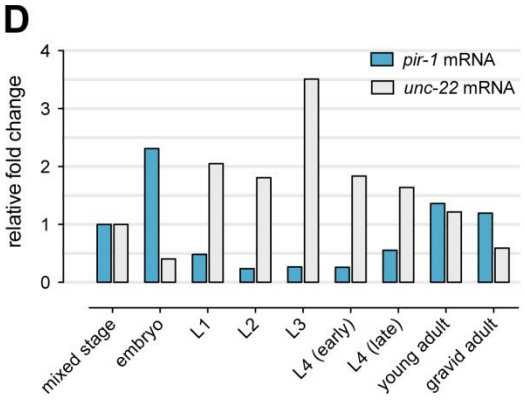



**Figure 2.2. *pir-1(tm3198)* homozygote animals are unable to reach sexual maturity and exhibit developmental defects. (A)** A gravid, balanced *pir-1* mutant heterozygote (above) next to a sterile *pir-1* mutant homozygote grown simultaneously for 96 hours at 20˚C. The two germline arms of each animal are highlighted in yellow and are partly concealed by intestinal tissue. The germline grows in a distal to proximal axis. In mature hermaphrodites, fertilized eggs accumulate proximally in the uterus and are expelled through the vulva. **(B)** Phenotypes associated with arrested mutants after seven days of growth. White arrowheads indicate the vulva and malformed oocytes. **(C)** Scoring of phenotypes in a population of arrested *pir-1* homozygotes grown simultaneously for seven days. **(D)** qRT-PCR of *pir-1* mRNA and *unc-22* mRNA across development in wild-type N2 synchronous populations. Levels of each mRNA were normalized to 18S rRNA and fold changes are relative to a mixed-stage population.

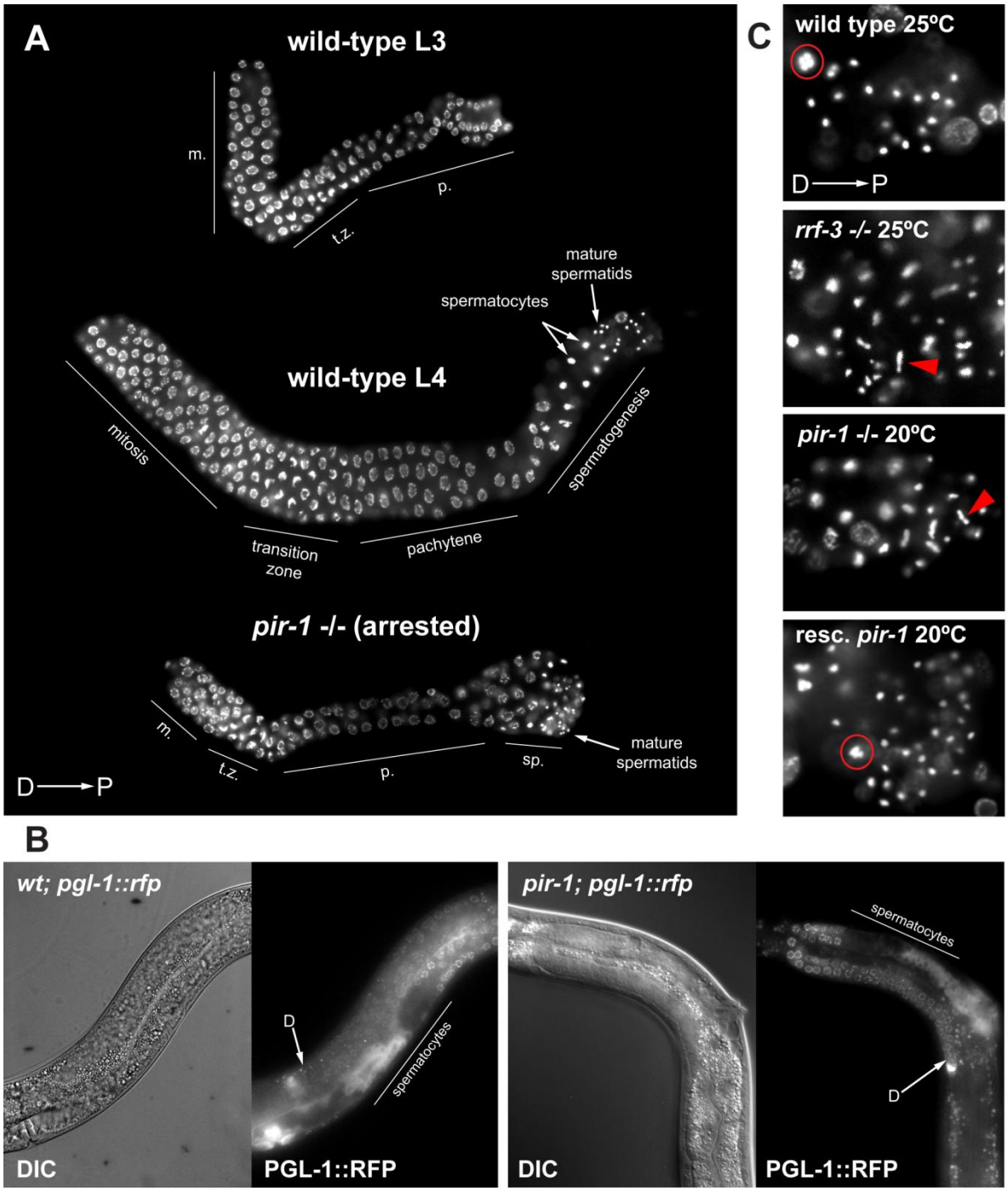



**Figure 2.3. The germline of *pir-1(tm3198)* arrested adults exhibits defects in proliferation and spermatogenesis. (A)** Comparison of fixed DAPI stained germlines of late L3 and late L4 wild-type N2 animals with a terminally arrested *pir-1* mutant (all grown at 20˚C). After mitotic proliferation, the transition zone marks entry into meiotic prophase, in which the chromatin is polarized in the nucleus, adopting a crescent-shaped appearance. Subsequently, nuclei enter pachytene during which chromosomes are redispersed around the nuclear periphery and are said to have a "bowl of spaghetti" appearance. During the L4 stage spermatogenic diakinesis, nuclei condense along with chromatin, leading to highly compacted cells termed spermatocytes which will undergo two consecutive divisions to produce mature spermatids. **(B)** P-granule distribution and morphology in live wild-type L4 and arrested *pir-1* animals. In both cases P granules can be seen surrounding undifferentiated mitotic and meiotic germ nuclei, marked by fluorescent PGL-1::RFP, which disappears as nuclei commit to spermatogenesis. **(C)** Spermatogenesis in L4s and in

an arrested *pir-1* mutant. The smaller DAPI stained dots are the nuclei of mature spermatids. Examples of healthy dividing spermatocytes are circled in red. In both the Eri mutant *rrf-3* at 25˚C or in *pir-1* at 20˚C spermatocytes fail to divide properly, forming chromosome bridges (red arrowheads). Spermatogenic divisions are fully rescued with a single-copy *pir-1::gfp* transgene (bottom image). 'D' is distal; 'P' is proximal.

## Defects in *pir-1(tm3198)* Mutant Animals Are Rescued by Tagged Wild-Type *pir-1* Transgenes

In order to rescue the aforementioned *pir-1* mutant phenotypes, we produced transgenic lines using four different constructs carrying N-terminal GFP or 3xFlag tags, and C-terminal GFP or 3xFlag tags fused to the *pir-1* gene. To ensure inclusion of essential regulatory sequences, the tagged genes were flanked on each side by 1,000 bp of native sequence. Multiple transgenic strains were first generated by biolistic bombardment, that gives rise to low-copy number, integrated transgenes with high frequency (Praitis *et al.*, 2001). Despite being more costly and labor intensive, we favored it over transformation by direct germline microinjection of plasmids. The latter method generates high-copy extrachromosomal arrays that tend to be silenced in the germline by RNAi mechanisms triggered by the highly repetitive nature of these recombined DNA assemblies (transgene silencing; Kelly *et al.*, 1997). Furthermore, integration of extrachromosomal arrays into the genome requires an extra step involving irradiation with highly mutagenic doses of UV- or X-rays.

The growth, fertility, and spermatogenesis defects of *pir-1* mutant animals were rescued upon introduction of C-terminally tagged *pir-1* transgenes into the *tm3198* deletion background by genetic crossing (all of six C-GFP and all of three C-3xFlag integrated lines that were crossed led to rescue; figures 2.4-2.7 all show rescued *pir-1* animals). Brood sizes were counted for a few independent lines, averaging 200-300 viable progeny per animal at 20˚C, in line with wild-type brood sizes. PIR-1::GFP integrated lines demonstrated that PIR-1 is expressed in most somatic and germline cells but that there was variation in the intensity of GFP expressed in germline *versus* soma between different lines. For further studies we chose a line that expressed GFP robustly in both somatic and germline cells (Fig. 2.4A), as well as a 3xFlag-tagged line. We used these lines for cell-biological and biochemical characterization of PIR-1. Interestingly, N-terminally tagged transgenes did not rescue any of the phenotypes (seven N-3xFlag and one N-GFP were crossed into *pir-1*), nor did we ever observe GFP signal from a total of 29 integrated lines. We later found that this occurred due to a misprediction of the *pir-1* start codon at the

onset of this project. At that time the *C. elegans pir-1* sequence was predicted to encode an extra 28 N-terminal amino acids, giving rise to a 261 aa protein. We later realized, through our own experiments (explained in Chapter III) as well as more recent curations of the *pir-1* gene based on high-throughput sequencing experiments, that PIR-1 is in fact 233 aa long, and consequently the tags should have been fused with the ATG at position 29 of the original gene model. Therefore, our N-terminal tagging likely disrupted the *pir-1* promoter.

During the course of this work, a new microinjection-based method for single copy insertion of transgenes was developed, termed MosSCI (*Mos*1-mediated Single Copy Insertion; Frokjaer-Jensen *et al*., 2008). Briefly, this method is based on the excision of the *Mos*1 transposon from *Drosophila melanogaster* inserted at known locations of the *C. elegans* genome by heterologous expression of the corresponding transposase. The excision, which takes place in the germline, happens in the presence of a plasmid carrying the sequence of interest flanked by sequences homologous to both sides of the site where *Mos*1 is inserted. The dsDNA break generated by excision of the transposon triggers repair of the DNA by homologous recombination with the genomic sequences in the plasmid, resulting in the integration of a single copy of the sequence of interest. We decided to use this technique in order to express PIR-1::GFP at a level closer to that of the native gene. Precise recapitulation of native expression is not guaranteed, however, as the genomic context of the insertion site is different than that of the original gene. With a single copy of *pir-1::gfp* on chromosome IV we were also able to rescue all phenotypes associated with *pir-1*. Although PIR-1::GFP intensity in the germline was similar to that of the bombardment line, somatic cells (and especially intestinal cells) expressed GFP more dimly (Fig. 2.4A vs Fig. 2.5A-C).

## PIR-1 Is Expressed in the Nucleus and Cytoplasm of Most Somatic and Germ Cells throughout Development

In GFP-tagged lines PIR-1 is visibly concentrated in the nuclei of both somatic and germline cells (Fig. 2.4.A and B). In non-integrated lines obtained by bombardment that overexpress PIR-1::GFP in mosaic patterns, PIR-1 is also evidently present in the cytoplasm, most visibly in neurons (Fig. 2.4C). The expression pattern throughout development is identical in the bombardment and single-copy MosSCI strains. Consistent with expression from multiple copies of the transgene, the GFP signal is generally stronger in the bombardment rescued strain, particularly in somatic nuclei. Somatic nuclei are brighter in early larvae, and gradually decrease intensity as they approach adulthood.

Interestingly, this is most noticeable in the nuclei of the intestine, which undergo a round of endoreduplication at each larval stage, progressing from a 2C (diploid) DNA complement at hatching to 32C in adulthood. Staining of PIR-1::GFP in fixed intestines shows that PIR-1 also co-localizes with chromatin, at the nuclear periphery (Fig. 2.4B). The fact that nuclear PIR-1::GFP is so abundant during intestinal development could suggest a role for it in promoting successful DNA replication. This does not explain, however, why somatic nuclei expression is retained in other cell types.

In the germline the difference in GFP intensity between multi-copy and single-copy strains is not very perceptible, perhaps as a consequence of tighter expression regulatory mechanisms in this tissue. It is in the germline, however, that PIR-1 expression dynamics is most interesting. During hermaphroditic germline development, PIR-1::GFP expression starts becoming visible in all germline nuclei during the L2/L3 stages, during which the germline expands through mitotic proliferation of its nuclei (Fig. 2.5A). In L4, as the germline further differentiates, so does the distribution of PIR-1::GFP along the distal to proximal axis. Distally, from the mitotic to roughly the mid-pachytene region, PIR-1 intensity is relatively high and uniform. After this it gradually fades out until a sudden increase occurs a little before nuclei begin to condense to become spermatocytes (Fig. 2.4A, 2.5C and 2.7). In adults, after spermatogenesis is complete, meiotic nuclei start giving rise to oocytes. Pachytene is followed by diplotene, during which nuclei start enlarging and becoming cellularized. As both cytoplasm and nuclei expand, diakinesis progresses to give rise to mature oocytes. Unlike spermatocytes, in which nuclear volume is reduced, the highly condensed chromosomes that result from diakinesis can be seen individually in the oocyte nuclei (Fig. 2.6A), and separate only after fertilization. Throughout this process nuclei are dark and contrast with a faint signal in the surrounding cytoplasm in both live and fixed/stained germlines, suggesting that oogenesis does not require nuclear PIR-1 (Fig. 2.5D and 2.6A). The cytoplasmic signal is too faint to make a conclusion about whether it represents diffuse cytoplasmic PIR-1::GFP or simply cytoplasmic autofluorescence.

In embryos PIR-1::GFP remains absent from nuclei until the eight-cell stage, after which it gradually increases its intensity in the multiplying nuclei (Fig. 2.4B and 2.5D). Once again, as we observed in the germline, PIR-1 expression overlaps with a highly proliferative stage of development, suggesting that it may have an important role in embryogenesis. Interestingly, in later embryos (past the 100-cell stage) the only cells that do not express PIR-1 in the nucleus are the two germline precursor cells, Z2 and Z3,

identified by the presence of PGL-1-containing P granules around their nuclei. These primodial germ cells remain PIR-1 negative well into the L1 larval stage, during which they begin to proliferate (Fig. 2.6B).

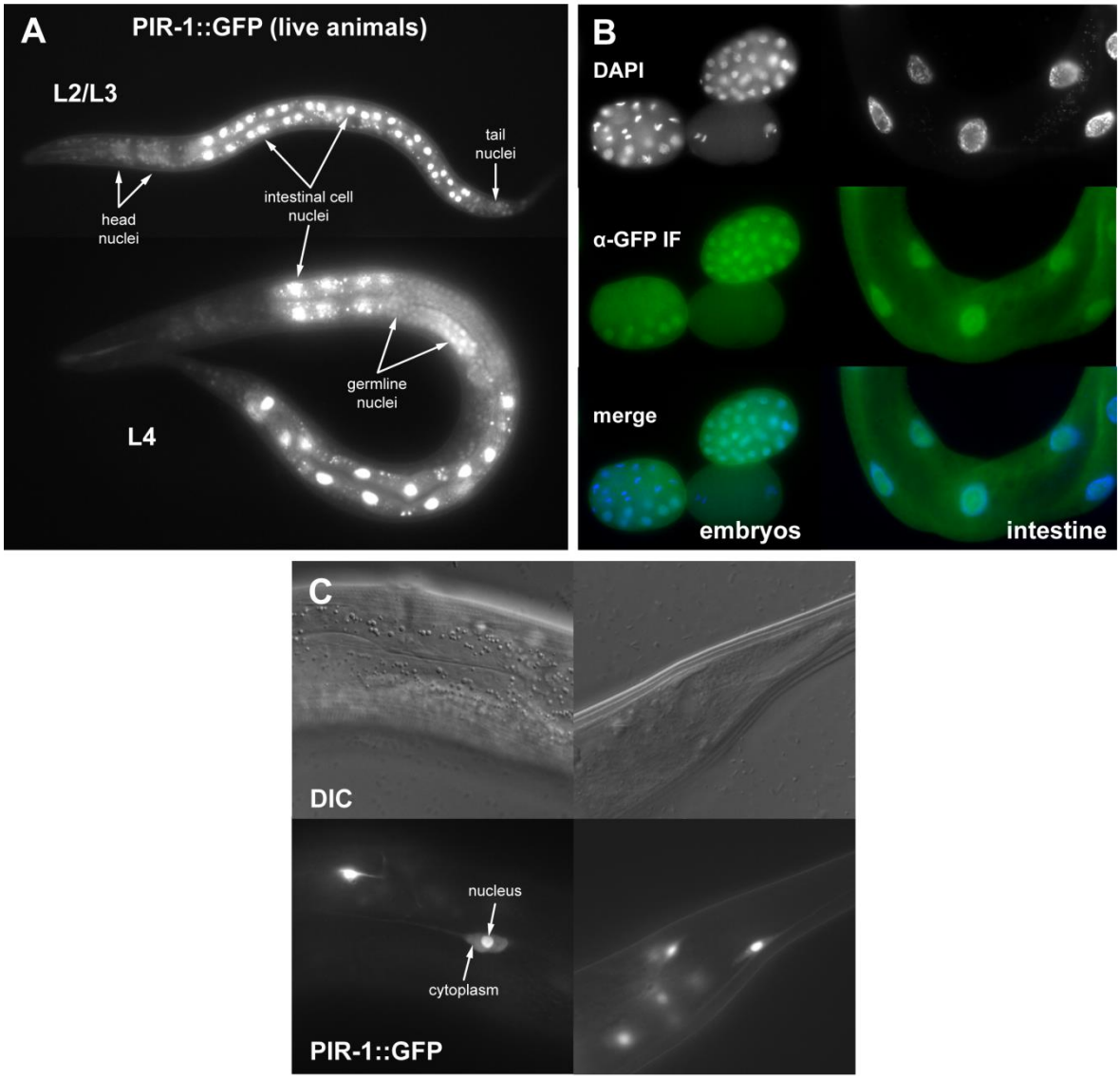



**Figure 2.4. PIR-1::GFP expression in bombardment lines. (A)** Live rescued *pir-1(tm3198)* with integrated *pir-1::gfp*. GFP expression can be seen in somatic and germline tissue, concentrating in nuclei. Intestinal cell nuclei are particularly bright in this strain. Other somatic nuclei, namely those of neurons in the head and tail regions accumulate PIR-1. Germline nuclei are visible in a mid-L4 larva and are brighter in the proximal region, where spermatogenesis occurs. The smaller bright spots in intestinal cells are autofluorescent intestinal granules that do not represent GFP signal. This strain was used for several experiments described in Chapter III. **(B)** Fixed embryos and adult intestine of the same strain followed by immunofluorescence against GFP and DAPI staining showing that PIR-1::GFP is also expressed in embryos and co-localizes with chromatin of intact nuclei, but not during cell divisions (note metaphase plates and anaphase in leftmost embryo). **(C)** Two live examples of non-integrated bombardment lines that overexpress PIR-1 in only a few cells. In this case, mostly neurons are visible in the mid and tail regions, showing that in addition to the nucleus, PIR-1 is also diffusely distributed throughout the cytoplasm.

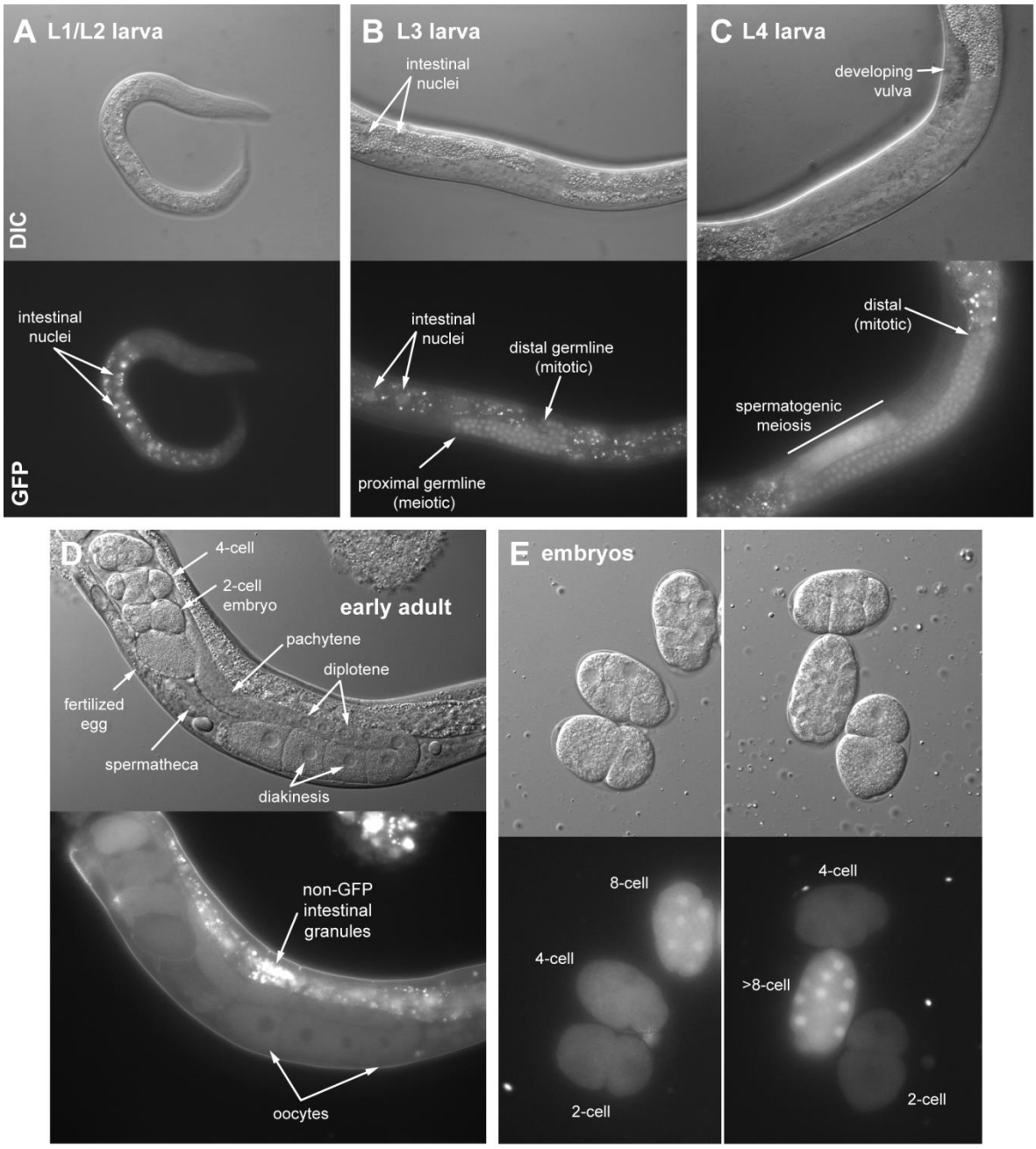



**Figure 2.5. Expression pattern of PIR-1::GFP across development in live *pir-1* animals rescued with a single-copy transgene. (A)** In early larvae the most conspicuous nuclei are those of intestinal cells. PIR-1::GFP intensity in these nuclei decreases along development but it can always be detected. **(B)** In an L3 animal whose germline is proliferating mitotically and just initiating meiosis at the proximal end, PIR-1 uniformly accumulates in all germ cell nuclei. **(C)** L4 animals continue to express PIR-1 in distal mitotic nuclei. In the proximal germline, where sperm development is taking place, PIR-1 accumulates more strongly in nuclei and cytoplasm. **(D)** In gravid hermaphrodites, PIR-1 is still in distal germ cell nuclei but gradually disappears as nuclei enter diplotene, and then diakinesis during oocyte formation. **(E)** During embryonic development PIR-1 gradually accumulates in nuclei and cytoplasm. In the first two divisions PIR-1 is excluded from the nucleus, but by the eight-cell stage it appears in nuclei and continues to build up in subsequent divisions.

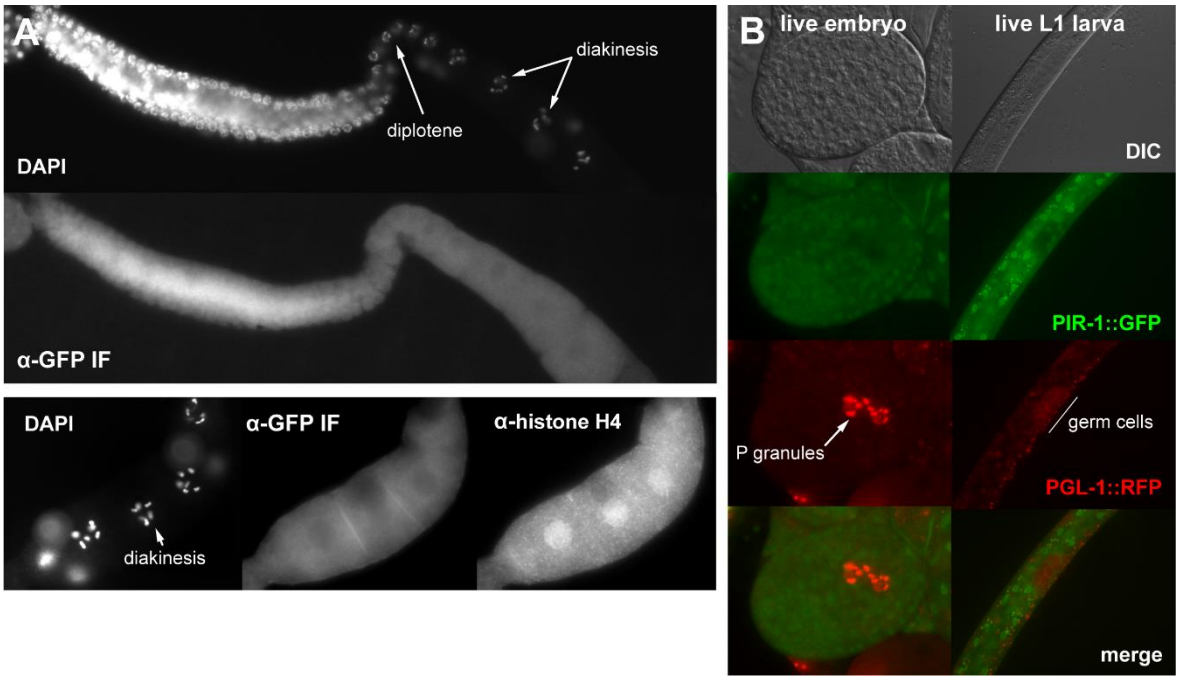



**Figure 2.6. PIR-1 does not accumulate in the nuclei of oocytes and primordial germ cells. (A)** Fixed germlines of rescued *pir-1; pir-1::gfp* young adult hermaphrodites were stained with anti-GFP antibody to further confirm that oocyte nuclei do not accumulate PIR-1. The lower panel shows fixed oocytes double stained with anti-GFP and an anti-histone H4 antibody. DAPI staining reveals highly condensed chromosomes in diakinesis. **(B)** A live post-100-cell stage embryo shows that the only nuclei not expressing PIR-1 are the two primordial germ cells Z2 and Z3, identified by perinuclear PGL-1::RFP protein, expressed from an integrated transgene. In L1 larvae, in which germline proliferation begins, PIR-1 is still absent from germ nuclei, but gradually builds up as proliferation continues. All depicted tissues and animals are bombardment-rescued *pir-1; pir-1::gfp*.

## Male and Hermaphrodite Germlines Exhibit Distinct PIR-1 Expression Patterns

Both male and hermaphrodite germlines robustly express PIR-1::GFP but exhibit curiously distinct expression patterns along the distal-proximal axis (Fig. 2.7A). In hermaphrodites, nuclear PIR-1::GFP is brightest at the distal mitotic region and transition zone, gradually fading away as nuclei go through pachytene. In L4/early adulthood, when all sperm cells are produced, nuclear and cytoplasmic GFP signal is abruptly turned on in the area where meiotic nuclei commit to spermatogenesis. Later in adulthood, as meiotic nuclei start being allocated for oocyte production, PIR-1::GFP is excluded from nuclei. This contrasts with male germlines – exclusively committed to sperm production – where nuclear PIR-1::GFP is weakly expressed in mitotic nuclei, but gradually accumulates without interruption as spermatogenesis progresses. PIR-1 clearly localizes to the nuclei

and chromatin of germ cells undergoing pachytene and diplotene and gradually weakens in spermatocytes (Fig. 2.7A and B). Unlike the germ cells that precede them, whose nuclei are only partially enveloped by a cell membrane and thus share a common cytoplasm, spermatocytes are fully cellularized. In spermatocytes, where chromatin starts a dramatic condensation process in preparation for the meiotic divisions that follow, PIR-1 no longer localizes to nuclei, but is still seen in the cytoplasm (Fig. 2.7B). Mature spermatids, on the other hand, seem completely devoid of PIR-1 (Fig. 2.7A and B). This implies that PIR-1 activity is most important in earlier stages of differentiation. Consistent with the spermatocyte division defects observed in *pir-1* mutants, these patterns of expression provide further evidence that PIR-1 may play an active role in sperm development.

Another noteworthy difference concerns the distal tip cells, or DTCs. In both males and hermaphrodites, these somatically derived cells are located at the apex of the gonadal arms, and actively signal nearby germ nuclei to proliferate mitotically (Germline Stem Cells or GSCs) through GLP-1/Notch (Abnormal *G*erm*l*ine *P*roliferation 1, a Notch family receptor) signaling (reviewed in Kimble and Crittenden, 2007). There are, however, important distinguishing features between hermaphrodite and male DTCs (hDTCs and mDTCs; Morgan *et al.*, 2010). Hermaphrodites possess one DTC per gonadal arm and share their capacity to induce mitotic proliferation with distal gonadal sheath cells, which envelope the hermaphrodite germline. Sheath cells support much of the GSC proliferation during larval development (Korta and Hubbard, 2010). By contrast, males do not possess gonadal sheath cells, so that mitotic proliferative signals are presumed to originate exclusively from the two DTCs present in their single gonadal arm. Additionally, hDTCs have a morphological function in guiding proper migration of the germline arms within the growing animal as proliferation occurs, so that it acquires the characteristic U shape, while in males a different cell assumes this role (Chesney *et al.*, 2009). We observed that both mDTCs express nuclear PIR-1::GFP very brightly, while the single hDTCs do not visibly express it. This difference may constitute an important clue linking PIR-1 to cell proliferation and to differences in how the process is orchestrated in the male and hermaphrodite germlines. In males PIR-1 may stimulate proliferation through the two distal tip cells, while in hermaphrodites this stimulatory effect may occur downstream, by intrinsic expression of PIR-1 in GSCs.

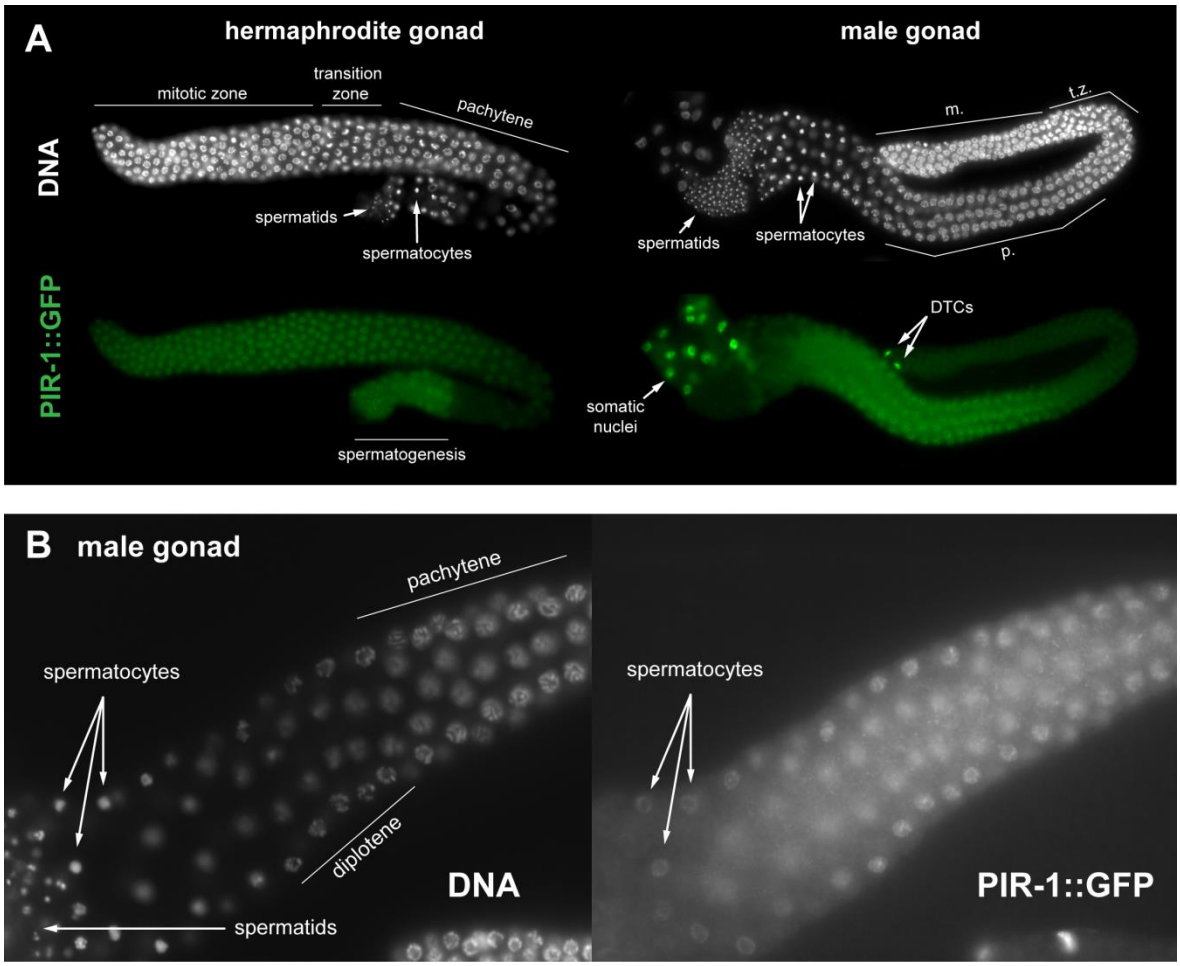



**Figure 2.7. The distribution of PIR-1::GFP in hermaphrodite and male germlines is not equivalent.** Germlines from bombardment rescued *pir-1; pir-1::gfp* L4 hermaphrodites and adult males were dissected, fixed and stained with anti-GFP. **(A)** At this stage of hermaphrodite development, spermatogenesis is still taking place and a burst of PIR-1 expression can be seen in the nuclei that have committed to sperm production. In males, all meiotic nuclei develop into sperm, which also coincides with increased PIR-1 accumulation. Expression decreases in the spermatocytes and spermatids. **(B)** Another example of an adult male germline showing a more detailed localization of PIR-1::GFP. The protein can be detected on chromatin during the pachytene and diplotene stages, but is no longer present in the nuclei of spermatocytes and spermatids.

## Targeted Mutagenesis of PIR-1 Shows that Its Phosphatase Activity Is Required for Rescue of All Mutant Phenotypes

In order to investigate functional parallels between *C. elegans* PIR-1 and its orthologs, we used the rescuing *pir-1::gfp* transgene to alter conserved residues by site-directed mutagenesis. We first replaced the critical active-site cysteine residue with a serine (C150S mutant; Fig. 2.8), which in baculovirus BVP and human PIR1 leads to a complete loss of RNA phosphatase activity (Takagi *et al.*, 1998; Deshpande *et al.*, 1999). The single-copy insertion line we obtained with the *pir-1(C150S)::gfp* transgene did not

rescue any of the mutant phenotypes, despite observing no change in GFP localization and still being able to detect full-length PIR-1::GFP by western analysis (shown in Fig. 3.6E and F, page 130). From this we concluded that the phosphatase activity of PIR-1 is required for its roles in promoting growth and fertility.

Besides the catalytic mutant, we also wanted to use some of the information gathered in previous structure/function studies of BVP to establish functional resemblances between the two proteins. Additionally we intended to generate a *pir-1* transgene with weaker rescuing activity (*i.e.*, a hypomorph) that would allow the extension of germline development and ideally give rise to embryos with compromised PIR-1 activity. This was an important goal considering our inability to use RNAi. We based our choice of residues to substitute on two studies from the laboratory of Stewart Shuman at the Sloan-Kettering Institute, in which the importance of conserved residues for enzymatic activity was determined *in vitro* and *in vivo* with mutated versions of BVP (Martins and Shuman, 2002a; Martins and Shuman, 2002b). In these studies, *in vitro* activity was assessed in two ways: (1) ATPase activity measured by the release of inorganic phosphate from an ATP substrate compared to wild-type BVP, and (2) the amount of covalent intermediate phosphoenzyme $^{32}$P-BVP formed by the reaction between [$\gamma$-$^{32}$P]ATP and BVP under conditions that stabilized this labile form. As an *in vivo* readout the authors used a heterologous *Saccharomyces cerevisiae* system in which the lethality associated with deletions in the yeast mRNA capping triphosphatase (*cet*) and guanylyltransferase (*ceg*) genes was rescued with a chimaera formed by BVP and the guanylyltransferase domain of the bifunctional mammalian capping enzyme (*i.e.*, BVP functionally replaced the triphosphatase activity of the capping enzyme).

The rationale for picking candidate mutations that could lead to partial PIR-1 activity was based on the ability of BVP to still rescue yeast capping *in vivo* while exhibiting some degree of compromised ATPase activity, phosphoenzyme formation, or both (Table 2.1). Also considered was temperature sensitivity (ts) defined by absence of yeast growth at 37˚C as opposed to normal (++) or slow (+) growth at 30˚C. We prepared mutant MosSCI constructs, but found that rescue of *pir-1* phenotypes did not require integration of the transgenes, as non-integrated lines of the wild-type *pir-1::3xflag* or *pir-1::gfp* constructs always led to rescue. Since multi-copy transgenic arrays tend to be silenced in the germline very easily, we were surprised to observe that *pir-1* was resistant to this phenomenon and speculate that it is related to its resistance to silencing by exo-RNAi.

With the exception of the R5A equivalent mutation, all equivalent BVP mutations that led to *in vivo* rescue in yeast also resulted in *C. elegans* rescue. R5A was the only studied case where phosphoenzyme formation was impaired without loss of ATPase activity. The fact that it did not rescue in *C. elegans* is interesting because the equivalent residue is necessary for triphosphatase activity of the mammalian capping enzymes (Changela *et al.*, 2005). This implies that in PIR-1 this conserved residue may have a different function, perhaps related to its unique diphosphatase activity. This activity is not taken into account in the yeast capping rescue assay, as capping strictly requires removal of the γ-phosphate only, but the essential activity of PIR-1 may require removal of both γ- and β-phosphates from its RNA substrates. We also tested another critical arginine residue outside the P-loop (R153) that, as in BVP, did not lead to rescue in *C. elegans* despite being expressed, and was therefore demonstrated to be functionally conserved.

The fact that the remaining mutants rescued PIR-1 activity robustly (both somatic growth and laying of viable embryos occurred, even at 25˚C), could mean that the approaches used to measure BVP activity should not be extrapolated to *in vivo* activity in *C. elegans*. Perhaps *in vivo*, even a weak PIR-1 activity is sufficient to restore function. Since this collection of mutants was not comprehensive, it is still possible that a more systematic approach could yield a version of PIR-1 with the desired partial rescue activity. Moreover, of the lines tested, only three were integrated as single copy insertions (C150S, T151S, and K118R). It is thus conceivable that in non-integrated lines, weakened PIR-1 activity is compensated by overexpression from multi-copy extrachromosomal arrays. This line of investigation could therefore still provide useful strains for further study.

Lastly, when we searched for additional motifs in *C. elegans* PIR-1 using the ScanProsite tool (De Castro *et al.*, 2006, available at ExPASy.org) we found that it contains a low-confidence bipartite nuclear localization signal (NLS) sequence near its C-terminus, comprising residues 194-208 (highlighted in Fig. 2.8). In human PIR1 a low-confidence bipartite NLS is also predicted near the N-terminus (residues 13-27). However such a sequence cannot be found in the orthologs of mouse, fly and the nematode *C. briggsae*. Since this could explain its nuclear localization we created a few mutants and assayed their ability to rescue and to express nuclear PIR-1::GFP. Typically, interfering with basic residues such as lysines and arginines results in disruption of the signal (Lange *et al.*, 2007). When we replaced the first two lysines with alanines (K194A; K195A), the mutant construct was unable to rescue. Conversely, a conservative substitution with arginines was able to restore rescuing activity. However, because we could never see GFP

signal or recover PIR-1::GFP by immunoprecipitation from the non-rescuing mutant, we concluded that significantly altering these two residues made the protein unstable and could not infer the ability of the sequence to act as an NLS. Predictably, deleting the entire sequence (Δ194-208) also did not lead to rescue or GFP expression. In contrast, when we changed the two lysines of the second portion to alanines (K204A; K205A) or when we deleted the entire second part of the motif (Δ204-208) these sequences were able to fully rescue *pir-1* mutants and accumulate in the nucleus. We therefore conclude that this sequence does not represent a true NLS.

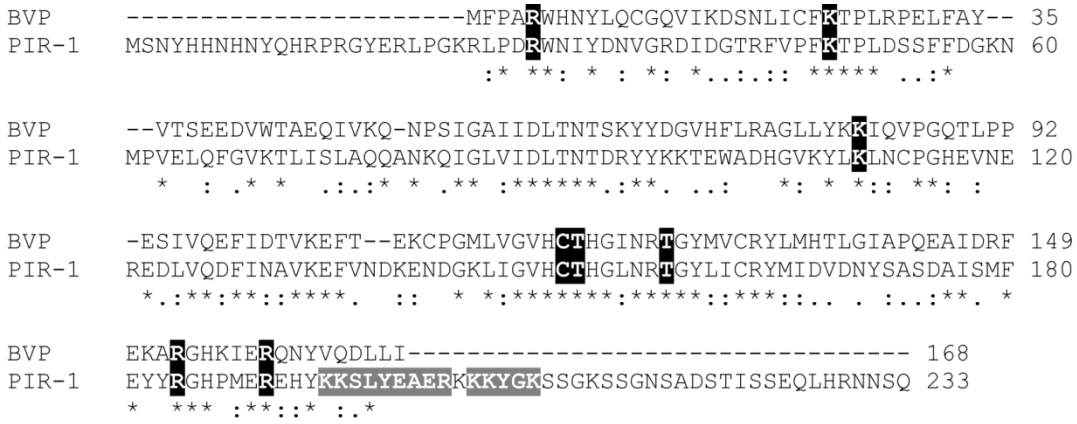



**Figure 2.8. Alignment of BVP and PIR-1 with residues used for mutagenesis.** Conserved residues are highlighted in black in the order that they appear in Table 2.1. Residues highlighted in gray show the putative bipartite NLS that was also mutagenized and tested for rescue.

**Table 2.1.** Rescue Assay of *pir-1(tm3198)* Defects by Different Point Mutants of PIR-1::GFP Based on Studies of Baculovirus Phosphatase

| BVP nutant | ATPase Activity (% of WT) [a] | $^{32}$P-BVP (% of WT) [b] | complemen-tation of *cet1Δ* yeast [c] | equivalent PIR-1 mutant | *C.elegans* rescue at 20 and 25˚C? | roles based on BVP structure and activity [d] |
|---|---|---|---|---|---|---|
| **R5A** | 95 | 8 | ++(ts) | **R28A** | No | lines the rim of the active site pocket perhaps interacting with substrate |
| **K25R** | 11 | 10 | ++(ts) | **K76R** | Yes | may stabilize structure around entrance to active site pocket |
| **K82A** | 7 | 83 | +(ts) | **K118A** | Yes | important for structure of loop that moves upon substrate binding |
| **K82R** | 17 | 100 | ++(ts) | **K118R** | Yes | " |
| **C119S** | <0.1 | n.d. | - | **C150S** | No | in the P-loop; forms covalent phospho-cysteinyl intermediate |
| **T120S** | 26 | n.d. | ++(ts) | **T151S** | Yes | in the P-loop |
| **T126A** | 8 | n.d. | ++ | **T157A** | Yes | in the P-loop |
| **R153A** | <0.1 | 2 | - | **R184A** | No | stabilizes P-loop in an active conformation |
| **R159K** | 28 | 22 | ++ | **R190K** | no lines obtained | lines the rim of the active site pocket perhaps interacting with substrate |

[a] From Martins and Shuman, 2000. [b] From Martins and Shuman, 2002b. [c] From Martins and Shuman, 2002a and Martins and Shuman, 2002b. [d] From Martins and Shuman, 2002a, Martins and Shuman, 2002b and Changela *et al.*, 2005. 'n.d.', not determined.

## Contributions

All experiments in this chapter were performed by Daniel Chaves.

## DISCUSSION

The results described in this chapter establish that the highly conserved RNA phosphatase PIR-1 is essential for *C. elegans* development. From the phenotypes of the novel loss-of-function allele *pir-1(tm3198)*, it is evident that PIR-1 is required for both somatic and germline growth. Consistent with this, the nearly ubiquitous expression pattern revealed by rescuing *pir-1::gfp* transgenes shows that PIR-1 is present throughout development. The germline temporal and spatial expression, in particular, is highly suggestive of PIR-1 being involved in the control of cell proliferation and sperm differentiation. Importantly, the ability of PIR-1 to rescue the severe mutant phenotypes requires a functional phosphatase active site. In this discussion we place this initial characterization of *C. elegans* PIR-1 in the context of previous studies and comment on additional observations.

### Maternal Load of *pir-1* mRNA and *pir-1* RNAi

A recent study in which the mRNA profiles of oocytes, one-cell and two-cell embryos are compared, shows that *pir-1* mRNA is very abundant in these three cell types (Stoeckius *et al.*, 2014) but that its levels continually decrease during development until the L3 stage. During and after L4, coinciding with germline development, they increase gradually (Marlon Stoeckius, personal communication). This pattern is consistent with our developmental qRT-PCR measurements of *pir-1* mRNA. Additionally, the fact that both oocytes and one to four cell embryos seem to express very little or no PIR-1 protein (Fig. 2.5, page 82) leads us to conclude that the maternal contribution occurs predominately at the mRNA level. The inherited *pir-1* mRNA in *pir-1(tm3198)* homozygote embryos is presumably translated during development. In the absence of zygotic *pir-1* transcription, the gradual dilution and degradation of maternal transcripts and resulting PIR-1 protein eventually lead to the developmental arrest we observe. The fact that *pir-1(tm3198)* homozygotes arrest at a variety of stages ranging from L3-like larvae to small young adult sterile animals likely reflects slight differences in the rate at which maternal PIR-1 is depleted from one animal to the other.

Despite having shown that *pir-1* mRNA can be specifically knocked down by RNAi through feeding (see Materials and Methods), the fact that we could not obtain phenotypes by RNAi even by injection, did not allow us to determine whether PIR-1 is required for development at earlier stages. In agreement with our experiments, three large

scale RNAi screens where *pir-1* RNAi was included found no associated phenotypes (Kamath *et al.*, 2003 and Rual *et al.*, 2004 both by feeding, and Soennichsen *et al.*, 2005 by injection). One exception was the study of Piano *et al.*, 2002, where injection of dsRNA against the genomic *pir-1* sequence led to 30-80% embryonic lethality and less than 10% of the progeny gave rise to adults that burst at the vulva. We have difficulty reconciling this result with our thorough attempts at silencing *pir-1* (including the same targeted sequence), but, if true, it is evidence that PIR-1 is also essential for embryonic development. This would agree well with PIR-1 protein expression onset in early embryonic development, with a gradual accumulation throughout embryogenesis.

### PIR-1 Is Required for Cell Proliferation

Our analysis of germlines from arrested *pir-1* mutant animals revealed that they are much smaller than what would be expected for the overall size of the animal. Germline growth seems to lag behind somatic growth, indicating that the highly proliferative germline tissue is very sensitive to PIR-1 loss. In arrested *pir-1* adults, the chromatin of germline nuclei undergoes normal changes during progression along the distal-proximal axis, qualitatively identical to those of an L4 germline. The distinction is the dramatically reduced number of nuclei present in each region of the germline. Additionally, we examined the distribution of P granules (marked by PGL-1) which associate with the outer nuclear membrane of every germline cell which has not yet committed to spermatogenesis or oogenesis. P granules had a wild-type appearance in all pre-spermatogenesis nuclei of *pir-1* germlines, suggesting that the arrest was not due to loss of germline cell identity. These results pointed to a defect in overall proliferation of the germline, and not to a block induced by aberrant early differentiation of germline cells. In addition to this germline defect, the larval arrested animals and the defective protruding vulvae of adults are highly reminiscent of defects observed in numerous mutants for proteins required for cell division, such as the cyclin E homolog of *C. elegans*, CYE-1 (Fay and Han, 2000; Brodigan *et al.*, 2003).

To complement our observations of *pir-1* mutants, we characterized the pattern of expression of rescuing PIR-1::GFP along development. We concluded that PIR-1 reaches its highest expression levels at stages of intense cell division, including not only the distal region of the hermaphrodite germline but also embryogenesis. Although our inability to silence *pir-1* by RNAi did not allow us to investigate whether it is essential for embryogenesis, its abundant and increasing expression during this stage suggests that it

may be. Moreover, since a large proportion of animals do not grow beyond L3-L4-like stages, it may also be required to promote post-embryonic cell divisions. Again, this correlates well with a high accumulation of PIR-1::GFP in somatic cells of early larval stages. In male germlines, rather than being strongly expressed in mitotic germ cell nuclei, PIR-1 is very highly expressed in the two distal tip cells (DTCs) of the male germline. This suggests that PIR-1 may stimulate male stem cell division via its DTCs. Considering how little is known about male DTCs, this is a difficult problem to address. Ideally, PIR-1 could be expressed from a male DTC-specific promoter, in both its wild-type and catalytically inactive versions, to investigate changes in germline proliferation of *pir-1* males and establish a direct cell-division stimulatory role. Additionally, the question of whether PIR-1 is expressed in the DTCs of hermaphrodites needs to be addressed with co-expression of DTC markers. Further exploration of this surprising difference could contribute to our sparse knowledge of how sexual dimorphism is achieved in the *C. elegans* germline.

In further support of a PIR-1 role in promoting cell proliferation we have made some observations regarding genetic interactions with key players in germline stem cell control (reviewed in Kimble and Crittenden, 2007). We have crossed the *pir-1(tm3198)* mutation into *glp-1* and *gld-1* mutant backgrounds that interfere with germline development. GLP-1 is a Notch-family receptor that along with its ligand, LAG-2 (*L*IN-12 *a*nd *G*LP-1 Phenotype 2), promote germline mitoses. In hermaphrodites, LAG-2 is expressed in the DTC, confining mitosis to the proximal germline. Losing GLP-1 activity leads cells to enter meiosis prematurely, while constitutive activity generates a germline tumor in which cells are unable to leave the mitotic cycle (Austin and Kimble, 1987; Crittenden *et al.*, 1994; Berry *et al.*, 1997; Pepper *et al.*, 2003). GLD-1 is an RNA binding protein that represses translation of GLP-1 and other mRNAs required for mitosis by binding to specific sites on their 5' or 3' UTRs. Its expression starts shortly before the transition zone, leading cells to enter meiosis by limiting, among others, GLP-1 translation (reviewed in Lee and Schedl, 2010). GLD-1 loss-of-function alleles lead to germline tumors by inducing meiosis-committed cells in pachytene to re-enter a mitotic cycle and proliferating uncontrollably (Francis *et al.*, 1995; Jones *et al.*, 1996). When combined with the *glp-1(ar202)* temperature-dependent gain-of-function allele (Pepper *et al.*, 2003), *pir-1* homozygotes grown at 15˚C, like *glp-1(ar202)* on its own, do not develop a germline tumor. However, when grown at 25˚C, at which GLP-1 becomes constitutively active, *pir-1* mutants form a large germline tumor, likely at the expense of meiosis and while maternal PIR-1 is still available. Conversely, the *gld-1(q485)* loss-of-function allele (Francis *et al.*,

1995), which leads to a germline tumor on its own, is not able to form one when combined with *pir-1* (*i.e.*, double mutants look like *pir-1* homozygotes). This genetic interaction suggests that PIR-1 may be a factor involved in regulation of mitotic cell divisions downstream of GLP-1 and upstream of GLD-1. The inability of germline cells to reenter mitosis in the *pir-1; gld-1* mutant is consistent with a positive role in mitotic proliferation. We believe that these observations should be expanded, namely by (1) investigating further genetic interactions with other key factors in germline stem cell regulation; by (2) directly comparing the expression of such factors by immunofluorescence or fluorescent reporters with PIR-1 expression; and (3) by studying their distribution in *pir-1* mutant germlines and, complementarily, PIR-1::GFP distribution in germline tumors.

Consistent with the genetic interaction between *pir-1* and *gld-1*, a study found *pir-1* mRNA to be significantly enriched among a set of RNAs isolated from non-cross-linked GLD-1 immunoprecipitates in young adult hermaphrodites (Wright *et al.*, 2011). A subsequent study that employed *in vivo* PAR-CLIP (i*n vivo P*hoto*a*ctivatable *R*ibonucleoside-Enhanced UV *C*ross-*L*inking and *IP*) to identify GLD-1 targets, also found *pir-1* mRNA among 439 reproducibly identified mRNAs and that, upon RNAi of *gld-1*, the protein levels of all 439 targets were specifically upregulated (Jungkamp *et al.*, 2011). Together, these findings established *pir-1* mRNA as a target for GLD-1-mediated translational repression. This correlates well with our observation that PIR-1::GFP expression is reduced in germ cell nuclei undergoing the pachytene stage of meiosis, where GLD-1 protein expression is known to be highest (Fig. 2.7A, page 85). Since exit from mitosis depends on GLD-1 repression, this implies that PIR-1 expression is particularly important in actively proliferating cells, but that it needs to be downregulated during the transition from mitosis to meiosis.

Somewhat contrary to the idea that in *C. elegans* PIR-1 promotes cell proliferation, a study of human PIR1 (or DUSP11) demonstrated that its over-expression in U2OS osteosarcoma cells inhibited cell proliferation, while RNAi knock-down stimulated proliferation, even during exposure to DNA damaging agents (Caprara *et al.*, 2009). The proliferation inhibitory effect depended on the phosphatase activity of PIR1, as a catalytically inactive mutant protein mutation failed to suppress cell division. The aforementioned study established that *PIR1* mRNA is directly bound by p53 and that, following DNA damage, the level of PIR1 protein increases in a p53-dependent manner. These results imply that, at least in the context of cancer, PIR1 assumes the role of a tumor suppressing enzyme. Paradoxically, *PIR1* mRNA levels obtained by mRNA-seq for 39

cancer cell lines from various organs and tissues are either very high for 18 cell lines or moderate for eight lines (data from the Cancer Genomics Hub deposited in the Expression Atlas database from the EMBL-EBI; Petryszak *et al.*, 2014). While this dispels the idea that a high level of PIR1 necessarily leads to a dampening of cell proliferation, it does not prove whether high *PIR1* expression in these lines is a cause or a consequence of unrestricted proliferation. Future studies using a wider range of conditions and cell lines will be necessary to address these intriguing results. As a final note, since p53 is conserved in *C. elegans* (CEP-1; *C.* e*legans p*53-Like Protein 1) it would be interesting to recapitulate these experiments in worms, by using the germline to measure effects on cell proliferation. In the course of our studies, we did produce a *pir-1; cep-1* loss-of-function double mutant, predicting that loss of p53-regulated cell-cycle checkpoints could lead to increased germline cell proliferation in the *pir-1* mutant background. These double mutant animals turned out to be phenotypically identical to the *pir-1* single mutant, ruling out that the proliferative arrest of *pir-1* is imposed by the activation of CEP-1/p53.

## The Lack of Nuclear PIR-1 in Specific Cells Suggests a Correlation with Transcriptional Activity

Nuclear PIR-1 is conspicuously absent from the nuclei of oocytes, in embryos until the four-cell stage, and in the two primordial germ cells, Z2 and Z3. One important common characteristic of these cell types is that they are transcriptionally quiescent. In oocytes, mRNA transcription is repressed as the cells enter diakinesis (Walker *et al.*, 2007) until the four-cell stage, after which mRNA transcription is resumed (Seydoux *et al.*, 1996; Seydoux and Dunn, 1997). In the developing embryo, the only cell that remains transcriptionally silent is the germ precursor cell, which at around the 100-cell stage divides into two cells, Z2 and Z3. These will only proliferate and become transcriptionally active post-embryonically, during the L1 stage, to give rise to all germline cells (Bender *et al.*, 2004; Schaner *et al.*, 2003). This striking overlap suggests that PIR-1 accumulates in nuclei undergoing active transcription, and reinforces the idea that its endogenous substrates are RNA molecules. Moreover, polyploid intestinal cell nuclei are the somatic cells with the highest accumulation of nuclear PIR-1. The intestine is the largest organ in *C. elegans* and one could assume that polyploidy evolved to support the demanding activity of this tissue by increasing transcriptional output. This would fit well with the hypothesis that PIR-1 accumulates in a transcription-dependent manner. The fact that we could see an overlap of PIR-1 with chromatin in intestinal nuclei is also in agreement with

this idea. Alternatively, PIR-1 could play a part in DNA replication during endoreduplication, compatible with its potential stimulatory effect on cell proliferation. Interestingly, a study established that genetically preventing or increasing ploidy in *C. elegans* leads to smaller or larger than wild-type adult sizes, respectively (Lozano *et al.*, 2006). Although the study focused on ploidy of hypodermal cells, future experiments in *pir-1* mutants could address the DNA content of both hypodermal and intestinal cells to determine whether DNA replication is impaired when PIR-1 is not present, thereby interfering with overall growth. Finally, in one additional provocative connection to DNA replication, a MudPIT-based proteomic survey of proteins associated with human TRF-2 (*T*elomeric *R*epeat-Binding *F*actor 2) found PIR1 as an interactor (Giannone *et al.*, 2010). TRF-2 is a crucial telomere repeat binding factor, and part of a complex involved in all aspects of telomere function, including telomere replication. Despite being just one among many TRF-2-interacting partners, *C. elegans* could be a good system in which to further validate and explore this association.

Several studies of human PIR1 have pointed to connections with RNA processing. Human PIR1 was found to localize to nuclei when overexpressed in HeLa cells and this association disappeared when cells were pretreated with RNase A, but not DNase I (Yuan *et al.*, 1998). This study also described interaction of PIR1 by yeast two-hybrid with splicing factors 9G8 (SRSF7, Splicing Factor, arginine/serine-rich 7), and SRp30C (SFRS9, Splicing Factor, arginine/serine-rich 9), both components of mRNA-splicing complexes. Another yeast two-hybrid screen from Caprara *et al.* (2009) revealed PIR-1 to interact with SAM68 (*S*rc-*A*ssociated Substrate in *M*itosis of *68* kDa), an RNA binding protein that has been implicated in multiple aspects of RNA metabolism, including alternative splicing, transcription and RNA transport (reviewed in Sanchez-Jimenez and Sanchez-Margalet, 2013). Importantly, this interaction was also validated *in vivo*. In another study, where the HeLa cell mRNA interactome was studied by mass-spectrometry of proteins associated with poly-adenylated RNA, PIR1 was recovered with relatively high abundance (Castello *et al.*, 2012). Taken together these studies heavily suggest activities in the control of gene expression, namely by mediating aspects of mRNA processing. The fact that *C. elegans* PIR-1 is expressed in most post-mitotic adult cells is compatible with a role in mRNA metabolism, and that perhaps this is the way in which it supports the proliferation of cells and growth.

We have performed experiments to address the involvement of PIR-1 in general transcription but have failed to obtain conclusive results (not shown). Specifically, we

immunostained Pol II in *pir-1* germlines, including the initiating form (CTD repeat phosphorylated on Ser5) and elongating form (Ser2 phosphorylation), to find that Pol II still localized to germline nuclei with a similar distribution and intensity as in wild-type germlines. This experiment was limited given that the *pir-1* arrest occurs prior to oogenesis, during which the most dramatic changes in transcription take place. We also knocked down *ama-1* (*Ama*nitin Resistant 1; encoding the large subunit of Pol II) by RNAi in the *pir-1::gfp* background. When L1s or L2s are fed *ama-1* dsRNA-expressing bacteria, animals arrest mostly as L3s or sterile adults, respectively. Upon arrest and a few days thereafter, we looked for decreased levels of nuclear PIR-1::GFP under the microscope but were unable to detect differences (in Chapter III further experiments are shown). In another approach, we exposed animals of different developmental stages to a range of concentrations and exposure periods of the transcription inhibiting drugs DRB (5,6-dichloro-1-β-D-ribobenzimidazole) and actinomycin D, but again failed to see an effect on PIR-1::GFP. Furthermore, we performed UV cross-linking followed by IP and RNA-seq (CLIP-seq) of PIR-1::GFP but obtained reads that resembled those of our negative control, deriving mostly from abundant tRNA fragments. Although the CLIP conditions were extensively optimized, it is possible that further optimization is required, or that the concentration of PIR-1 binding to RNA is low compared to the unbound fraction (in both nucleus and cytoplasm), severely limiting the sensitivity of this approach.

Finally, it is important to consider that human PIR1 was detected in a thorough proteomic study of the nucleoli of HeLa cells (Andersen *et al.*, 2005). PIR1 was gradually depleted from nucleoli with increasing times of exposure to actinomycin D, implying not only that the association is transcription-dependent but also that human PIR1 may participate in ribosomal RNA (rRNA) transcription or processing. Another, more recent, study showed that this is indeed the case (Tafforeau *et al.*, 2013). In that study, 625 proteins associated with the nucleoli of HeLa cells were individually downregulated via RNAi and the effects on pre-rRNA processing were assessed by quantitative northern blotting. Knockdown of PIR1 led to a strong accumulation of the 47S primary transcript (containing the 18S, 5.8S and 28S rRNAs), of a 34S RNA intermediate, and of a 5' external transcribed spacer fragment. However, as the levels of mature 28S and 18S remained similar to wild type, and silencing of 50 other nucleolar proteins led to the same pattern of accumulated rRNA precursors, it is likely that some proteins act redundantly at these processing steps. Consistent with this possibility is the fact that knockdown of PIR1 in cultured cells leads to increased cell proliferation, which is imcompatible with a loss of

functional ribosomes (Caprara *et al.*, 2009). Regarding *C. elegans*, when we looked at 28S, 18S and 5S rRNA bands from total RNA of seven-day old *pir-1* mutant animals, their intensity and migration on an agarose gel stained by ethidium bromide and visualized under UV light was no different than in wild-type animals (not shown). While this indicates that there are no gross abnormalities in rRNA accumulation, a more complete study of rRNAs and its intermediates should be performed in *pir-1* mutant animals. A deficiency in the processing of rRNA could explain the growth and proliferation defects that we observe.

Curiously, a study of the *C. elegans* nucleostemin homolog *nst-1* shows that the deletion mutant phenotype as well as the developmental expression pattern of this protein is highly reminiscent of PIR-1 (Kudron and Reinke, 2008). Mammalian nucleostemin is a predominately nucleolar GTP-binding protein that actively shuttles between the nucleolus and the nucleoplasm, and is crucial for several aspects of development (reviewed in Tsai and Meng, 2009), cancer progression (Lin *et al.*, 2010) and regulation of stem cell pluripotency (Qu and Bishop, 2012), among other processes. In *C. elegans*, *nst-1* loss-of-function mutant animals arrest as L1 and L2 larvae, but, similarly to *pir-1*, retain a wild-type lifespan during which they feed and move normally. NST-1::GFP concentrates in nucleoli and is diffusely present in the nucleoplasm at all developmental stages. Its expression is reduced in the most proximal oocyte and shuts off in early embryos, only to be turned back on around the 18-cell stage embryo. During this time window, despite continuing rRNA transcription, no processing of rRNA or ribosome assembly takes place (Saijou *et al.*, 2004). Consistent with the nucleolar localization of NST-1 and absence from cells where rRNA processing is halted, *nst-1* mutant animals were shown to have a marked decrease in rRNA levels. This effect is thought to lead to fewer mature ribosomes and cause the developmental arrest. Similar effects have been reported in *S. cerevisae*, the fission yeast *Schyzosaccharomyces pombe* and flies. Studies of vertebrate nucleostemin, on the other hand, are contradictory when it comes to its role in promoting high levels of rRNA and ribosome assembly, which may be carried out instead by the paralog GNL3L. Mounting evidence now points to a role for vertebrate nucleostemin in maintaining genome integrity during DNA replication in the nucleoplasm via the p53 pathway and even to a telomere-protective role (reviewed in Tsai, 2014). It remains possible that in *C. elegans*, in addition to ribosome biogenesis, NST-1 takes on similar roles. When we placed the *nst-1::gfp* transgene in the *pir-1::3xFlag* background and performed reciprocal IPs, there was no indication that the two proteins stably interact. Nonetheless, given the results

gathered from human cells, and the striking similarity between the mutant phenotypes and pattern of expression of *C. elegans* NST-1 and those of PIR-1, a potential role for this enzyme in rRNA biogenesis should be thorougly investigated.

### PIR-1 Expression Patterns and Defects in *pir-1* Mutant Animals Uncover a Role in Sperm Development and Link It to the Eri Pathway

One of the clearest phenotypes that result from the *tm3198 pir-1* deletion is the inability of spermatocytes to correctly divide into mature spermatids. This is manifested by the formation of chromatin bridges between dividing spermatocytes, in everything similar to what has been reported for temperature-sensitive sterile Eri (enhancement of RNAi) mutants (Gent *et al.,* 2009; Pavelec *et al.,* 2009). This phenotype was not detected in the earlier *pir-1(tm1496)* arrested mutants because they never reached spermatogenesis. We have tested *pir-1(tm3198)* mutant animals for enhancement of RNAi by feeding them bacteria expressing dsRNA against genes that only manifest a phenotype in Eri RNAi-hypersensitive strains. As explained in Chapter I, the Eri phenotype results from a loss of ERGO-1 26G-RNAs, leading to a release of downstream 22G-RNA synthesis proteins and Argonautes that become available to mediate a more potent exogenously triggered RNAi response. Since *pir-1* animals are sterile, we were limited to testing Eri targets for which the phenotype appeared in the parental (P0) generation by feeding animals starting at the L1 stage. We tested *unc-73* and *dpy-13*, which give rise to an uncoordinated movement phenotype and a severe dumpy phenotype, respectively, in Eri mutants (Simmer *et al.*, 2002; Kennedy *et al.*, 2004; Duchaine *et al.*, 2006). Unlike the positive controls (*eri-1* and *rrf-3* mutants), *pir-1* mutant animals did not exhibit the phenotypes, arguing that PIR-1 may not participate in the ERGO-1 pathway. Due to maternally derived PIR-1, however, and because RNAi at later developmental stages generally does not lead to phenotypes in Eri mutants, we could not definitively conclude that *pir-1* mutant animals are not Eri.

Together with the sperm defect, the elevated expression of PIR-1 during spermatogenesis in both hermaphrodites and males make a very strong case for its direct involvement in sperm development. The ERI complex is required for ALG-3/4-dependent 26G-RNAs, which were shown to be essential for proper sperm maturation (Han *et al.*, 2009, Conine *et al.*, 2010, Conine *et al.*, 2014). Unlike *ergo-1* mutants, *alg-3; alg-4* mutants are not hypersensitive to RNAi, consistent with a role confined to spermatogenesis (Han *et al.*, 2009). This potential association is explored in Chapter IV. Importantly, PIR-1 may be required for sperm development beyond *C. elegans*. In *Drosophila melanogaster*

the mRNA of the PIR-1 ortholog CG13197 was found to be expressed in almost every adult fly tissue tested, but was particularly elevated in testes, followed by ovaries (from FlyAtlas Anatomical Expression Microarray data and modENCODE tissue expression mRNA-seq data; Chintapalli *et al.*, 2007; Celniker *et al.*, 2009). In collaboration with Keith Boundy from the laboratory of Phillip Zamore in our department, we found that a P-element insertion fly mutant (*P{wHy}CG13197*$^{DG30703}$) that leads to a pronounced decrease in CG13197 mRNA expression resulted in sterile adult males. Until the phenotype is properly rescued this observation remains preliminary, but it is nonetheless indicative that PIR-1 function in spermatogenesis may be evolutionarily conserved. Further exploration of the expression databases shows that in vertebrates *PIR1* is expressed to varying degrees in most tissues, including testes. Extended studies of vertebrate PIR1 are therefore worthwhile, not only regarding sperm development, but also general development and cancer.

## MATERIALS AND METHODS

Note: As many of the basic techniques for handling *C. elegans* are available through several printed and online sources (namely the WormMethods section in WormBook.org) the methods described in this thesis are not meant to exhaustively describe general *C. elegans* techniques. Detailed information is given for procedures that have either been created during this project or adapted from existing protocols, or where it is thought to help the reader understand and interpret the described results.

### Outcrossing and Balancing of the *pir-1(tm3198)* Deletion Allele

Due to its sterility phenotype, the new *tm3198* deletion allele was sent to us as a mixed population of wild-type heterozygotes. In general, when mutations are generated by chemical mutagenesis other potential mutations in the background must be eliminated by successive genetic crosses with a wild-type strain, in a process termed outcrossing. This allows replacement of most of the original genome by intact genome sequence through recombination events at each cross. The deletion of interest is followed throughout the process by PCR-based genotyping. Primers flanking the deletion were designed (primers O1/O2, see Appendix B for sequences) to identify individual worms carrying the mutation by single worm PCR. For this single worms were lysed in 5 µl of worm lysis buffer (10 mM Tris-Cl pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$, 0.45% NP-40, 0.45% Tween-20, 0.01% gelatin) with 0.4 µg/µl Proteinase K in PCR tubes at 65˚C for 1 hr, followed by 15 minutes at 95˚C. Lysates were mixed with 5 µl of ultrapure water and 1 µl was used as template in a 25 µl PCR reaction using Roche Taq polymerase. A 56˚C primer annealing temperature and a 68˚C extension temperature for 30x cycles were used. 20 µl of PCR reaction were run on a 1% agarose/1X TAE buffer gel (1X TAE is 40 mM Tris, 20 mM acetic acid, 1 mM EDTA, with a final pH of 8.0) to assess presence or absence of the deletion according to band size. This genotyping method was used to generate all strains carrying deletion mutations or transgenes that did not have an easily identifiable phenotype or visible fluorescent tags. Deletion heterozygote hermaphrodites were then mated with N2 males (first outcross). Cross-progeny males were used to establish several cross plates with one male and three N2 hermaphrodites per plate. After mating, hermaphrodites were transferred to single plates (or "singled") and males were genotyped for the deletion. From the cross-progeny of heterozygote males at least 10 F1 males were selected to set up more individual crosses with N2 hermaphrodites. This process was repeated three more times for

a total of five outcrosses. The final outcross was done by mating heterozygote males with already outcrossed *pir-1(tm1496)/mnC1** balanced hermaphrodites. This genetic balancer carries a bright, ubiquitously expressed nuclear GFP ($P_{sur\text{-}5}$*::gfp*) that makes its carriers easily traceable under a fluorescence dissection microscope. GFP-positive (carrying *mnC1**) hermaphrodite F1cross-progeny were then singled, allowed to have progeny and genotyped for *tm3198*. Heterozygote animals were the final 6x outcrossed balanced strain that was the base for making all other *pir-1* strains in this study.

## Worm Culture, Synchronization and Developmental staging

Large scale cultures of worms were generally grown on 15 cm Nematode Growth Medium (NGM; 1.7% (w/v) agar, 50 mM NaCl, 0.25% (w/v) peptone, 1 mM $CaCl_2$, 5 µg/ml cholesterol, 25 mM $KH_2PO_4$, 1 mM $MgSO_4$) with a 1:1 mix of agar and agarose to prevent burrowing of the worms into the medium when growing them for extended time periods. For most experiments worms were grown at 20˚C and fed with concentrated OP50 *E. coli*. For cleaning and synchronizing populations, starved small plates were chunked onto large 15 cm plates with OP50 and allowed to grow until a large fraction of worms were gravid. They were then harvested with 1X M9 buffer (22 mM $KH_2PO_4$, 42 mM $Na_2HPO_4$, 85 mM NaCl, 1 mM $MgSO_4$) into a 15 ml conical tube and bleached to recover embryos with bleach solution (10% bleach, 0.25 M KOH). Embryos were washed 3x with M9 buffer, and allowed to hatch in M9 buffer while shaking overnight at room temperature or 20˚C. To obtain synchronous populations at different stages, live hatched L1 worms were counted and the desired number of worms was transferred to fresh 15 ml tubes and centrifuged at 800x *g* for 1-2 minutes. The resulting pellet was mixed with OP50 and spread on a fresh NGM plate. As a rule, for a large plate of gravid adults, no more than 150,000 worms were plated, since higher densities would cause animals to enter the dauer stage and not reach adulthood. For developmental staging of N2 animals at 20˚C the different stages were generally harvested at the following times: L1 larvae at 6 hours after plating, L2s at 15-18 hours, L3s at 24-26, L4s at 34-36 and young adults at 46-48. Embryos were obtained from just-bleached gravid adults. These timings varied with strain, and staging was always visually monitored for features that characterize each stage. Animals were harvested and washed 3x with 15 ml of M9 buffer in 15 ml conical tubes with 20-second centrifugations at 800x g. They were then incubated with 10 ml of M9 buffer for 15-30 minutes in a rocking platform to allow full digestion of ingested bacteria, followed by one wash with ice-cold distilled water to remove salts and paralyze worms to

form a compact pellet that allowed aspiration of as much water as possible. Pellets were flash-frozen in a dry ice/ethanol bath and stored at -80˚C or processed immediately for RNA or protein extraction.

### Ivermectin-Based Counter-Selection

Counter-selection using ivermectin was employed in this work in order to obtain large numbers of *pir-1* arrested homozygote animals to provide enough RNA and protein for many of the experiments performed. Ivermectin belongs to a class of macrocyclic lactones called avermectins, which interact with vertebrate and invertebrate GABA receptors and invertebrate glutamate-gated chloride channels. The *pir-1* deletion alleles were crossed into an ivermectin-resistant triple mutant genetic background comprising the genes *avr-14*, *avr-15* and *glc-1*, which encode glutamate-gated chloride channel subunits (Dent *et al.*, 2000). Throughout this thesis this mutant combination is referred to as 'avr3x.' Expression of the wild-type sequence of any one of the three genes is sufficient to restore sensitivity to ivermectin. The genetic balancer *mnC1** used to propagate *pir-1* mutants contains a wild-type copy of *avr-15* under the control of the pharyngeal *myo-2* promoter. This causes heterozygote carriers and balancer homozygote worms to become sensitive to ivermectin, due to compromised pharyngeal muscle activity that makes them unable to feed. Conversely, balancer-free *pir-1* homozygotes are able to grow in the presence of the drug.

In order to place *pir-1(tm3198)* in the avr3x background, avr3x males were crossed with *pir-1(tm3198)/mnC1**. Several (~10-15) GFP-negative hermaphrodite cross-progeny (balancer-free *pir-1* heterozygotes) were singled into small plates, and allowed to self-fertilize until the plates were starved (about two generations). These plates were then chunked onto large ivermectin plates with OP50. After three days plates were screened for healthy gravid adults, which were singled onto small plates, allowed to have progeny and genotyped for the presence of the deletion allele. Progeny from positive mothers were mated with *pir-1(tm1496)/mnC1*;* avr3x males, and GFP-positive (with balancer) cross progeny were singled and allowed to have progeny. Animals carrying *tm3198* were identified by the arrested phenotype, confirmed by PCR and tested for ivermectin resistance.

Balanced *pir-1* populations were expanded in regular NGM in order to obtain sufficient synchronous L1 larvae to plate on NGM supplemented with 25 µg/L of ivermectin. After plating, only *pir-1* homozygotes grew while all others never passed the

L1 stage. Many of these arrested larvae remain alive, however, making it important to eliminate them when harvesting the counter-selected population. For this, worms were centrifuged in M9 buffer for up to 10 times at lower speeds than usual (~600x *g* for brief seconds). While this led to loss of some *pir-1* arrested worms, it eliminated most live L1s.

### RNA Extraction and qRT-PCR

Worms were collected from plates and washed 3x with 15 ml of M9 buffer in 15 ml conical tubes with 20-second centrifugations at 800x *g*. They were then incubated with 10 ml of M9 buffer for 15-30 minutes in a rocking platform to allow full digestion of ingested bacteria, followed by one wash with 4˚C distilled water to wash away salts and paralyze worms thereby allowing formation of a compact pellet in order to remove as much water as possible. At least 5 volumes of TRI Reagent (MRC, Molecular Research Center) were added to pellets, which were either flash frozen in a dry ice/ethanol bath or processed immediately. Lysis was performed by crushing with a metal dounce with ~30 strokes at room temperature. Samples were aliquoted into 1.5 ml tubes and processed according to the manufacturer. BCP separation reagent (MRC) was used instead of chloroform. RNA pellets were dissolved in ultrapure water or TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA) depending on downstream experiments.

Before cDNA synthesis, 100 µg of RNA were pre-treated with 6 U of Turbo DNase (Ambion) in the presence of 60 U Superasin RNase inhibitor (Ambion) in a 100 µl volume at 37˚C for 1 hour. After phenol:chloroform extraction and precipitation, RNA was resuspended in ultrapure water and quantified in a NanoDrop spectrophotometer. For each 20 µl cDNA reaction, 2 µg of RNA were used, using Superscript III Reverse Transcriptase (Invitrogen) with random hexamers according to the manufacturer's instructions. qRT-PCR reactions were carried out using the ABI Prism 7500 Fast Sequence Detection System with Fast SYBR Green PCR Master Mix (Applied Biosystems). Each 15 µl reaction contained 7.5 µl of SYBR Green reagent, 400 nM of each primer and 2 µl of cDNA. The standard fast thermocycling program was used and for each primer pair a standard curve with 1:5 cDNA dilutions was generated to calculate mRNA amounts in samples. Per sample, 2-3 technical replicates were run. Primers used were O3/O4 (18S rRNA), O5/O6 (*gapdh/gpd-2* mRNA), O7/O8 (*sec-5* mRNA), O9/O10 (*pir-1* mRNA), and O11/O12 (*unc-22* mRNA), as listed in Appendix B.

### RNAi of *sec-5* and *pir-1*

To prepare dsRNA, primers containing T7 promoter tails were designed to amplify 200-400 nt sequences from cDNA using iProof high-fidelity DNA polymerase (Bio-Rad). Products were gel purified and used as templates for dsRNA synthesis using the MEGAscript T7 kit (Ambion) according to the manufacturer's instructions. dsRNA was verified on a 1% agarose/1X TAE gel, quantified, and microinjected in the body cavity or intestines of young adult worms at a concentration of 20 ng/µl and 1 µg/µl. In the case of *pir-1*, germline injections were also performed in an attempt to increase the potency of the treatment in this tissue. Phenotypes were scored in the progeny a day later at 20˚C after the injections. Primers used were O13/O14 for *pir-1* and O15/O16 for *sec-5*.

In addition to the available *pir-1* dsRNA construct for RNAi by feeding available from the Ahringer Lab RNAi library (Kamath *et al.*, 2003), containing a 1158 bp sequence amplified from genomic DNA in the L4440 double T7 vector from the Fire Lab *C. elegans* Vector Kit (1999, unpublished), two new constructs were generated. One was designed comprising the majority of the unspliced sequence of *pir-1* with flanking T7 promoters in the pCR-2.1-TOPO vector backbone (Invitrogen) and transformed into the RNase III-deficient *E. coli* strain HT115 (DE3), used for dsRNA feeding. The other construct used the same primers to amplify the spliced sequence from cDNA and cloned into the pCR-Blunt II-TOPO vector (Invitrogen; primers O17/O18).

RNAi food was prepared by growing a 3 ml overnight starter culture from a single HT115 *E. coli* colony on a tetracycline/ampicillin plate (or kanamycin for the pCR-Blunt vector) in LB medium (the tetracycline selects for a plasmid in HT115 encoding the T7 polymerase). 1 ml of starter culture was added to 1 liter of TB medium (Terrific Broth; 1 liter contains 12 g of tryptone, 24 of yeast extract, 4 ml of glycerol and 100 ml of 1 M Potassium Phosphate Buffer buffer pH 6.0 (0.72 M $KH_2PO_4$, 0.28 M $K_2HPO_4$)) supplemented with ampicillin (50 µg/ml) or kanamycin (50 µg/ml) and grown overnight at 37˚C. The following day IPTG was added to a concentration of 2 mM for induction of T7 polymerase expression and production of dsRNA and incubated for 3 hours at 37˚C. The bacteria were collected by centrifugation at 3,000x *g* for 15 minutes at 4˚C and resuspended in four pellet volumes of sterile M9 buffer. RNAi food was plated on NGM plates supplemented with 50 µg/ml of ampicillin and 1 mM of IPTG and left to dry and further induce at room temperature overnight and used immediately or stored at 4˚C.

Validation of the new dsRNA-expressing food was performed by feeding the rescued bombardment *pir-1; pir-1::gfp* in which a two-day exposure led to depletion of GFP signal in the bright intestinal nuclei, but not in the germline (further exposure did not change expression). Furthermore, qRT-PCR of *pir-1::gfp* mRNA following *gfp* RNAi by feeding showed a ~50% and ~40% reduction in mRNA level when using *pir-1* or *gpf* primers (O19/O20), respectively. Again, only somatic GFP signal was down-regulated, and no phenotypes were observed. RNAi targeting *pir-1* was therefore specific but was not enough to knock it down to a level that interfered with the activity of the protein.

### Generation of Tagged *pir-1* Constructs

In order to fuse expression tags to the *pir-1* gene, we adopted a strategy based on homologous recombination in *S. cerevisiae* of the tagged portion of the sequence of interest into a yeast artificial chromosome (YAC) carrying the full genomic locus. N- or C-terminally tagged versions of the *pir-1* gene (with either *gfp* or *3xFlag* sequences) were generated by amplifying 1 kb of sequence from genomic DNA surrounding the start or the stop codon of the gene with iProof high-fidelity enzyme (Bio-Rad). The primers were designed to make products containing terminal *Bam*HI sites (for start sequence; primers O21/22) or terminal *Xba*I sites (for stop sequence; O23/O24) that are not present in these sequences or the tag cassettes. The fragments were TOPO cloned into vector pCR-Blunt II TOPO (Invitrogen), insertion was verified by PCR of colonies using Roche Taq polymerase and some positive clones were sequenced. They were then cut with *Not*I, treated with Klenow (NEB) in the presence of dNTPs to fill-in the overhangs and religated with T4 DNA ligase (NEB) in order to destroy the vector's unique *Not*I site. After religation and transformation, the vectors were verified by *Not*I digestion (which should now not linearize the plasmid) then used for circular PCR with divergent primers that were designed to introduce a new *Not*I site immediately downstream of the start (primers O25/O26 and O27/O28) or upstream of the stop codon (primers O29/O30 and O31/O32). The PCR reaction was digested with *Dpn*I to destroy the original vector and visualized after agarose gel electrophoresis for verification of the linear product. The blunt ends (generated by the iProof polymerase) were phosphorylated with PNK (NEB) in the presence of ATP, then ligated and transformed. Plasmids from a few clones were digested with *Not*I to verify the creation of the new site. Two positive clones of each fragment were then digested with *Not*I followed by CIP (NEB) dephosphorylation. In parallel, tagged *Not*I site-flanked cassettes with *gfp* or *3xFlag* sequences were digested from vectors with *Not*I

and gel-purified. These cassettes contain the yeast ochre supessor tRNA, *sup4o*, embedded in a synthetic *C. elegans* intron, which allows selection of recombined sequences in a specific auxotrophic yeast strain (Rocheleau *et al.*, 1999). The purified fragments were phosphorylated with PNK, ligated to the vectors and transformed. Positive clones were identified by *Not*I digestion to release the inserted cassette and proper orientation was assessed by digestion with *Eco*RI, which cuts once within the cassettes and twice in the vectors. Clones were sequenced to screen for PCR errors and to confirm that the *pir-1* ORF was in the same frame as the *gfp* and *3xFlag* sequences.

Correct constructs were linearized with *Bam*HI or *Xba*I (to release *pir-1* sequences fused to tag cassettes from the vector backbones) and transformed into a yeast strain carrying the Y51C5 YAC that contains the *C. elegans pir-1* gene. Recombinant clones were selected through suppression of an ochre mutation in a lysine synthesis pathway gene, allowing it to grow on Ura- Lys- medium (unrecombined cells can only grow on Ura- medium). Genomic DNA was isolated from positive strains and recombination was confirmed with PCR using primers outside the insert to be recombined into the *pir-1* locus (O21/O22 for N-terminal and O23/O24 for C-terminal). Although these genomic DNAs were intended for direct injection into *pir-1(tm1496)/mnC1* animals for rescue, attempts to do it were unsuccessful. In the meantime the laboratory adopted the technique of transformation by bombardment, and we decided to use the tagged YACs to make bombardment-compatible constructs carrying the tagged *pir-1* sequence only.

## Generation of Constructs for Transformation by Bombardment

Using 250 ng of recombinant YAC genomic DNA per 25 µl reaction, sequences were amplified with iProof at a low cycle number (20) to include 1 kb upstream and 1 kb downstream of the *pir-1* start and stop codons, respectively (primers O33/O34). The sequences were then cloned into pCR-Blunt II-TOPO. Primers carrying Gateway (Invitrogen) cloning *att*B1 and *att*B2 sites (O35/O36) were used for PCR from the TOPO vectors (direct amplification from YAC DNA with the Gateway primers was not successful). PCR products were digested with *Dpn*I to eliminate the original vector (which carries a kanamycin resistance gene that is also present in the Gateway vector), gel purified, eluted with ultrapure water and quantified. The BP clonase (Invitrogen) recombination reaction to introduce the *pir-1* sequences into pDONR-201 (Invitrogen) was performed according to the manufacturer. Recombinant DH5α transformants were selected with kanamycin agar plates and through loss of the *ccdB* gene which is present only in the

unrecombined vector and causes bacterial death when expressed. For confirmation, the plasmids of clones were isolated by minipreparation and digested with *Not*I to release the *gfp* or *3xFlag* cassettes. Two pDONR clones of each type were sequenced.

For the next step, the pDONR vectors were recombined with pCG150 (pDESTR4-R3 Destination vector carrying a 2.2 kb *C. elegans unc-119* rescuing fragment (Merritt *et al.*, 2008) using LR clonase (Invitrogen) according to the manufacturer's instructions. Colonies were selected on ampicillin and recombined plasmids were confirmed by digestion with *Not*I. Plasmid minipreparations of the correct clones were performed to provide enough material for bombardment.

### Transformation of *C. elegans* by Biolistic Bombardment

Instead of the traditional *C. elegans* DNA transformation method by germline microinjection (Mello *et al.*, 1991), we preferred transformation by biolistic bombardment (Praitis *et al.*, 2001). The advantage of this technique over microinjection is that it generates integrated, low-copy number transgenic lines with high frequency, lowering the possibility of transgene silencing in the germline. Germline transgene silencing occurs very frequently with the multi-copy extrachromosomal arrays that assemble when transformation is performed by microinjection.

The bombardment transformation protocol we used is an adaptation of a protocol provided to our laboratory by John Reece-Hoyes then at Ian Hope's laboratory at the University of Leeds, UK. A Bio-Rad PDS-1000/He machine with a hepta adaptor was used. 60 mg of 0.3-3 µm gold beads (Chempur) were prepared by vortexing in 2 ml of 70% ethanol for 5 minutes and then soaked for 15 minutes. After spinning, they were washed 3x with sterile distilled water. The pellet was resuspended in 1 ml of sterile 50% glycerol for a final concentration of 60 mg/ml. Coating of the particles with DNA was done on the day of the bombardment. The beads were vortexed for 20 minutes and, for each hepta shot (seven different spots on a 10 cm plate are hit with the gold microparticles), 7 µg of plasmid DNA in 30 µl of ultrapure water was added dropwise while slowly vortexing (to prevent bead clumping) to 70 µl of gold suspension in a siliconized 1.5 ml tube. Then 300 µl of 2.5 M $CaCl_2$ and 112 µl of 0.1 M spermidine were added in the same manner. This mixture was vortexed for 5 minutes, the gold was pelleted and resuspended in 800 µl of 70% ethanol, pelleted again, and resuspended in 70 µl of 100% ethanol. The mixture was kept vortexing until ready to use.

The strain used for bombardment was the *unc-119(ed3)* mutant which exhibits almost complete paralysis. To grow large numbers of animals, starved populations from several small (3 cm) plates were collected and washed with M9 buffer, pelleted by centrifugation and plated on large (15 cm) plates with concentrated OP50 and grown until most animals were fully gravid. They were then bleached to recover eggs, grown synchronously and expanded until large numbers were reached. For harvesting, 200,000 hatched L1 larvae were plated per 15 cm plate with 3 ml of concentrated OP50. Per bombardment plate, 1 ml of packed young adults (with at most 2-4 eggs) are required (400,000 to 500,000 worms). Immediately before the procedure, worms were washed with M9 buffer and pelleted. Then 150 µl of concentrated worms were pipetted on each of the seven spots to be hit on a 10 cm NGM plate without OP50. The plate was pre-chilled on ice to minimize dispersion of the worms and kept over ice until ready to bombard.

Macrocarrier discs were coated in the center with 10 µl of gold beads in a flow hood and allowed to dry. The macrocarriers were assembled into the hepta adaptor along with 1350 psi rupture disks and a stopping screen to hold the macrocarriers in the adaptor after they were shot (all from Bio-Rad). Inside the chamber the vacuum was kept at a pressure of 27 inches of Hg and the plate was placed on the second shelf from the bottom. Operation was carried out according to the machine's manual. After bombardment, 1 ml of sterile M9 buffer was added to the plate and the worms were allowed recover for 1-2 hours at room temperature. 4 ml of M9 buffer were added and worms were equally distributed into 8x 10 cm plates pre-seeded with OP50. Worms were grown at 20˚C for 3-4 days and four wild-type moving gravid transformants (*unc-119* rescued) were picked from each plate and placed on a drop of bleach solution in a 3 cm plate with OP50 to decontaminate. From each of these plates, 6-8 progeny (F1) animals were singled onto individual plates and transmission of the *unc-119* construct was assessed. Whenever the transmission was 100% in any of the plates, the transgene was assumed to have been integrated and was checked for a few generations. Both integrated and non-integrated lines were observed under the microscope for GFP signal and multiple lines were kept. Integrated lines were crossed with *pir-1(tm3198)/mnC1** males and hermaphrodite progeny not carrying the nuclear $P_{sur-5}$::GFP signal from the balancer were allowed to self-fertilize. Several F2 animals were singled, allowed to produce offspring and subjected to PCR for the wild-type allele, transgene and homozygous deletion allele, using primers O21/O24. Animals in which only the transgene and deletion allele PCR products could be detected, were *pir-1* homozygotes rescued by the transgene.

## Generation of Constructs and Strains for Transformation by MosSCI, and Transformation Protocol

Single copy *pir-1* transgenic strains were generated using the *Mos*1-Mediated Single Copy gene Insertion (MosSCI) system based on the protocols described in Frokjaer-Jensen *et al.*, 2008. Improved protocols and reagents are available at wormbuilder.com. Previously generated pDONR-201-*pir-1::gfp* plasmids, which had rescued the *pir-1* phenotype were recombined with LR clonase into the pCFJ201 destination vector, which carries a *C. briggsae unc-119* rescuing fragment and has homology sites surrounding the *cxTi10882 Mos*1 insertion on chromosome IV to allow insertion of the transgene by homologous recombination following *Mos*1 transposon excision. Correct clones were screened by PCR with primers flanking the *Mos*1 insertion site (O37/038) and plasmids for injections were purified by minipreparation. At the time this was the only other genomic insertion site available besides the preferred chromosome II site, which could not be used due to its proximity to the *pir-1* locus. The chromosome IV site was shown to have a lower efficiency of transformation compared to the chromosome II site, which did not allow insertion of all of the transgenes generated. Since then, a *pir-1* mutant strain carrying the chromosome I *ttTi4348* site was made, which should be used for future transgenic experiments.

The *cxTi10882; unc-119(ed3)* strain (EG5003) was crossed with bombardment-rescued *pir-1(tm3198); unc-119(ed3); In[pir-1::gfp, unc-119];* avr3x in order to place it in the avr3x background. Heterozygotes were allowed to self-fertilize and *unc-119* progeny (indicating loss of the transgene) was singled and checked for homozygous *Mos*1 insertions by PCR with O37/O38. Positive animals were then nearly starved and plated on ivermectin plates to select for resistant fertile animals. These were then crossed to *pir-1(tm3198)/mnC1*;* avr3x, and F1 progeny siblings were crossed using wild-type-moving GFP balancer positive males and balancer negative hermaphrodites to rebalance *pir-1*. From the sibling cross, balanced *unc-119* cross-progeny was PCR tested for presence of the *pir-1* deletion and of the *Mos*1 insertion. The resulting *pir-1/mnC1*; unc-119; cxTi10882;* avr3x strain was used for microinjection of the rescuing constructs.

For germline microinjections, a mix of the following plasmids and respective concentrations were used: pCFJ201-*pir-1::gfp* (50 ng/µl), pCFJ90 ($P_{myo-2}$*::mCherry* which labels the pharynx; 2.5 ng/µl), pCFJ104 ($P_{myo3}$*::mCherry*, which labels body wall muscle; 5 ng/µl); pJL43.1 ($P_{glh-2}$*::transposase*, which constitutively expresses the Mos1 transposase in the germline; 30 ng/µl), and pJL43.1 ($P_{hsp-16.48}$*::transposase*, for heat-shock

induction of transposition; 30 ng/µl). Inclusion of both types of transposase constructs was meant to increase the frequency of insertion. Per construct, 40-60 animals were injected, isolated and allowed to grow at 20˚C for 2-3 days. Wild-type moving transformants were picked and assessed for high transmission. Transmitting lines were grown to a mixed stage population and heat-shocked for 75 minutes at 34˚C, following recovery at 15˚C for 90 minutes and 20˚C for 1 hour. From each line ~400 motile young adults were picked and screened under a dissection fluorescent microscope to eliminate GFP-negative animals (which have lost the *mnC1** balancer). About 25 worms were transferred to a 10 cm plate with OP50 for a total of 16 plates per line and grown at 20˚C for 2-3 days, then transferred to 25˚C until they starved. Then a 1/4 piece of the agar of each plate was chunked into a fresh 10 cm plate with OP50 and allowed to starve at 25˚C. Plates were screened for wild-type moving worms that lost the mCherry red fluorescence and had therefore likely lost the extrachromosomal arrays. Animals were isolated and screened by PCR for the presence of the transgene at the *Mos*1 locus and later by immunoprecipitation (IP) of PIR-1::GFP.

### Site-Directed Mutagenesis of *pir-1* Transgenes for Transformation by MosSCI

For generation of *pir-1* mutant transgenes, instructions from the QuikChange site-directed mutagenesis kit (Stratagene) were followed using either the kit's PfuUltra or iProof (Bio-Rad) DNA polymerase. Mutagenesis was performed on the smaller pDONR201 vectors, after which the *pir-1* gene was sequenced and recombined into pCFJ201 with LR clonase. The deletions in the putative NLS (Δ194-208 and Δ204-208) were generated by divergent PCR primers lacking the targeted nucleotides followed by ligation. The PCR used iProof DNA polymerase for 15 cycles, and the product was digested with *Dpn*I, gel purified and phosphorylated with PNK. Following purification with Qiagen Minelute columns, 200 ng of each vector was ligated in a 10 µl volume with LigaFast T4 DNA ligase (Promega) for 15 minutes at room temperature. As negative controls, the same amount of unligated vectors were transformed in parallel with ligated mixtures (5 µl per reaction). Colonies were screened by PCR with Roche Taq polymerase with primers O21/O22. Primers for PIR-1 mutagenesis (O39-O71) and the respective mutations can be found in Appendix B.

### Preparation of Tissues for Microscopy

For visualization of live animals, washed worms were mounted on slides with a 2% agarose pad with M9 buffer containing 0.4% levamisole to induce paralysis. For live embryos, gravid adults were placed on the agarose pad and cut around the vulva with a fine hypodermic needle. Preparation of tissues for DAPI staining and immunofluorescence were carried out largely according to Phillips *et al.*, 2009. For gonad dissection 40 to 50 L4 to young adult worms were picked from plates and washed extensively with 1X Egg Buffer (25 mM HEPES-NaOH pH 7.4, 118 mM NaCl, 2 mM EDTA, 0.5 mM EGTA, 0.1% Tween-20) to eliminate bacteria. Then the buffer was replaced with Egg Buffer containing 0.4 mM levamisole (15-30 µl) and transferred onto an 18x18 mm coverslip. By cutting the animals with the tip of a fine hypodermic needle at either the head (below the pharynx) or at the tail, gonads, and intestines were released. An equal volume of fixative solution (3.7% formaldehyde in 1X Egg Buffer without Tween-20) was added and pipetted up and down to further extrude and dissociate germline tissue from the rest of the animals. Fixation was allowed to occur for 5 minutes at room temperature. All but about 10 µl of solution were removed from the coverslip. The coverslip was picked up by touching the drop at the center of a positively charged slide (VWR VistaVision HistoBond) leaving a small corner of the coverslip protruding from the edge of the slide. Excess solution was wicked away from the edge of the coverslip using torn strips of absorbent filter paper to promote adherence of the tissues to the slide. The sample was then freeze-cracked by placing it on a pre-cooled aluminum block on dry-ice for at least 10 minutes and quickly flicking the coverslip from the slide using the protruding corner. The slide was immediately dipped in cold (-20˚C) methanol in a Coplin jar for 1 minute, and then transferred to 1X PBS buffer (10 mM Phosphate pH 7.4, 137 mM NaCl, 2.7 mM KCl) containing 0.1% Tween-20 (PBST) at room temperature. For DAPI staining only, the slide was washed in PBST for 10 minutes, followed by another 10-minute wash with PBST containing 0.5 µg/ml DAPI, and a final 30-minute wash in PBST, all at room temperature. Slides were mounted by first removing excess buffer from the slides without letting the sample dry completely and then inverting the slide and touching the sample on a drop of 10 µl of Vectashield mounting medium placed at the center of a 22x22 mm coverslip. Excess medium was removed by pressing the inverted mounted slide on a paper towel, and the edges were sealed with transparent nail polish.

For immunofluorescence, the wash step after methanol was followed by 3x 10-minute washes in PBST, followed by a blocking step with 0.5% BSA in PBST. For this,

100 µl of the solution were added onto the worms and covered with a square Parafilm coverslip to hold the liquid in the sample area. Slides were incubated in a humid chamber at room temperature for at least 30 minutes. The Parafilm slides were removed by dipping the slide in PBST. 100 µl of primary antibody diluted in blocking solution were placed on the sample and covered with a Parafilm coverslip, and incubated in a humid chamber for 2 hours at room temperature or overnight at 4˚C. After 3x 10-minute washes in PBST at room temperature the slide was incubated with secondary antibody as described for the primary, followed by DAPI staining and mounting of the slide as described above.

For enhancement of the PIR-1::GFP signal, worms were incubated with a 1:100 dilution of anti-GFP mouse monoclonal antibody (WAKO) overnight at 4˚C. Secondary antibody incubation was performed for 2 hours at room temperature with a 1:500 dilution of FITC-conjugated donkey anti-mouse (Jackson). For histone H3 a rabbit polyclonal anti-H3 (Cell Signaling) was used at a 1:100 dilution and incubated overnight at 4˚C, followed by a 2-hour, room-temperature incubation with 1:500 TRITC-conjugated anti-rabbit antibody (Jackson). Images were acquired with a Zeiss Axioplan 2 microscope using Zeiss AxioVision software.

# CHAPTER III

Proteomic and Biochemical Studies
of PIR-1 and Interacting Proteins

## INTRODUCTION

In parallel with our characterization of PIR-1 at the cellular and developmental levels, we undertook biochemical approaches to find PIR-1 protein partners *in vivo*. Since DCR-1 integrates multiple complexes to produce distinct sRNA species, one of the main objectives of the project was to understand precisely which Dicer complexes PIR-1 associates with. Our point of departure was the analysis of PIR-1 immunoprecipitates from the rescuing transgenic lines by Multidimensional Protein Identification Technology (MudPIT), the same approach that led to the discovery of PIR-1 as a partner of Dicer. With this method we confirmed the interaction of PIR-1 with DCR-1, and defined a new set of interactors that points to a role in Dicer-dependent sRNA biogenesis. Further, we validated these interactions by IP/western blotting across development. Unexpectedly, we found that PIR-1 is expressed as two isoforms with intriguing differences concerning the proteins they interact with, subcellular distribution, and developmental expression. These findings constitute an important groundwork from which to dissect the apparently wide range of PIR-1 functions.

# RESULTS

## MudPIT of PIR-1 Immunoprecipitates Reveals Interactions with Proteins of the ERI Complex

MudPIT is a highly sensitive technique for large-scale identification of proteins in highly complex mixtures (Wasburn *et al.*, 2001; Wolters *et al.*, 2001; reviewed in Yates *et al.*, 2009). It combines fractionation of a peptide digest of the protein mixture by multidimensional microscale liquid chromatography, electrospray ionization, and tandem mass spectrometry. The origin of each peptide is then computationally determined by comparison to a protein database. In preparation for MudPIT, we performed silver-staining of PIR-1::3xFlag IPs, as the clear presence of specific bands in silver staining is a requisite for generating good-quality data through MudPIT. Three prominent bands that were present only in the transgenic rescued IP were sent for mass spectrometry (Fig. 3.1A). Two of them were identified as DCR-1 and the helicase DRH-3. The other band could not be identified due to contaminants, but was later confirmed to be the Eri RdRP RRF-3, in agreement with the molecular weight of the band on the gel. The IP material used for the first MudPIT experiment was tested by western blotting with antibodies available in the laboratory (Fig. 3.1B). This confirmed DCR-1 and DRH-3 to interact specifically with both PIR-1::3xFlag and PIR-1::GFP, and showed that ERI-1b, but not the shorter isoform ERI-1a, also interacts with PIR-1. ERI-1a has been shown to have a role in cytoplasmic 3'-end processing of 5.8S rRNA which is distinct from the role of ERI-1b in 26G-RNA biogenesis (Gabel and Ruvkun, 2008).

The IPs used for MudPIT were performed on a 1:1 mixture of extract from gravid adult transgenic animals fed normal bacteria and extract from gravid adult wild-type N2 animals that were fed bacteria grown on $^{15}$N-containing (heavy) growth medium. This allowed us to discern specific from unspecific interactions that occur with the affinity matrix, through analysis of heavy-to-light peptide ratios obtained for each protein. In other words, when only light peptides (from the transgenic extracts) for a given protein are recovered, it means that the interaction is PIR-1-specific. As an additional criterion for specificity we considered proteins that were detected in both PIR-1::GFP and PIR-1::3xFlag IP experiments. Table 3.1 summarizes the results of this experiment, revealing that most of the proteins in common to both IPs are components of the ERI complex, namely DCR-1, DRH-3, RRF-3, RDE-4, and ERI-3. They rank among the highest abundance proteins recovered, as judged by the obtained spectral counts, the measurement

that most strongly correlates with relative protein abundance (Liu *et al.*, 2004). As the IP of PIR-1::GFP was generally cleaner and free from immunoglobulin contaminants, we obtained higher spectral counts and detected additional peptides corresponding to ERI-1 and ERI-5, not found in the PIR-1::3xFlag IP. Additionally, ERI-9 was detected in both IPs, albeit at a very low abundance. All of these proteins have previously been defined as components of the ERI complex (refer to Chapter I, page 43; Duchaine *et al.*, 2006; Thivierge *et al.*, 2012). To confirm these results, an independent PIR-1::GFP IP/MudPIT experiment was performed without a heavy-isotope control (Table 3.2). Again, we obtained high spectral counts for DCR-1, DRH-3, RRF-3, RDE-4, and ERI-3, and could still detect some peptides from ERI-1, ERI-5, and ERI-9.

We then proceeded to verify some of these interactions by western blotting, as well as to test other important RNAi factors for which antibodies were available. In Figure 3.2A and B, IP of PIR-1::3xFlag robustly pulled down DCR-1, DRH-3, RRF-3 and ERI-1b (the RRF-3 antiserum cannot detect the protein in inputs and it normally detects an unspecific band above RRF-3). Surprisingly, we could also easily detect RDE-8 despite its very low abundance in the second PIR-1::GFP IP MudPIT run. This Zc3h12a-like endoribonuclease was discovered in our laboratory and shown to be required for exo- and endo-RNAi (Tsai *et al.*, 2014). This constituted the first evidence that RDE-8 may integrate the fully assembled ERI complex. Other proteins tested included the essential Argonaute CSR-1 (also detected at low abundance in the same MudPIT experiment as RDE-8), the 22G-RNA RdRPs RRF-1 and EGO-1, and the 3'-5' exonuclease required for 22G-RNA accumulation MUT-7, none of which could be detected. When analyzing the post-IP supernatants we could not detect a depletion of the proteins recovered in the PIR-1 IP, despite an almost complete depletion of PIR-1::3xFlag (Fig. 3.2A and B). This suggests that PIR-1 stably interacts with only a small fraction of the total available pool of these proteins, consistent with the fact that DCR-1 and DRH-3 are known components of other complexes. Interestingly, the IPs reveal that PIR-1 may exist as two isoforms migrating very closely within ~2-4 kDa, regardless of the tag present on its C-terminus (compare Fig. 3.2B and C). We named the isoforms PIR-1a (lower, lighter band) and PIR-1b (upper, heavier band).

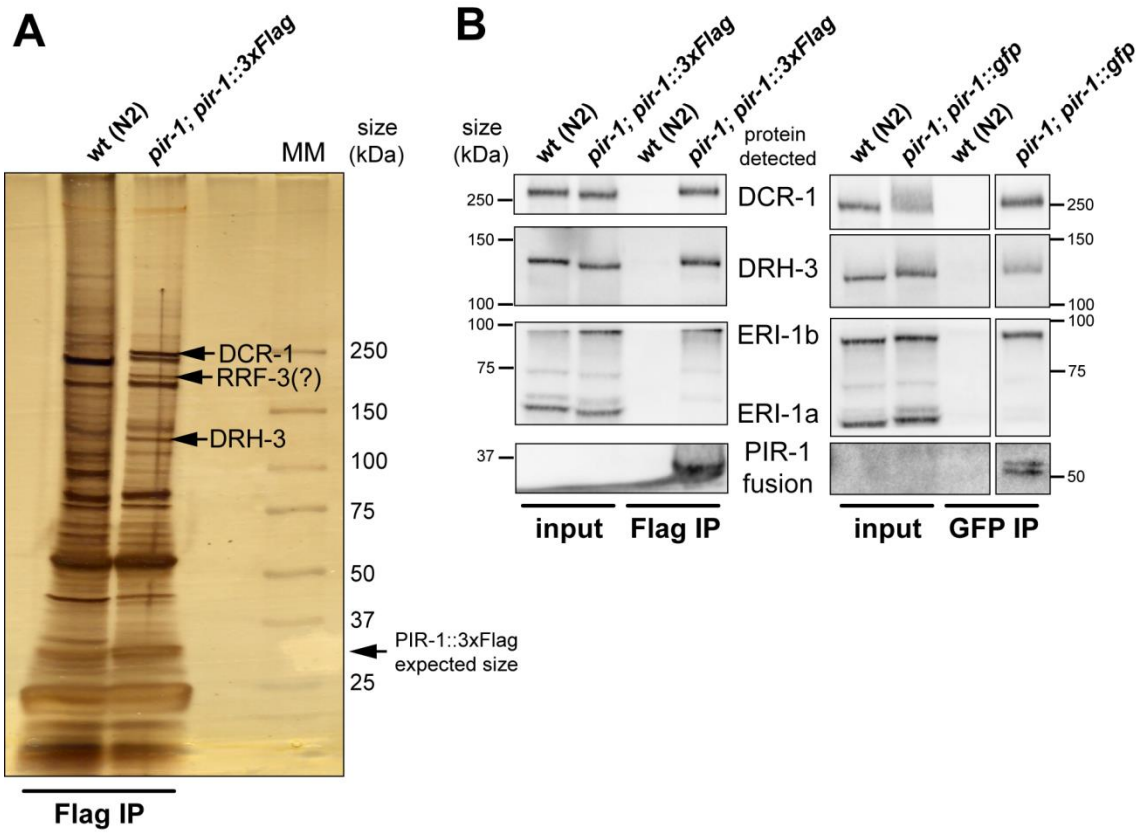



**Figure 3.1. Validation of PIR-1 immunoprecipitations before MudPIT. (A)** Silver staining of a PIR-1::3xFlag IP. Arrows point to bands found exclusively in the IP from the transgenic animals but not in the wild-type negative control. The bands were excised and sent for identification by mass spectrometry. The middle band was not identified due to contaminants, but migrated at the predicted molecular weight for RRF-3. **(B)** Western blotting of IPs from mixed extracts of rescued transgenic animals and $^{15}$N-labeled wild-type animals used for MudPIT analysis. 'MM' is the molecular weight marker.

**Table 3.1.** List of Specific Interactors Obtained by MudPIT of PIR-1::GFP and PIR::3xFlag Immunoprecipitates from a Mixture of Unlabeled Transgenic with $^{15}$N-labeled Wild-Type Protein Extracts from Gravid Adults

| Protein | Length in aa/ MW in kDa | Spectral Count[a] (GFP/Flag) | Peptide Count (GFP/Flag) | Protein Coverage (GFP/Flag) |
|---|---|---|---|---|
| **High Confidence PIR-1 Interactors[b]** | | | | |
| PIR-1[c] | 233 / 27.1 | 202 / 12 | 29 / 7 | 42.9% / 22.3% |
| DCR-1 | 1845 / 210.9 | 137 / 27 | 50 / 12 | 31.2% / 9.0% |
| RRF-3 | 1780 / 203.1 | 68 / 15 | 27 / 9 | 18.0% / 6.1% |
| DRH-3 | 1119 / 129.1 | 59 / 19 | 24 / 7 | 23.1% / 7.1% |
| ERI-5 | 531 / 61.6 | 24 / 0 | 11 / 0 | 22.0% / 0% |
| RDE-4 | 385 / 43.4 | 13 / 8 | 8 / 3 | 28.6% / 15.1% |
| ERI-3 | 578 / 66.4 | 16 / 2 | 7 / 2 | 10.7% / 5.7% |
| **Additional Interactors Previously Found in Association with DCR-1[d]** | | | | |
| ALG-2 | 910 / 101.6 | 4 / 3 | 2 / 2 | 3.7% / 5.2% |
| ERI-9 | 635 / 73.5 | 2 / 3 | 2 / 2 | 5.4% / 6.5% |
| ERI-1 | 582 / 67.2 | 7 / 0 | 3 / 0 | 12.7% / 0% |
| ALG-1 | 1002 / 110.9 | 3 / 0 | 2 / 0 | 3.3% / 0% |
| B0001.2 | 926 / 105.1 | 0 / 3 | 0 / 2 | 0% / 2.7% |

[a] Spectral counts represent the total number of tandem mass spectra that match peptides to each protein and is the best estimate of the protein abundance in a mixture. [b] High confidence interactions were defined as proteins that had >10 spectral counts in either IP and for which no heavy $^{15}$N peptides were recovered. [c] The data refer to the fusion proteins PIR-1::GFP (476 aa / 54.3 kDa), and PIR-1::3xFlag (263 aa / 30.5 kDa). [d] Interactors that did not meet the specified spectral count cutoff but that have been found to interact with DCR-1 or ERI-1 in previous studies (Duchaine *et al.*, 2006; Thivierge *et al.*, 2012).

**Table 3.2.** List of Interactors Obtained by MudPIT of PIR-1::GFP Immunoprecipitates from Young Adult Extracts

| Protein | Length in aa/ MW in kDa | Spectral Count (abundance) | Peptide Count | Protein Coverage |
|---|---|---|---|---|
| **ERI Complex Proteins** | | | | |
| PIR-1::GFP | 476 / 54.3 | 142 | 29 | 35.9% |
| DCR-1 | 1845 / 210.9 | 118 | 41 | 22.8% |
| RRF-3 | 1780 / 203.1 | 54 | 28 | 17.9% |
| DRH-3 | 1119 / 129.1 | 50 | 25 | 24.8% |
| ERI-3 | 578 / 66.4 | 23 | 10 | 14.9% |
| RDE-4 | 385 / 43.4 | 15 | 6 | 17.1% |
| ERI-1 | 582 / 67.2 | 11 | 6 | 11.3% |
| ERI-5 | 531 / 61.6 | 9 | 4 | 8.7% |
| ERI-9 | 635 / 73.5 | 2 | 2 | 3.8% |
| **Proteins with Roles in Exo- and Endo-RNAi Pathways** | | | | |
| RDE-8 (ZC477.5) | 339 / 38.9 | 5 | 2 | 6.8% |
| **Other Dicer-Associated Proteins** | | | | |
| B0001.2 | 926 / 105.1 | 8 | 6 | 8.6% |

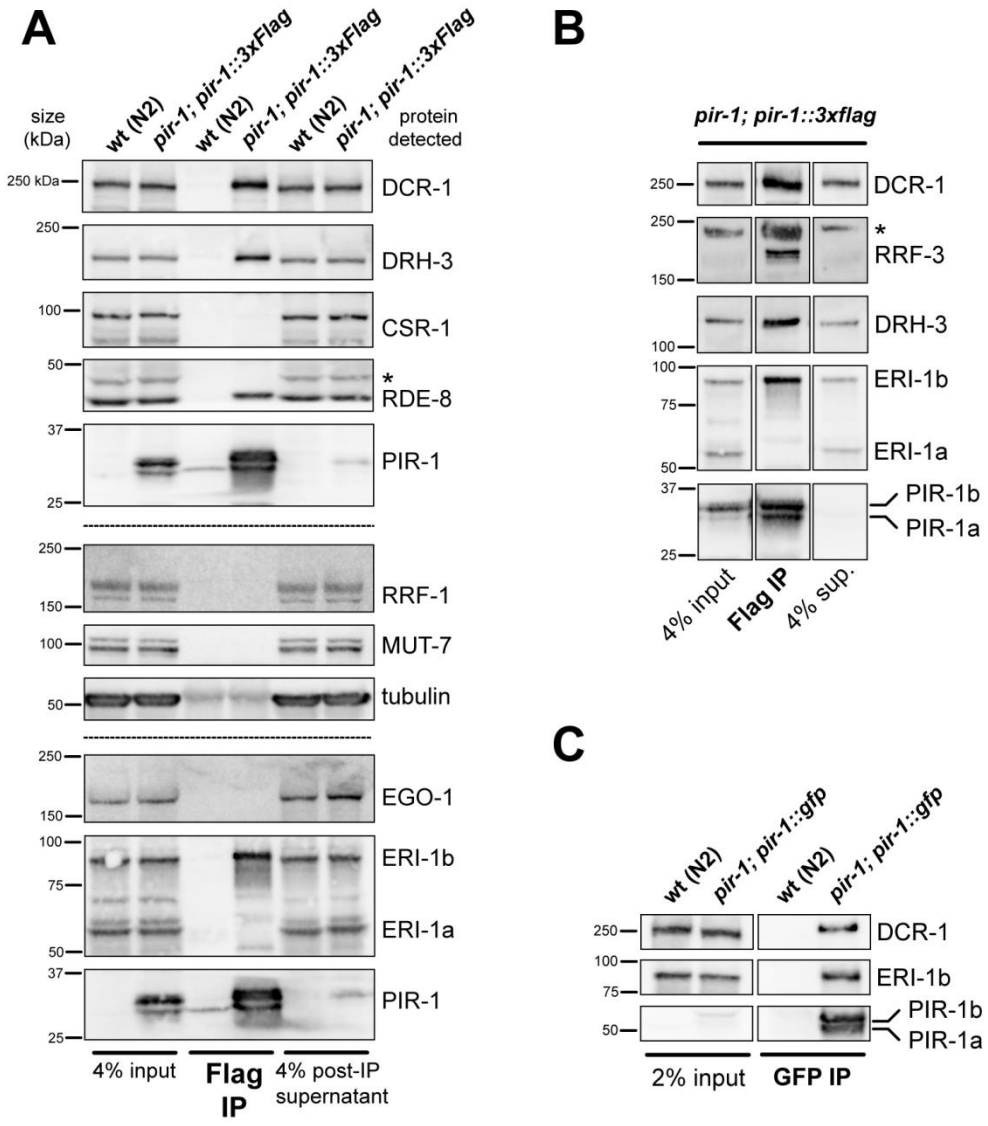



**Figure 3.2. Immunoprecipitation/western analyses confirm interactions with ERI complex proteins and demonstrate that PIR-1 is expressed as two distinct isoforms. (A)** PIR-1 interacts with DCR-1, DRH-3 and ERI-1b but not with other sRNA proteins not found in MudPIT. RDE-8 is abundantly pulled down by PIR-1 in spite of having been detected in only one MudPIT run at very low abundance. Analysis of supernatants after IP shows depletion of PIR-1::3xFlag but not of its partners. **(B)** Western blot confirmation of the interaction with RRF-3, which can only be detected upon IP enrichment. Two different PIR-1 isoforms are clearly enriched after IP. **(C)** The two isoforms are maintained when the PIR-1 is C-terminally tagged with GFP. PIR-1::GFP was normally difficult to detect in inputs, perhaps due to lower sensitivity of the antibody used or lower expression compared to PIR-1::3xFlag. Dashed lines separate the same samples run on separate lanes or gels. The asterisks signal unspecific bands detected by the antibody used.

## PIR-1 Co-Fractionates with the Fully Assembled ERI Complex and Maintains Most Interactions throughout Development

In order to further characterize the association of PIR-1 with ERI-complex proteins, we performed gel filtration followed by western analysis of extracts from young gravid adult *pir-1::3xFlag* rescued animals (Fig. 3.3A). The fully assembled ERI complex is estimated to be ~850 kDa (Thivierge *et al.*, 2012), and consistent with PIR-1 assembling with the complex it is abundantly detected in high molecular weight fractions (to the left of the 440 kDa marker). Curiously, PIR-1b strictly overlaps with DCR-1, becoming undetectable in lower MW fractions where DCR-1 can also no longer be detected. In these lower molecular fractions, PIR-1a becomes enriched and still overlaps with DRH-3, ERI-1b and RDE-8. This experiment strongly supports PIR-1b as the predominant form associating with the core ERI complex but it also points to the existence of lower MW complexes integrating primarily PIR-1a. This uneven distribution could reflect the involvement of PIR-1 in distinct processes.

During the optimization of IP and washing conditions for PIR-1 CLIP-seq, we made some observations regarding the properties of PIR-1 interactions. The inclusion of an ultracentrifugation step at ~130,000x *g* for one hour, for instance, did not deplete the extracts of PIR-1 and some of its partners, suggesting that the majority of PIR-1 molecules are not stably associated with very high MW complexes such as ribosomes, spliceosomes or other large RNP complexes (Fig. 3.3B). We increased the stringency of post-IP washes by first varying the amount of salt and then the detergent concentration and composition from the standard lysis and IP buffer (25mM HEPES-KOH pH7.5, 150 mM KCl and 0.5% NP-40). The inclusion of a five-minute washing step at 4ºC with different salt concentrations demonstrated that the tested interactions are stable in up to 1 M KCl, beyond which they begin to weaken (Fig. 3.3C). Using 1 M KCl, we tested different concentrations of non-ionic (NP-40) and anionic (SDS and sodium deoxycholate) detergents. When we included sodium deoxycholate, in addition to NP-40 and SDS, the interactions with DRH-3 and ERI-1b were partially lost, while the interaction with DCR-1 remained as strong as with NP-40 alone (Fig. 3.3D, conditions 5 and 6). While this is far from being a systematic study of the stability of the complex, it does suggest that the interaction of PIR-1 with DCR-1 is particularly stable and helps explain why PIR-1 was so abundantly detected in DCR-1 proteomics, but less so in proteomics of ERI-1 and ERI-5 (Duchaine *et al.*, 2006; Thivierge *et al.*, 2012). Lastly, incubating PIR-1 immunoprecipitates with RNases followed by washes did not disrupt the association with

DCR-1 or DRH-3 (Fig. 3.3E). Despite not entirely conclusive, together with the high tolerance of the interactions to harsh salt and detergent conditions, this result argues that PIR-1 complexes likely do not require RNA to assemble.

In agreement with the expression pattern of PIR-1::GFP described in Chapter II, we could detect PIR-1::3xFlag at every developmental stage by western blotting (Fig. 3.4A). PIR-1 is expressed at a slightly higher level in embryos than in postembryonic stages, during which expression is relatively constant. DCR-1, DRH-3 and ERI-1b are highly expressed in embryos and L1 larvae compared to L2 and L3 larvae. During the L4 stage, at which the germline greatly expands and spermatogenesis takes place, the expression of the three proteins is again increased, only to suffer another slight decrease during adulthood. PIR-1 IPs at every developmental stage showed that both PIR-1 isoforms are recovered at all post-embryonic stages. Interestingly, only one form is expressed in embryos. An independent PIR-1::GFP IP clearly shows that the enriched form in embryos is PIR-1b (Fig. 3.4B), which we have shown in the previous section to preferentially co-fractionate with Dicer at high MW. At this developmental stage, the highest amount of all tested interactors is pulled down, including RRF-3, suggesting that the activity of a PIR-1-containing ERI complex is particularly important in the developing embryo. During the rest of development, interactions are maintained throughout but are stronger in L1s, reaching a low during L3. They then start gradually increasing at L4 to coincide with intense germline development and, later, with embryogenesis in gravid adults. Interactions with RRF-3 and RDE-8 are also detected in both young and older gravid adults (Fig. 3.4C).

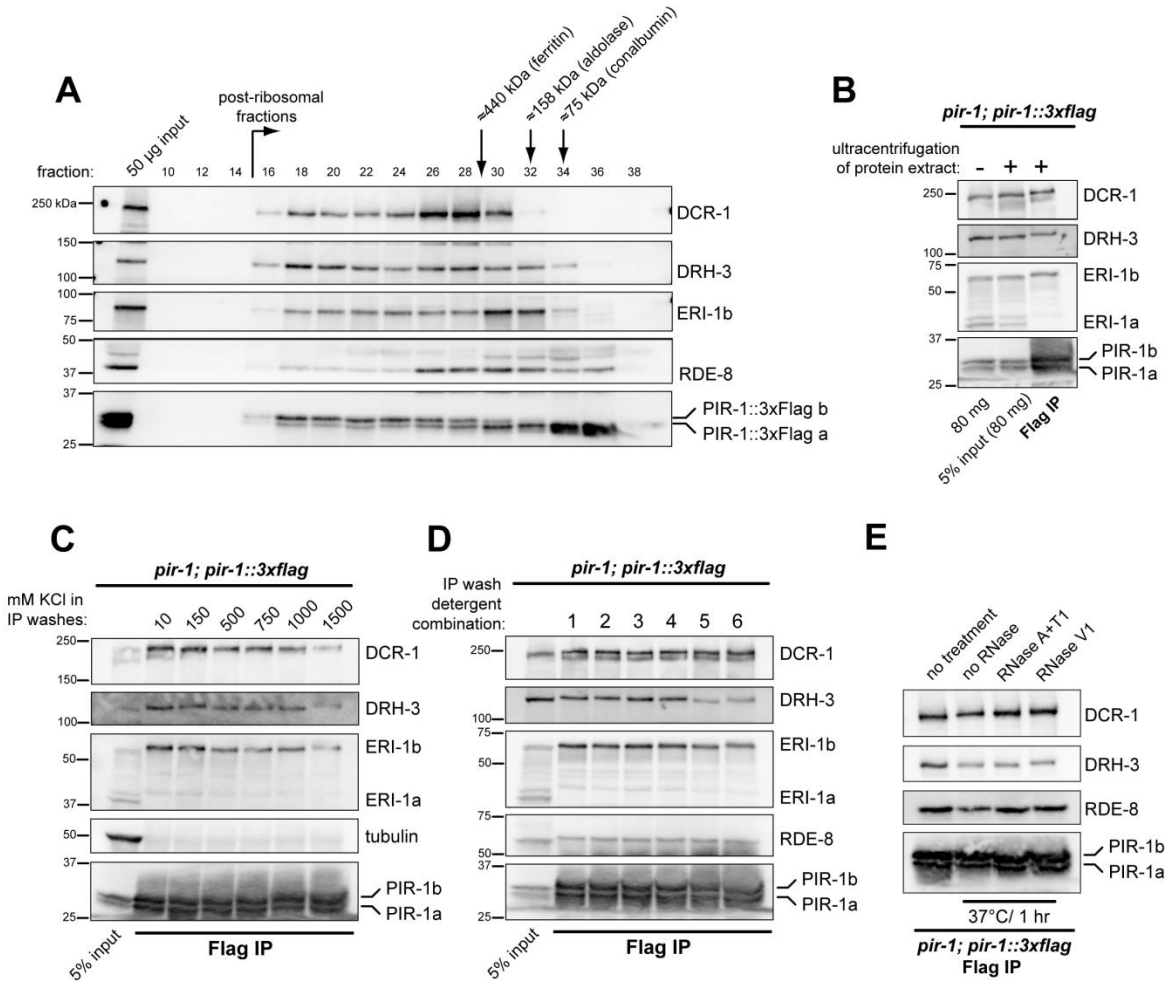



**Figure 3.3. PIR-1 co-fractionates with the fully assembled ERI complex and interacts very stably with Dicer and other proteins of the complex. (A)** Gel filtration of PIR-1 immunoprecipitates from young gravid adults followed by western blotting. PIR-1a is found in a very broad range of molecular weights, but PIR-1b tends to reside in MW fractions where the ERI complex is also found. **(B)** PIR-1a, PIR-1b and interactors remain in extracts subjected to ultracentrifugation, precluding stable and abundant association with very high MW complexes. **(C)** PIR-1 interactions are resistant to low and high salt washes after IP. **(D)** Under high salt conditions, the PIR-1/DCR-1 interaction is resistant to washing with high concentrations of anionic and ionic detergents. The buffer is 1M KCl, HEPES-KOH pH 7.5 with the following combinations of detergents in each lane: (1) 0.5% NP-40; (2) 0.5% NP-40, 0.1% SDS; (3) 1% NP-40; (4) 1% NP-40, 0.1% SDS (5) 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate; (6) 1% NP-40, 0.1% SDS, 1% sodium deoxycholate. **(E)** PIR-1 interactions are maintained after treatment with RNases specific for single-stranded RNA (A and T1) and for double-stranded RNA (V1).

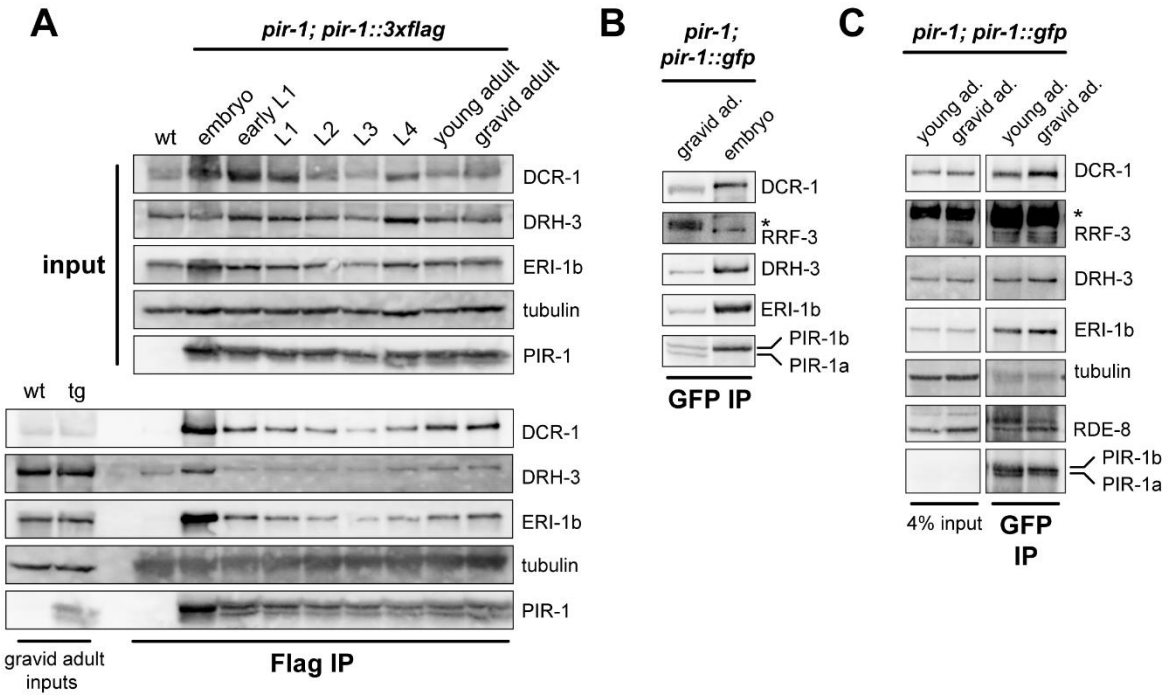



**Figure 3.4. PIR-1 is expressed throughout development as two isoforms, except in embryos. (A)** IP and western blotting of PIR-1::3xFlag at various developmental stages. 'wt' (wild-type) is N2; 'tg' (transgenic strain) is *pir-1; pir-1::3xflag*. The signal in the IP lanes detected with the tubulin antibody coincide with the IgG heavy chain from the antibody used for IP, and therefore reflect non-specific binding of the secondary antibody. **(B)** IP of PIR-1::GFP from the MosSCI single-copy rescued animals to show that only PIR-1b is expressed in embryos and that the RRF-3 interaction is maintained. **(C)** IP of PIR-1::GFP from an independent rescued integrated bombardment line (not the bombardment line used for MudPIT). Non-gravid young adults and gravid adults maintain all interactions, including RRF-3 and RDE-8. Interactions are stronger in gravid adults, likely due to embryos where higher amounts of interactors co-immunoprecipitate with PIR-1 (in all panels). The asterisk represents an unspecific band detected by the RRF-3 antiserum.

## PIR-1 Isoforms Likely Arise from Post-Translational Modifications

Given the intriguing differential expression and interaction pattern of PIR-1a and PIR-1b we sought to understand how these forms originate, with the ultimate goal of asking whether they are functionally distinct. In *C. elegans* about 70% of mRNAs begin with a 22-nt spliced leader (SL) donated by a 100-nt precursor RNA. The addition of this sequence, which has a trimethyl guanosine (TMG) cap at its 5' end, is spliced to the first exon to replace an "outron" sequence with a canonical monomethyl guanosine cap. This process is termed trans-splicing. There are a variety of SL sequences, but the predominant are SL1 and SL2. SL1 typically gets added to the mRNAs of single genes or to the first mRNA of polycistronic RNAs transcribed from multi-gene operons (~15% of *C. elegans*

genes are clustered in operons). SL2 gets spliced exclusively to pre-mRNAs that follow the first gene of an operon (reviewed in Blumenthal, 2012). We designed primers for SL1, SL2, and for a position ~180 nt downstream of the predicted *pir-1* and *sec-5* mRNA start codons. PCR products of a different size would indicate that trans-splicing occurred at a site other than the predicted start site for each mRNA. *sec-5*, being the first gene of an operon, served as our SL1 positive control. RT-PCR followed by TOPO cloning and sequencing demonstrated that *sec-5* is indeed SL1 trans-spliced at the predicted site, whereas *pir-1* RT-PCR gave rise to a product shorter than the expected 180 nt (Fig. 3.5A). We found that the SL1 got spliced seven nt upstream of a second ATG, located 83 nt downstream of the originally predicted ATG. This result was later confirmed by RNA-seq in our laboratory (Fig. 3.5B). The *pir-1* gene depicted is the old model predicted to encode a 261 aa protein. Cap-seq, a method developed in our laboratory that primarily clones the 5' ends of capped RNAs (Gu *et al.*, 2012), confirmed that *pir-1* mRNA is SL1 trans-spliced right before the second start codon. mRNA-seq is in agreement with the trans-spliced mRNA being the prevalent form. Additionally, data from the laboratory of Andrew Fire at Stanford University independently confirmed the size and exon usage by mRNA-seq, as well as the mRNA segments that are protected by ribosomes (*i.e.*, undergoing active translation) by ribo-seq (Stadler *et al.*, 2012). Finally, when we assembled all the PIR-1 peptides detected in our MudPIT experiments, we never obtained sequence coverage upstream of the second ATG (not shown). Together, these results indicate that the maximum size of PIR-1 is 233 aa (and at least 27.1 kDa) in developmental stages where both PIR-1 isoforms are expressed.

In order to complement the aforementioned experiments, we individually mutagenized the three potential start sites in PIR-1::GFP to assess changes in rescuing ability and isoform expression (Fig. 3.6A and B). Abrogating ATG1 (in the original gene model) did not affect the transgene's ability to rescue the *pir-1* mutant phenotype and expression of the two isoforms. Mutagenizing ATG2, the start codon found immediately downstream of the SL1 sequence, led to no expression and no rescue. Interfering with ATG3 did not affect rescue and both isoforms were still expressed. Both ATG1 and ATG3 mutations did not interfere with PIR-1 localization in the nuclei of live animals (Fig. 3.6C). These results define ATG2 as the sole PIR-1 start site, implying that the size difference between PIR-1a and PIR-1b is not due to alternative start sites. We examined the potential for protein variants generated by alternative splicing, including exon skipping and intron inclusion, but they either gave rise to ORFs of different sequences with premature stop

codons or to proteins too small to explain the small size difference between the isoforms. On SDS-polyacrylamide gels, the PIR-1::GFP protein runs just above 50 kDa, in line with the 54 kDa size predicted from translation initiating at ATG2. It is therefore possible that the isoforms arise from post-translational modifications such as N-terminal cleavage or addition of small modifier molecules. Since PIR-1 is predicted to have at least nine phosphorylation sites (five at serine residues, three at threonine residues, and one at a tyrosine residue; predicted using the ScanProsite tool from Expasy.org), we digested immunoprecipitated PIR-1::3xFlag with calf intestinal phosphatase (CIP) or with Lambda protein phosphatase (λ PP). While CIP can remove phosphates from most residues, it prefers phosphotyrosine residues, whereas λ PP is a dual specificity protein phosphatase with high activity not only towards phosphorylated tyrosines, but also phosphoserines and phosphothreonines. Upon treatment with λ PP, but not CIP, the bands lost sharpness but did not collapse into one lower band, as one would expect if the only modification was phosphorylation and digestion was complete (Fig. 3.6D). This result suggests that PIR-1 is phosphorylated, but that it may not be the only post-translational modification it is subjected to.

In order to directly identify post-translational modifications that would justify the existence of two isoforms, we performed MudPIT analysis of PIR-1 in the *drh-3* mutant background. As detailed in the next section, when DRH-3 is not expressed, only PIR-1b is stable. We therefore took advantage of this fact to maximize peptide coverage of the heavier PIR-1 isoform. We intended to use the results to individually substitute every modifiable residue and find how each modification affects PIR-1 isoform composition, interactions it retains with protein partners, and its ability to rescue the mutant phenotypes. We were only able, however, to cover about 50% of the protein, precluding us from undertaking a systematic analysis of its modifications. We were able to confirm that peptides for which phosphorylations were predicted were indeed phosphorylated *in vivo*. This was observed for seven of the nine predicted phosphorylations at positions T49, S54, T90, T92, Y164, S196 and S210. Peptides containing the remaining two predicted phosphoserines were not recovered. Other predicted modification sites included an amidation site (residues 21-24), two N-myristoylation sites (residues 147-152 and 153-158), and three N-glycosylation sites (155, 170 and 230). With the exception of residue N230, none of the other residues were covered. Interestingly, the peptide with N230 contained a modification with a measured mass of 359 Daltons, coinciding with the sum of two mannose molecules. Thus the asparagine at position 230 of PIR-1 is confirmed to be

N-glycosylated *in vivo,* supporting the idea that the two isoforms result from differential post-translational modifications. Lastly, we considered other modifications such as ubiquitination and sumoylation, despite the fact that they usually add significantly more mass to the modified protein than the size difference we observe between PIR-1a and PIR-1b (typically ~8.5 kDa for ubiquitin and ~12 kDa for SUMO). Nonetheless, we silenced key components of both pathways by RNAi and examined the effect on the size of the isoforms by IP/western blotting. In all cases both PIR-1a and b exhibited no differences relative to mock-RNAi treated animals (not shown).

As mentioned in Chapter II, we analyzed immunoprecipitates of non-rescuing catalytically inactive PIR-1::GFP by western blotting. Intriguingly, we saw that in this strain we could only recover the highest MW isoform PIR-1b (Fig. 3.6E, upper and lower blots). The interaction with DCR-1 and the nuclear localization did not change in arrested animals, indicating that the lack of rescue is solely due to the loss of catalytic activity of PIR-1 (Fig. 3.6F). When we compared two other integrated rescuing transgenic strains carrying mutations that do not affect the ability of PIR-1 to rescue, they also only expressed PIR-1b (Fig. 3.6E, upper panel), and their association with DCR-1 was also not perturbed. This result bears the significant implication that expression of PIR-1b is sufficient to rescue the *pir-1* mutant phenotypes. We speculate that the disappearance of PIR-1a upon mutagenesis could be due to slight structural changes that either render it inherently unstable or cause it to lose stabilizing interactions with crucial protein partners.

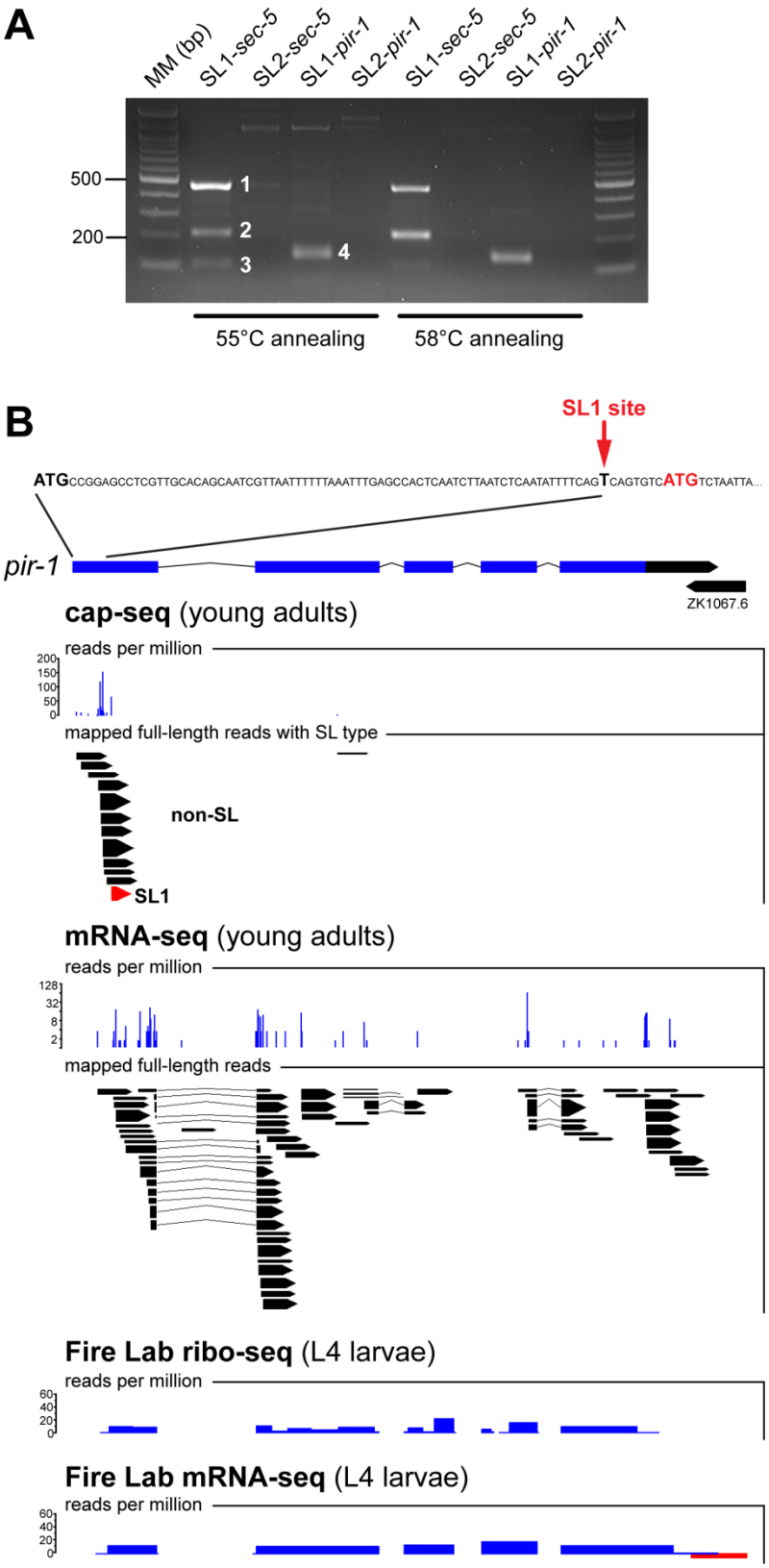



**Figure 3.5. *pir-1* mRNA is trans-spliced to SL1 at its 5' end, giving rise to an open reading frame of 233 amino acids. (A)** RT-PCR of a 5' segment of the *sec-5* and *pir-1* mRNAs with SL1 or SL2 forward primers. Bands 1-4 were purified, TOPO-cloned and sequenced. Bands 1 and 3 were unspecific products, band 2 corresponded to the predicted SL1-*sec-5* size, and band 4

corresponded to a shorter SL1-*pir-1* sequence excluding the start codon predicted in the original gene model. **(B)** Deep-sequencing of PIR-1 mRNA by different methods. The depicted gene model gives rise to a 261 aa protein. In blue are exons separated by introns, and in black is the predicted 5'-UTR. Sequencing of capped RNAs (cap-seq) demonstrates that most reads map to a position downstream of the predicted ATG. The normalized number of reads is represented as blue bars that map to the first nucleotide of each read. Below, full-length reads are mapped, with the thickness of the arrows proportional to the amount of each unique read. Black arrows represent non-SL capped reads, and the red arrow indicates that the sequence is preceded by SL1. The exact site where the SL1 sequence is fused is mapped above the gene diagram. mRNA-seq captured sequences from all exons and the predicted 5' UTR. The Fire laboratory experiments show that ribosomes start translation after the SL1 site and cover all predicted exons, but not the 5' UTR, as expected. mRNA-seq correlates well with our experiment, but defines exons more clearly. Sense reads are colored in blue and antisense are red.

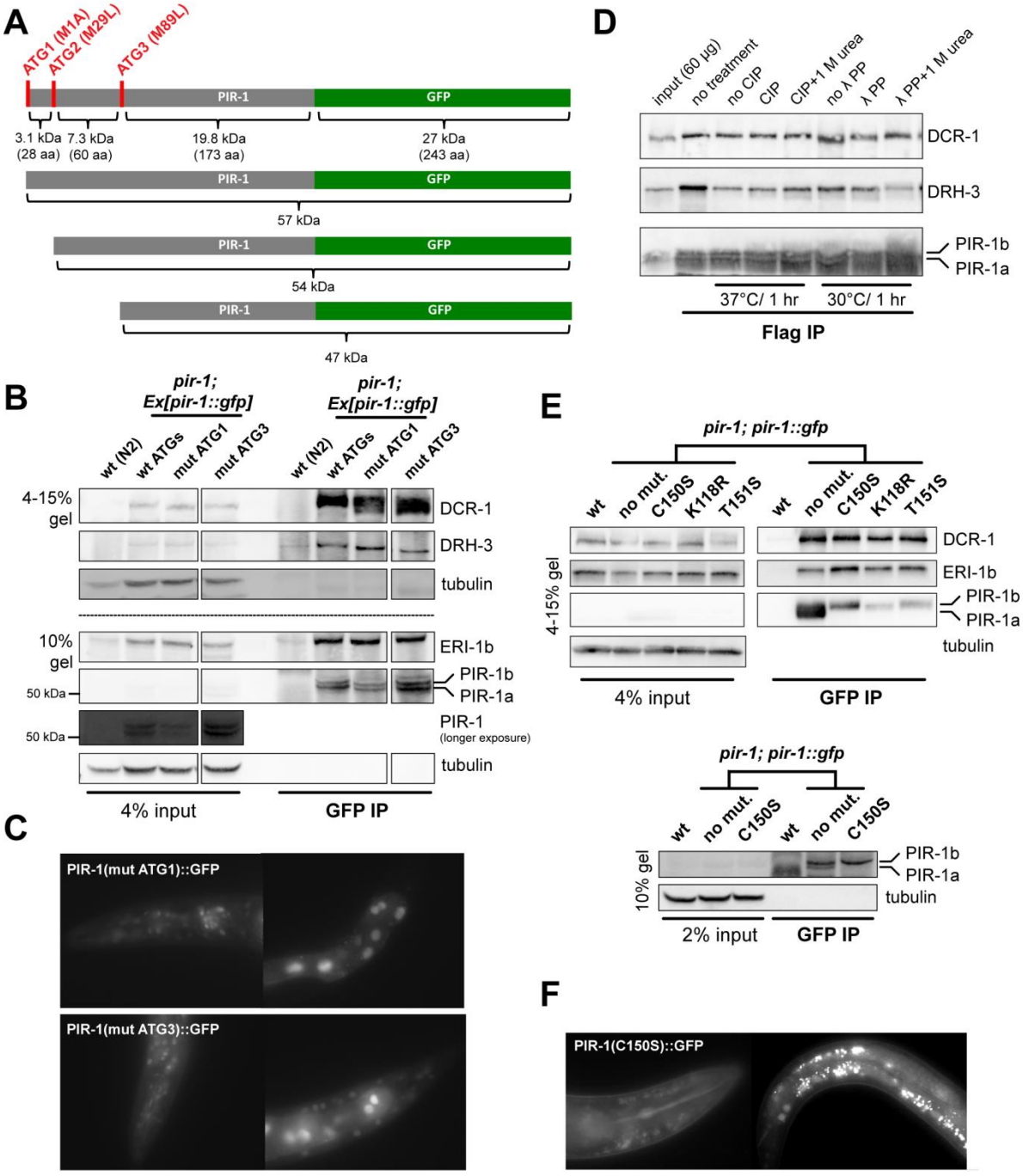



**Figure 3.6. PIR-1 isoforms do not arise from alternative start sites and only PIR-1b is expressed when certain amino acids are substituted. (A)** Diagram of PIR-1::GFP with potential start sites and mutations used to annul them (in red). Sizes are given for each of the three potential forms, including the original 261 aa version. **(B)** IP from extracts of non-integrated rescued transgenic young adults for PIR-1(mut ATG1)::GFP and PIR-1(mut ATG3)::GFP. "Ex" stands for extrachromosomal arrays. Transgenic lines with the PIR-1(mut ATG2)::GFP did not rescue the mutant phenotypes or expressed the protein. The dashed line separates two different gels with the same samples. **(C)** Microscopy of rescued PIR-1 ATG mutants shows unaltered nuclear PIR-1::GFP expression. Head neurons can be seen on the left images and intestinal nuclei are prominent on the right. **(D)** Phosphatase treatments of PIR-1::GFP immunoprecipitates with CIP (Calf Intestinal Phosphatase) or λ PP (Lambda Protein Phosphatase). 1 M urea was added to create slightly denaturing conditions that could facilitate digestion of phosphorylated residues. **(E)** IP of

PIR-1::GFP from integrated single-copy lines with three different mutant sequences (refer to Table 2.1, page 89), including the catalytically inactive C150S, which was counter-selected on ivermectin for seven days. The lower panel is an independent experiment with improved resolution to confirm that only the PIR-1b isoform remains in the phosphatase mutant. **(F)** Localization of PIR-1(C150S)::GFP mutant protein in arrested animals. The left image shows mostly neuronal nuclei and the right shows the germline and intestinal nuclei (the brightest spots are non-GFP intestinal granules).

## Dicer Stabilizes and Preferentially Interacts with the PIR-1b Isoform

Given the distribution differences of PIR-1a and PIR-1b on the gel filtration experiment in Figure 3.3 (page 123), we wanted to better understand the contribution of each form to the interactions we uncovered, particularly with DCR-1. Since we did not have enough antibody to IP endogenous DCR-1, we created a *pir-1; dcr-1* deletion mutant strain rescued with integrated *pir-1::gfp* and non-integrated *3xflag::dcr-1* transgenes (from Welker *et al.*, 2010). While IP of PIR-1 pulled down PIR-1a and PIR-1b, IP of DCR-1 pulled down only one PIR-1 isoform (Fig. 3.7A, top panel). This form appears to be PIR-1b because when we performed a PIR-1 IP on an extract previously used for DCR-1 IP, we observed a decrease in the intensity of the top PIR-1b band relative to PIR-1a (Fig. 3.7A top and lower panels). In contrast, IP of DRH-3 recovered both PIR-1 isoforms (Fig. 3.7B). These results are consistent with DCR-1 co-fractionating predominately with PIR-1b, and with DRH-3 co-fractionating with both isoforms.

We next asked how disruption of the ERI complex through genetic deletion of some of its components would affect interactions with PIR-1. We crossed the *pir-1::gfp* transgene and *pir-1* deletion with *dcr-1*, *drh-3*, *eri-1* and *rde-4* null mutants. The animals carrying *dcr-1* and *drh-3* deletions were genetically balanced, as homozygotes develop into sterile adults. In order to obtain pure homozygous populations, we counter-selected animals for seven days on ivermectin, as described for *pir-1* mutants in Chapter II. IP of PIR-1 from *rde-4* animals led to no changes in interactions or changes in PIR-1 isoforms when compared to IP from a wild-type background (Fig. 3.7C). When ERI-1 was not expressed, interactions with DCR-1 and DRH-3 were maintained and both PIR-1 isoforms were present (Fig. 3.7D). In the absence of DRH-3 the association with both RRF-3 and RDE-8 was lost, and DCR-1 and ERI-1b interactions were weakened (Fig. 3.7C and D). Interestingly, the lighter PIR-1a isoform became nearly undetectable in the *drh-3* background. Conversely, in the absence of DCR-1, we observed a depletion of PIR-1b (Fig. 3.7E). Also in the *dcr-1* mutant background, the association with DRH-3 became

nearly undetectable and the interaction with ERI-1b was weakened. The apparent differential stabilization effect of DRH-3 and DCR-1 on the two PIR-1 isoforms again correlates with the gel filtration result and suggests that PIR-1 can interact separately with each of them. In Figure 3.7E, however, it is difficult to reconcile the presence of PIR-1a with the absence of DRH-3 in the *dcr-1* mutant IP lane. In this case, it is possible that enough DRH-3 is interacting with PIR-1a to stabilize it, albeit at an undetectable level. Alternatively, PIR-1a may only be able to fold into a stable conformation in the context of a transient interaction with DRH-3. Finally, we examined the interactors recovered from the MudPIT of PIR-1 complexes in the *drh-3* mutant background. As stated in the previous section, the goal of this experiment was to maximize coverage of PIR-1b to identify posttranslational modifications. The extracts and IPs were therefore subjected to ultracentrifugation and high stringency wash conditions for maximum enrichment of PIR-1b (see Materials and Methods). Under these restrictive conditions, only the interactions with DCR-1, RDE-4 and ERI-3 remained (Table 3.3). This confirmed the strong association between PIR-1 and DCR-1, and suggested the existence of a strong core complex of these four proteins, independent of the DRH-3/RRF-3 RdRP module proposed by Thivierge *et al.* (2012). Supporting our result, the same study also demonstrated that ERI-3 directly interacts with DCR-1.

Together, the results obtained from gel filtration, DCR-1 and DRH-3 IPs, and PIR-1 IPs in *dcr-1* and *drh-3* mutants, point to potential functional distinctions between PIR-1a and PIR-1b. Because the two isoforms still overlap in high MW fractions, it is possible that slightly different complexes exist or even that both forms are present in the same complex. Since DRH-3 can pull down both isoforms and DCR-1 only pulls down one, it cannot be ruled out that DRH-3 is also able to bind PIR-1. Even though the destabilization of PIR-1 in mutant backgrounds suggests direct binding, future experiments should specifically address this by resorting, for instance, to recombinant protein pulldown assays or targeted yeast two-hybrid screens.

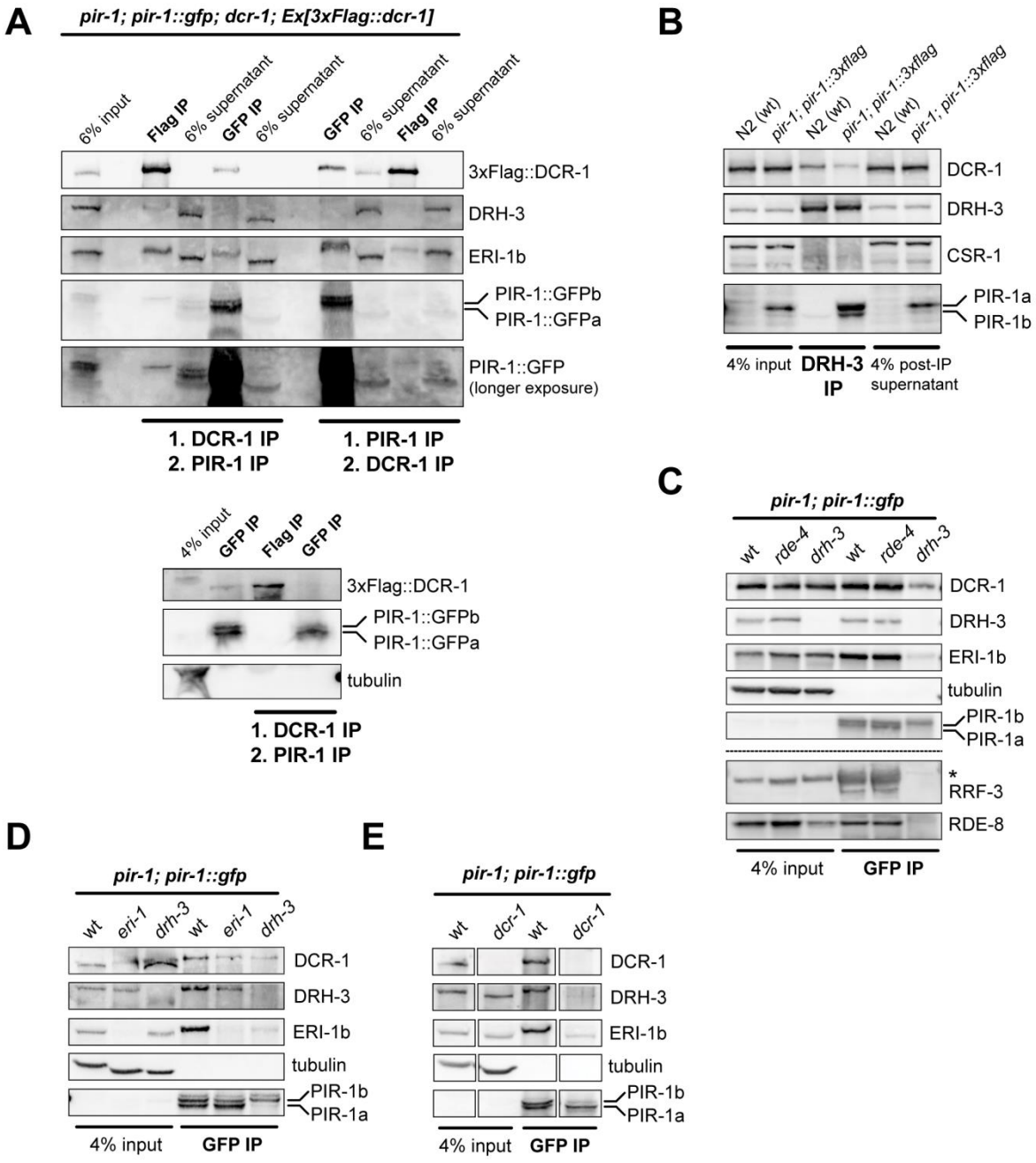



**Figure 3.7. PIR-1 isoforms do not interact equally with DCR-1. (A)** Sequential IPs of PIR-1::GFP and 3xFlag::DCR-1. The order in which the IPs were performed is indicated below each panel. The second IP used the supernatant from the first IP. The two panels constitute independent experiments to emphasize the depletion of PIR-1b when DCR-1 IP is performed first. In the top panel the DCR-1 IP recovers only one PIR-1 band. **(B)** IP of endogenous DRH-3 recovers PIR-1a and PIR-1b. CSR-1 served as a negative control. **(C)** PIR-1::GFP IPs in *rde-4* and *drh-3* mutant backgrounds. Note that only one PIR-1 isoform is maintained in *drh-3* and interactions with other proteins are lost or reduced. The asterisk signals an unspecific band. **(D)** PIR-1::GFP IPs in *eri-1* and *drh-3* mutant backgrounds. This confirms that only PIR-1a is lost in the absence of DRH-3. **(E)** PIR-1::GFP IPs in the *dcr-1* mutant background, in which only PIR-1b is not expressed. In all experiments animals were non-gravid young adults or, in the case of *drh-3* and *dcr-1*, mature sterile adults.

**Table 3.3.** List of PIR-1 Interactors Obtained in an Independent MudPIT Experiment on PIR-1::GFP IP Complexes in the *drh-3* Mutant Background with High Stringency[a]

| Protein | Length in aa/ MW in kDa | Spectral Count (abundance) | Peptide Count | Protein Coverage |
|---|---|---|---|---|
| **ERI Complex Proteins** | | | | |
| PIR-1::GFP | 476 / 54.3 | 160 | 66 | 59.9% |
| DCR-1 | 1845 / 210.9 | 21 | 18 | 11.4% |
| RDE-4 | 385 / 43.4 | 16 | 7 | 17.4% |
| ERI-3 | 578 / 66.4 | 16 | 9 | 17.8% |

[a] IP from sterile *drh-3* animals counter-selected for seven days. Only the PIR-1b isoform was recovered. High-stringency conditions included ultracentrifugation of extracts and high-salt, high-detergent post-IP washes (see Materials and Methods).

## PIR-1b Is Enriched in the Nucleus Where It Interacts with Dicer and ERI Complex Proteins and Associates with Chromatin

The localization of PIR-1 in cell nuclei raised the question of whether its association with the ERI complex occurs within the nuclear compartment. Initial purifications of nuclei followed by western blotting revealed that PIR-1b preferentially accumulates in nuclei, while PIR-1a is detectable primarily in total extract (Fig. 3.8A and B). This made it clear that a pool of cytoplasmic PIR-1 exists. However, because obtaining pure cytoplasmic extracts from *C. elegans* is not feasible, we were unable to determine whether only PIR-1a or both isoforms coexist in the cytoplasm. Next, we prepared a large amount of pure nuclei from non-gravid young adults to test protein interactions by IP of PIR-1::3xFlag in the nucleoplasm (*i.e.*, the soluble portion of lysed nuclei, excluding chromatin). As a result DCR-1, DRH-3 and ERI-1b co-purified with PIR-1b from nuclear extracts (Fig. 3.8C). This was particularly obvious for Dicer, which could not be detected in the nuclear extract input sample, but became visible after IP. This finding strongly suggests that a PIR-1-containing ERI complex, and perhaps other types of PIR-1 complexes operate in the nucleus.

When we stained fixed tissues of *pir-1::gfp* rescued animals with DAPI, some of the GFP signal overlapped with chromatin (Chapter II, Fig. 2.4B). We therefore purified chromatin from transgenic animals and tested for the presence of PIR-1 and ERI proteins by western blotting. We found that PIR-1b (but again not PIR-1a) was easily detected in purified chromatin, as were DRH-3, ERI-1b, and DCR-1 (Fig. 3.8D and E). Tubulin and histone H3 served as negative and positive markers of pure chromatin, respectively. Since

we suspect that PIR-1 activity may correlate with transcription, we took advantage of its robust association with chromatin to ask whether interfering with Pol II transcription could disrupt any of these interactions. AMA-1 (the Large Subunit of RNA polymerase II in *C. elegans*) was easily detected on chromatin, but not in total protein extracts (Fig. 3.8C and D). We were able to knock it down to a nearly undetectable level by feeding the animals *ama-1* dsRNA-expressing bacteria such that they arrested development as sterile young adults (Fig. 3.8E). RNAi of *ama-1* did not lead to significant changes in the levels of any of the tested proteins in total extracts or on chromatin. In fact, the amount of DCR-1, DRH-3, ERI-1b, RDE-8 and PIR-1 relative to histone H3 was higher in *ama-1* RNAi chromatin than in control chromatin. This argues that PIR-1, as well as all the other proteins, may interact with chromatin in an RNA-independent manner. We cannot, however, completely exclude the possibility that even after most Pol II is depleted, some RNA remains associated with chromatin and permits attachment of these factors. Whichever is the case, the stable association of PIR-1 and its interactors with chromatin suggests that the ERI complex may be constitutively positioned to operate on nascent transcripts, perhaps using them as templates for sRNA biogenesis.

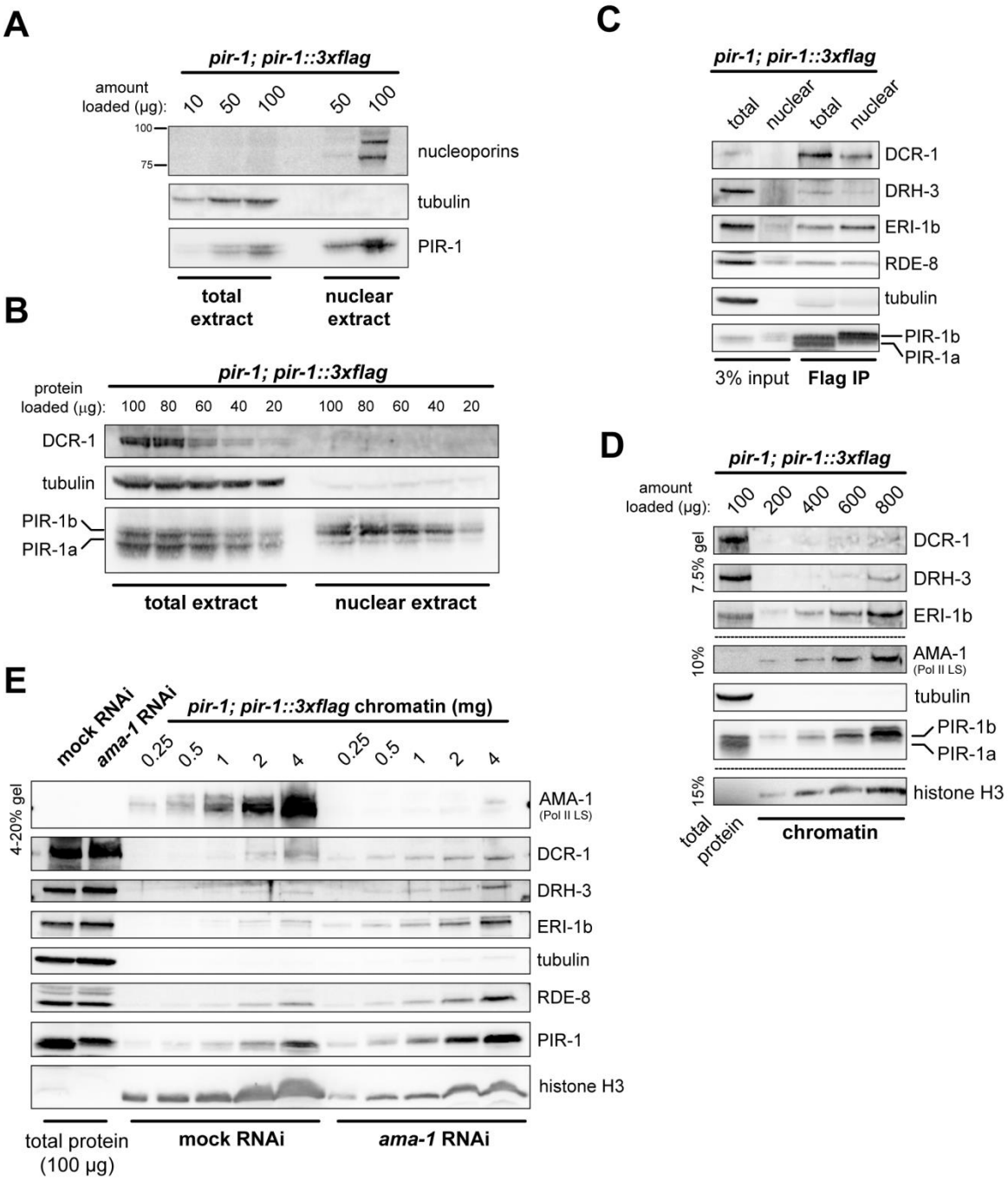



**Figure 3.8. PIR-1b interacts with the ERI complex in the nucleus and it associates with chromatin. (A)** Western blot of total and nuclear extracts. As a positive control for nuclear enrichment we used the mAb414 antibody, which recognizes the FG repeats of several nucleoporins. Tubulin served as a negative control. **(B)** PIR-1b is enriched in nuclear extracts. Most DCR-1 is found in the cytoplasm. **(C)** IP of PIR-1::3xFlag from total and nuclear extracts. PIR-1 interactions are maintained in the nucleus. **(D)** Western blot on increasing amounts of chromatin. AMA-1 is the Large Subunit of RNA Polymerase II, and along with histone H3 serve as positive markers for chromatin. Only PIR-1b associates with chromatin along with most tested interactors. Dashed lines separate different gels. **(E)** Western blot of chromatin upon *ama-1* RNAi. Dicer can be seen to associate with chromatin more clearly than in panel C. Mock RNAi consisted of feeding with bacteria carrying the empty double-T7 promoter vector L4440 used to clone RNAi target sequences.

### Human PIR1 May Also Interact with Dicer

At the onset of this project we asked whether human PIR1 (DUSP11) also interacts with Dicer. We started by testing the only commercially available anti-PIR1 antibody at the time on HeLa and HEK293T cell extracts (*H*uman *E*mbryonic *K*idney transformed with the SV-40 large *T* antigen). A band just above 35 kDa was detected, consistent with the expected size of ~39 kDa (Fig. 3.9A). The higher amount of PIR1 detected in HEK293T cells compared to HeLa cells is in agreement with *pir-1* mRNA expression levels reported in the Human Protein Atlas (mRNA-seq data; proteinatlas.org; Uhlen *et al.*, 2015). We could not detect this band in simian COS cell extracts (*C*V-1 in *O*rigin and transformed with *S*V-40 virus), consistent with the manufacturer's claim that the antibody does not cross-react with non-human homologs. We attempted validation by RNAi against *PIR1* using two synthetic siRNAs, but only obtained a reduction of 40% of the *PIR1* mRNA measured by qRT-PCR (not shown).

Since the antibody was not able to IP endogenous PIR1, we cloned the PIR-1 full-length cDNA into an expression vector in fusion with a 3xFlag tag. Simultaneously, we also obtained a full-length GST-Dicer fusion construct (Tahbaz *et al.*, 2004). We then used these plasmids for transient transfection of HEK293T cells, in which both fusion proteins were robustly expressed (Fig. 3.9B and C). In order to detect both endogenous and fusion proteins, we used the anti-PIR1 and an anti-Dicer antibody for western blots. IP of 3xFlag-PIR1 did not pull down endogenous Dicer and only pulled down a very small amount of GST-Dicer, possibly non-specifically (Fig. 3.9B). Upon reciprocal IP using the anti-Dicer antibody, however, we were able to recover 3xFlag-PIR1 even in the absence of GST-Dicer (Fig. 3.9C). Despite this encouraging result, it should be noted that there was no enrichment for endogenous PIR1. Perhaps the fraction of Dicer associating with PIR1 is too low to detect in this manner, and only in conditions of overexpression can this association be observed due to out-competition of existing endogenous Dicer partners. These results hint that PIR1 and Dicer may also interact in vertebrates. Because the interactions were only seen in conditions of overexpression, further experiments should be conducted to conclusively validate the interaction.

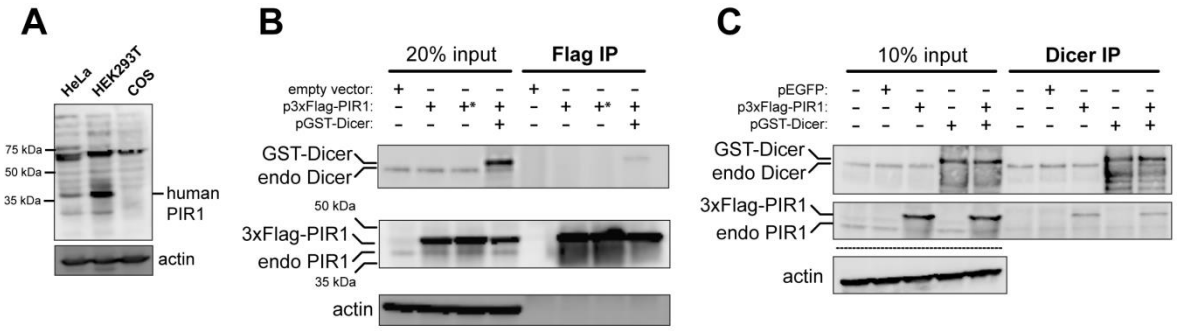



**Figure 3.9. Human PIR-1 is expressed in HEK293T cells and associates with Dicer when overexpressed. (A)** Western blot of endogenous PIR1 (DUSP11) in human cell lines. The antibody does not crossreact with simian COS kidney cell extract. **(B)** IP/western of FLAG-PIR1 in cells transiently transfected with both FLAG-PIR1 and a full-length human Dicer GST fusion. Actin was used as an input loading control. The asterisks indicate that cells were transfected with double the amount of vector as the previous lane. **(C)** IP/western of Dicer following transfections as in (B).

## Contributions

Processing, running, and analysis of immunoprecipitated protein samples by MudPIT were performed by James Moresco, in the laboratory of John Yates III, at the Scripps Institute, La Jolla, USA. Gel filtration was performed with assistance from Hsin-Yue Tsai in our laboratory. All other procedures and analyses were performed by Daniel Chaves.

## DISCUSSION

Through the work reported in this chapter we have confirmed the association of PIR-1 with Dicer in *C. elegans*. We further widened the scope of its interactions to include most proteins known to compose the ERI complex, pointing to an endogenous sRNA biogenesis function. Interestingly, at a time when evidence that sRNA pathways operate in the nucleus is steadily accumulating, we showed for the first time that the ERI complex is present in the nucleus. Below we discuss the implications of these discoveries and speculate about additional functions that may help us understand why *pir-1* mutants exhibit such severe developmental phenotypes, beyond what can be explained from loss of ERI activity alone.

### The Stable Association of PIR-1 with the ERI Complex Suggests a Role in Dicer-Dependent 26G-RNA Biogenesis

The first clue that PIR-1 could be part of the ERI complex came from MudPIT analysis of ERI-1, in which PIR-1 peptides were detected at very low abundance (Duchaine *et al.*, 2006). Our characterization of PIR-1 immunocomplexes by MudPIT and western defined very clearly that this is indeed the case. From our results it is also clear that PIR-1 does not associate with proteins required for the canonical exo-RNAi response, contradicting initial studies which implicated PIR-1 in Dicer-mediated cleavage of dsRNA into effector siRNAs (Duchaine *et al.*, 2006). We cannot rule out, however, that associations with exo-RNAi factors besides Dicer occur transiently and therefore could not be detected by the methods we employed.

In the most complete validation to date of the interactions occurring within the ERI complex, MudPIT of ERI-5 immunoprecipitates also uncovered PIR-1 at very low abundance (Thivierge *et al.*, 2012). The study resulted in the proposal of a two-module system for the ERI complex: the RdRP module, consisting of RRF-3, DRH-3 and ERI-5, and another consisting of DCR-1. The combination of the two modules is made possible by ERI-5, which was shown to bind directly to DCR-1. Our western analysis of extracts depleted of PIR-1 made it clear that PIR-1 must interact with only a small fraction of DCR-1 and other components (Fig. 3.2, page 120). This is not surprising, given our knowledge that some of these factors integrate other complexes, namely DRH-3 in RdRP complexes consisting of RRF-1/EKL-1 or EGO-1/EKL-1 for production of 22G-RNAs, and DCR-1 interacting with RDE-1/RDE-4/DRH-1 for synthesis of primary siRNAs (Tabara *et al.*,

2002; Duchaine *et al.*, 2006; Gu *et al.*, 2009; Thivierge *et al.*, 2012). An important result of our study was the discovery that the endoribonuclease RDE-8 is part of the ERI complex. Since both the RdRP RRF-3 and RDE-8 cease to associate with PIR-1 in the *drh-3* mutant background (Fig. 3.7, page 133), we propose that RDE-8 interacts with the RdRP module of the complex. RDE-8 has been shown to catalyze the cleavage of mRNAs at regions targeted by RNAi. The 3' ends of the resulting fragments are then stabilized via RDE-3-mediated 3' polyurydilation, to be subsequently used as templates for multiple rounds of RdRP 22G-RNA production (Tsai *et al.*, 2015). The addition of an endoribonuclease activity to the ERI complex has important implications to our current understanding of the molecular events that lead to the production of 26G-RNAs. In Chapter IV we will present a working model for 26G-RNA synthesis by the ERI complex.

Western blots of PIR-1 indicated that it is found as two distinct isoforms throughout post-embryonic development. The indication that the isoforms may not be functionally equal came from our analysis of PIR-1 complexes in different genetic deletion backgrounds (Fig. 3.7, page 133). Strikingly, the heavier form PIR-1b was not detectable in *dcr-1* deletion extracts. This isoform preference was corroborated by the differential co-fractionation of the isoforms by gel filtration, where PIR-1b co-fractionates strictly with DCR-1 (Fig. 3.3, page 123). Together, these results demonstrate that PIR-1b is an obligatory Dicer cofactor. This protein stabilization effect is not unprecedented for DCR-1 interactors, as *dcr-1* deletion extracts have been shown to be depleted of the primary Argonaute RDE-1 and the helicase DRH-1 (Thivierge *et al.*, 2012). The fact that we observe the PIR-1b and DCR-1 proteins cofractionating over a large range of molecular weights supports the notion that PIR-1 may not only be required to assist DCR-1 in 26G-RNA biogenesis but also in other Dicer-dependent processes. The lighter isoform PIR-1a could not be detected in IPs from *drh-3* deletion extracts, pointing to the possibility that PIR-1a is stabilized through a direct interaction with DRH-3. This is in agreement with the ability to recover both isoforms upon DRH-3 IP, but only PIR-1b upon DCR-1 IP.

The possibility that PIR-1 interacts directly and separately with both DCR-1 and DRH-3 is not farfetched, as they are considered to be paralogs based on sequence and domain topology. They share four domains: a DEAD-box helicase domain, an ATP-binding helicase superfamily 1/2 domain, a C-terminal helicase domain and a P-loop with nucleoside triphosphate hydrolase activity. It is possible that the functions that PIR-1 assumes when interacting with DRH-3 are distinct from those dictated by its association with DCR-1. Intriguingly, only PIR-1b is detected in embryos (Fig. 3.4, page 124) further

highlighting the potential for functional differences between the isoforms. It is possible that PIR-1b is the form required to support intense cell division. This form alone is capable of recovering all of the ERI complex proteins we could detect. Additionally, when we mutagenized PIR-1 residues that did not affect the ability of the protein to rescue, only PIR-1b was expressed (Fig. 3.6, page 130). Although we cannot explain why PIR-1a was absent, this implies that PIR-1b may be necessary and sufficient for rescue. Perhaps PIR-1a represents a non-functional form that is processed from the heavier protein in order to induce delocalization or to sequester important protein partners, thus regulating the activity of PIR-1b.

In trying to understand the origin of the two PIR-1 isoforms, one of our goals was to be able to create transgenic animals expressing them individually, to start teasing apart potential functional differences. Although it seems likely that the isoforms arise from post-translational modifications, namely in the form of multiple phosphorylations and N-glycosylations (some of which we have confirmed by mass spectrometry), we did not completely determine their identity and were therefore unable to address this issue. It would also be useful to know whether PIR-1 completely disappears in a *dcr-1; drh-3* double mutant. Should this be the case, then further study of PIR-1 and its isoforms could help uncover new biological functions for DCR-1 and DRH-3 that could explain their own essential roles in development and fertility.

### The Implications of a New Cellular Address for the ERI Complex

Through isolation of nuclei from transgenic animals, we were able to show that PIR-1b is the isoform that primarily localizes to the nucleus. Having established that this isoform is only stable in the presence of DCR-1, we presume that PIR-1 functions in the nucleus in association with DCR-1. Additionally, PIR-1a is depleted from nuclear extracts, implying that PIR-1 also operates in the cytoplasm. Since we could not obtain pure cytoplasmic extracts, it is possible that PIR-1b is also present and functional in the cytoplasm. When we immunoprecipitated PIR-1 from nuclear extracts we found it to still associate with DCR-1, DRH-3, ERI-1b and RDE-8, suggesting for the first time that the ERI complex is assembled in the nucleus. One caveat of these experiments is the potential contamination of nuclear extracts with germline P granules, which assemble outside of the nuclei in association with nuclear pores. Although we do not know whether this association resists our nuclear isolation procedure, we believe that heavy contamination is unlikely. First, DRH-3, which normally concentrates heavily in P granules, is not detectable in

nuclear inputs; second, we never saw PIR-1 concentrating at P granules; and third, the isolated nuclei are a mixture of germline nuclei and somatic nuclei, which do not have P granules in their outer periphery, diluting any residual P granules. Although we would have benefited from additional controls, we assume that most interactions recovered from nuclear extracts occur with nucleoplasmic PIR-1. The fact that these factors, including PIR-1, were found to also associate with chromatin, provides additional proof that they all reside in the nuclear compartment, independently of any association with cytoplasmic perinuclear domains.

Current knowledge regarding the subcellular localization of most proteins of the ERI complex is fragmented and incomplete. ERI-1b, for instance, has been shown to localize primarily to the cytoplasm in the soma (Kennedy *et al*., 2004; Gabel and Ruvkun, 2009), but nothing is known about its localization in the germline. DRH-3 has been assumed to localize in the cytoplasm of somatic cells based on its interactions with RRF-1 but directly shown to localize to perinuclear germline P granules with EGO-1 throughout development (Gu *et al.*, 2009; Claycomb *et al*., 2009). However, its reported functions in maintaining chromosomal integrity upon exposure to DNA-damaging agents (Nakamura *et al.*, 2007) and in ensuring proper chromosome segregation in germline and embryos (Duchaine *et al.*, 2006) argue that a nuclear pool of DRH-3 must exist. Regarding DCR-1, though most of its activity has also been assumed to occur in the cytoplasm, namely in exo-RNAi and miRNA processing, it has also been directly detected in the nuclei of germline cells (Beshore *et al.*, 2011). In the germline, RDE-8 was found to localize to the cytoplasm and to co-localize with perinuclear Mutator foci (Tsai *et al.*, 2015). Mutator foci were defined by the presence of several factors required for 22G-RNA synthesis, localizing to the cytoplasm of somatic cells, and adjacently to P granules around the nuclei of the germline (Phillips *et al.*, 2012). Finally, no studies exist regarding the cellular localization of the RdRP RRF-3. Unfortunately, we were unable to detect RRF-3 when reprobing our western blots of nuclear extracts and IPs. We attribute this to an inneficient antibody combined with the partial loss of protein from the blotted membrane during the harsh stripping procedure.

Numerous studies have shown the occurrence of Dicer-dependent processes in the nucleus. Examples include Dicer association with chromatin in *S. pombe* to promote centromeric heterochromatin assembly and to silence euchromatic genes co-transcriptionally (Colmenares *et al.*, 2007; Motamedi *et al.*, 2004; Woolcock *et al.*, 2011); nuclear Dicer associating with Pol II in human cells to restrict the accumulation of

endogenous dsRNA, thus preventing interferon-induced (IFN) apoptosis (White *et al.*, 2014); and the generation of transcription start site-associated sRNAs around the promoters of protein-coding genes in human cells (Zamudio *et al.*, 2014; Rybak-Wolf *et al*., 2014). A few reports have begun to address the properties of Dicer that allow nuclear localization. Despite lacking classical nuclear localization signals, the putative double-stranded RNA binding domain (dsRBD) of *S. pombe* Dicer was shown to be required not for substrate binding, but rather to allow nuclear retention of the enzyme (Emmerth *et al.*, 2010; Barraud *et al.*, 2011). The dsRBD domain was also found to constitute a non-canonical NLS that enables human Dicer to enter the nucleus (Doyle *et al.*, 2013). Although overexpressed human Dicer has always been reported as being restricted to the cytoplasm, endogenous Dicer has recently been abundantly detected in the nucleus by immunofluorescence and western analysis of cytoplasmic and nuclear fractions from HEK293 cells (White *et al*., 2014). Another study went further to demonstrate that nuclear Dicer and other RNAi factors exist in the nuclei of several human cell lines at abundances close to those of cytoplasmic extracts and, importantly, that nuclear Dicer is catalytically active and found in multiprotein complexes (Gagnon *et al.*, 2014). In *C. elegans*, a rescuing GFP::DCR-1 transgene revealed germline DCR-1 to be expressed predominately in the cytoplasm surrounding the nuclei of mitotic germ cells. Subsequently, during pachytene, it was shown to acquire a more prominent nuclear localization, including in oocytes (Drake *et al.*, 2014). Moreover, DCR-1 has been found in germline nuclei to associate with the inner-facing nuclear membrane at nuclear pores connected to external P granules (Beshore *et al.*, 2011).

We have shown that PIR-1 and its interactors are stably associated with chromatin (Fig. 3.8, page 136). With the goal of finding target loci for PIR-1, we optimized and performed chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) of PIR-1::GFP from embryos and compared it to Pol II ChIP-seq from the same material. Although the initial result showed a strong overlap of PIR-1 with Pol II occupancy at numerous genes, we were not able to reproduce these results (not shown). Perhaps the association of PIR-1 with chromatin is too indirect, or the sensitivity of the method is compromised by the abundance of PIR-1 in the nucleoplasm and cytoplasm. Upon RNAi knockdown of Pol II the interactions with chromatin were not perturbed, which led us to infer that they may not depend on mRNA transcription. Consistent with this, we have not been able to co-IP Pol II with PIR-1, even from formaldehyde-crosslinked extracts (not shown). We cannot completely rule out, however, that enough residual Pol II activity is

still taking place so as to allow mRNA-mediated tethering of the examined proteins to chromatin or that, alternatively, PIR-1 complexes interact with Pol I or Pol III RNA.

Given the strong interaction between PIR-1 and DCR-1, studies of chromatin-associated Dicer provide some of the few links that allow us to ask how PIR-1 may function there. In *S. pombe*, for instance, Dicer has been shown to promote transcription termination at highly transcribed protein coding genes (Pol II), tDNA (Pol III) and rDNA (Pol I), in order to avoid the DNA damage that results from the collision of DNA replication forks with these highly transcribed regions (Castel *et al.*, 2014). Intriguingly, this function does not depend on the formation of siRNAs, as a catalytically-inactive Dicer still leads to Dicer-dependent termination at these loci. This suggests a novel mechanism for chromatin-localized Dicer, distinct from the co-transcriptional Dicer-dependent generation of siRNAs at centromeric repeats that leads to deposition of repressive histone H3K9me2, also in *S. pombe* (reviewed in Castel and Martienssen, 2013). In human cells, knockdown of Dicer leads to accumulation of dsRNA resulting in apoptotic cell death (White *et al.*, 2014). In this study, the association of human Dicer with chromatin was found to be transcription-dependent, interacting with Pol II via double-stranded RNA (dsRNA). Small RNAs overlapping with regions of dsRNA accumulation were abundantly detected, implying that an RNAi mechanism may aid in the prevention of deleterious dsRNA accumulation. There are currently no reports providing evidence for a direct interaction of Dicer with chromatin. Considering the proliferation and growth arrest that we observe upon loss of PIR-1 in *C. elegans*, as well as its pervasive expression and interaction with DCR-1 in the nucleus, we speculate that these proteins could be degrading excess dsRNA in the nucleus. In this scenario, the absence of PIR-1 and DCR-1 from nuclei would lead to genomic instability in dividing cells. In turn, upon DNA replication, DNA damage checkpoints would be activated, and overall development would cease. We believe that determining whether dsRNA accumulates in *pir-1* and *dcr-1* mutants is a worthwhile avenue of future research.

Lastly, a proteomic survey of proteins associated with *C. elegans* spermatogenic (spermatocytes and spermatids) and oogenic chromatin failed to find PIR-1, DCR-1 or any of the other ERI factors (Chu *et al.*, 2006). While for oocytes this correlates well with the absence of nuclear PIR-1, the degree to which PIR-1 localizes to nuclei or chromatin in spermatogenic cells is not clear from our microscopy analyses. Based on the aforementioned study, PIR-1 is therefore not likely to associate with spermatocyte and spermatid chromatin. Rather, its spermatogenesis-promoting activity may be taking place

at earlier meiotic stages. This concurs with the expression pattern we found in male germlines (Fig. 2.7, page 85), where PIR-1 expression is highest in nuclei preceding spermatocytes. At our current level of knowledge, we can only speculate about the assembly of the complete ERI complex on chromatin. Assuming that it does occur, a good candidate for direct binding to chromatin would be the Tudor domain ERI-5 protein. The conserved Tudor structural motif has been shown to bind methylated histone tails to facilitate DNA damage repair and regulate transcription (reviewed in Lu and Wang, 2013). This association could poise the ERI complex to directly engage nascent transcripts and use them as templates for 26G-RNA synthesis in the nucleus. This would allow for rapid triggering of 26G-RNA-dependent synthesis of 22G-RNAs for efficient silencing of transcripts. Such a model is supported by data presented in Chapter IV, where we will expand this discussion point.

### Clues to PIR-1 Functions beyond the Nonessential ERI Complex

The activity of the ERI complex is required for the biogenesis of 26G-RNAs that allow spermatogenesis to proceed at high temperatures (ALG-3/4 class) and silencing of a few genes in somatic cells (ERGO-1 class), but it is dispensable for critical aspects of development (Han *et al.*, 2009; Pavelec *et al.*, 2009; Gent *et al.*, 2010; Vasale *et al.*, 2010; Conine *et al.*, 2010; Fischer *et al.*, 2011; Conine *et al.*, 2013). Since DCR-1 and PIR-1 are both essential for normal development, their activity must extend to other ERI-unrelated processes. DCR-1, in particular, has long been known to associate with other proteins to produce miRNAs. Many of the developmental defects observed in *dcr-1* loss-of-function mutants are attributed to a global depletion of miRNAs which coordinate expression at a genome-wide level. Since PIR-1 associates with DCR-1, we hypothesized that it could aid miRNA production. However, as we demonstrate in Chapter IV, *pir-1* mutants do not exhibit defects in miRNA accumulation. We therefore presume that PIR-1 must partner with different proteins to participate in other pathways. Moreover, given the abundance of the PIR-1/DCR-1 interaction, and that the PIR-1b isoform only accumulates when DCR-1 is expressed, it is plausible that some of the essential processes may require the joint action of both proteins.

Since IP experiments cannot reveal transient interactions or interactions that occur in a small number of cells, we cannot exclude that further important PIR-1 or PIR-1/DCR-1 partners exist. Additionally, as an example of the limitations of adopting a single experimental approach, we noted that despite the low abundance of ERI-1b and RDE-8

peptides obtained through MudPIT, we were able to detect them robustly by western blotting. This discrepancy could reflect protein specific properties that make them less suitable for detection by mass spectrometry combined with a very high affinity of the antibodies developed against them. On the other hand, some proteins detected in some of the MudPIT experiments as abundantly as ERI-1b and RDE-8 turned out to be undetectable by western, despite the effectiveness of available antibodies (including the essential Argonaute CSR-1 and the nucleostemin homolog NST-1, discussed in Chapter II). With such examples we should not rule out that many of the low abundance proteins we detected may constitute true PIR-1 interactors, warranting validation with additional experiments. We therefore curated some of the most interesting factors, mostly related to RNA metabolism and processing, from all the MudPIT runs (Table in Appendix C, page 297). We excluded typical contaminants such as ribosomal proteins, actins, or vitellogenins (abundant components of *C. elegans* egg yolk). Our list includes some factors that were only found in PIR-1 IPs, but it also includes proteins that were recovered from the negative control. This takes into account the fact that reassociation of RNA binding proteins between RNA molecules has been proven to occur in lysates used for immunoprecipitation (Mili and Steitz, 2004). One could thus imagine that in some cases factors indirectly associated to PIR-1 via common RNA targets could be replaced by the same $^{15}$N-labelled proteins derived from the negative control extract. Below we discuss potential interactors which we think are worthy of further pursuit.

The proteins comprising the list in Appendix C function in processes that could explain some of the most severe *pir-1* mutant phenotypes, as several of those proteins are essential themselves. Some interesting examples include the redundant miRNA Argonautes ALG-1 and ALG-2. Both proteins were found at low abundance in DCR-1 and ERI-1 immunoprecipitates by MudPIT (Duchaine *et al.*, 2006; Thivierge *et al.*, 2012), suggesting that a small subset of ERI complexes may incorporate these Argonautes. The fact that we can recover them through PIR-1 corroborates this and suggests that ALG-1/2 may have miRNA unrelated roles. Another interesting example is the worm-specific nuclear RNAi component NRDE-1, which was recovered in MudPIT under high stringency. This protein localizes to the nucleus and is required, along with other factors, to inhibit Pol II elongation and deposit repressive histone marks in genes targeted by exogenous and endogenous 22G-RNAs (Burkhart *et al.*, 2011). Even though this process is not immediately essential for development, mutants for any of its factors, including the nuclear Argonaute HRDE-1, exhibit a loss of gamete formation and function over generations,

culminating in completely sterile animals (mortal germline phenotype; Buckley *et al.*, 2012). PIR-1 could conceivably participate in this process by dephosphorylating triphosphorylated 22G-RNAs in germline nuclei to regulate the deposition of repressive histone marks that maintain germline immortality.

We have also found components of important protein complexes found in germline P granules and cytoplasmic granules equivalent to P-bodies. So far, more than 40 proteins found to associate with P granules are proteins with RNA-binding domains or are predicted to regulate translation (Updike and Strome, 2010) and, similarly to P-bodies, also contain Argonautes. In *C. elegans*, the Argonautes ALG-3, CSR-1, PRG-1, and WAGO-1 were found to accumulate in P granules, establishing them as regions of intense sRNA-based regulation of transcripts (Batista *et al.*, 2008; Claycomb *et al.*, 2009; Gu *et al.*, 2009; Conine *et al.*, 2010). Among the factors we found by MudPIT, are the helicase CGH-1 (*C*onserved *G*ermline *H*elicase 1) and the RNA-binding Sm domain protein CAR-1 (*C*ytokinesis, *A*poptosis, *R*NA-Associated 1) which are known to interact via RNA (Audhya *et al.*, 2005). Both proteins are conserved and are essential for formation and maintenance of P granules and P-bodies. Curiously, we also found the conserved poly(A)-binding proteins PAB-1 and PAB-2, which regulate the stability and translation of mRNAs. PAB-1 has been shown to interact with AIN-1, the *C. elegans* ortholog of the Argonaute-interacting P-body GW182 protein, which in turn interacts with the miRNA Argonaute ALG-1 and with deadenylase complexes (Eulalio *et al.*, 2009; Kuzuoglu-Ozturk *et al.*, 2012). Mutants of PAB-1 display impaired germline mitotic proliferation, and the protein was found to physically associate with CGH-1 and CAR-1 in P granules (Ko *et al.*, 2010; Ko *et al.*, 2013). While we have not seen PIR-1 concentrating in P granules, we cannot exclude that some of it exists in these important ribonucleoprotein centers. Alternatively, these interactions may be taking place in cytoplasmic granules of somatic and germline cells.

Another P-granule component that should be validated is the germline VASA-related DEAD-box helicase GLH-1 (*G*erm*l*ine *H*elicase 1). This constitutive P-granule protein is essential for *C. elegans* fertility by promoting germline proliferation and P-granule integrity (Kuznicki *et al.*, 2000; Schisa *et al.*, 2001). Importantly, it has been shown to localize in proximity to nuclear DCR-1 located at nuclear pores, and demonstrated to physically interact with DCR-1, the loss of which also disrupts P granules (Beshore *et al.*, 2011). Since *pir-1* mutants do not exhibit visible P-granule defects as judged by the localization of PGL-1, (Chapter II, Fig. 2.3), it is possible that germline P

granules and other DCR-1/GLH-1 RNP granules still assemble normally, but that regulation of specific mRNAs is compromised. It will therefore be important to validate these interactions and deepen the study of these granules in the absence of PIR-1.

In line with the discoveries that human PIR1 interacts with pre-mRNA splicing factors, we detected peptides for some conserved *C. elegans* splicing proteins, all with essential roles in development or fertility. HRP-1 (Human *H*n*RNP* A1 Homolog 1), HRP-2 and SQD-1 (homologous to *Drosophila SQD*/Squid Protein) are orthologs of various heterogeneous nuclear ribonucleoproteins (hnRNPs) that coordinate not only aspects of pre-mRNA processing, mRNA metabolism and transport, but also DNA replication, DNA damage repair and telomere maintenance (reviewed in He and Smith, 2009). We also found UAF-1 (*U2AF* Splicing Factor 1), an ortholog of the large subunit of splicing factor U2AF which recognizes specific RNA polypyrimidine sequences preceding 3' splice sites. U2AF has been shown to determine the accumulation of alternatively-spliced variant transcripts, depending on the strength of its association with pre-mRNA (Zorio and Blumenthal, 1999; Ma and Horvitz, 2009). The participation of PIR-1 in pre-mRNA splicing would be in agreement with the ubiquitous nuclear localization in somatic and germ cells throughout development and with the deleterious effects on development and fertility described for splicing protein mutants. In light of the precedent for interaction with splicing factors in humans, we believe that these interactions should also be the focus of future studies. Specifically, the occurrence of splicing defects in *pir-1* mutants should be examined in genes for which splicing variants have been well defined, but also at a global level by RNA-seq approaches. Lastly, we have identified a variety of nucleolus-associated proteins. Considering that human PIR1 was found to localize to nucleoli in a transcription-dependent manner (Andersen *et al.*, 2005), and that its silencing leads to the accumulation of some rRNA precursor intermediates (Tafforeau *et al.*, 2013), we do not rule out that PIR-1 could have a direct role in rRNA processing or metabolism in *C. elegans*. Alternatively, PIR-1 could indirectly affect rRNA metabolism by assisting small nucleolar RNA processing. Such roles would be consistent with the impairment of growth and cell proliferation.

## MATERIALS AND METHODS

### Immunoprecipitations

Protein extraction was generally performed in ~3 worm-pellet volumes of lysis/IP buffer containing 25 mM HEPES-KOH pH 7.5, 2 mM EDTA, 0.5% NP-40, and 150 mM KCl, supplemented with protease and phosphatase inhibitors (Roche protease and phosphatase inhibitor cocktail tablets, 1 tablet of each per 10 ml of buffer). All steps were performed on ice or 4˚C. Animals were crushed with 75-100 strokes in a metal dounce. Lysates were cleared by centrifugation at 10,000x *g* for 10 minutes. Protein concentration was measured using the Lowry-based DC Protein Assay from Bio-Rad according to the manufacturer's instructions. We generally obtained concentrations of 10-12 mg/ml. Per 1.5 ml microtube, 5 mg of protein in 1 ml of IP buffer were pre-cleared with 25 µl of Protein A/G PLUS-Agarose beads (Santa Cruz Biotechnologies) for 30 minutes to 1 hour. Antibody was then added to the cleared lysate and incubated for 1 hour. The lysate was transferred to a fresh tube containing 25 µl of protein A/G beads and incubated for 1 hour. Beads were washed at least 3x 5 minutes with lysis buffer. After complete removal of the buffer, the immunoprecipitates were eluted by incubating the beads with 25-30 µl of 2X protein sample buffer (1X is 50 mM Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 0.01% (w/v) bromophenol blue, 100 mM DTT) for 5 minutes at 95˚C. Generally, the entire sample was loaded on 4-15% SDS-polyacrylamyde gels. Inputs were typically diluted to 5 µg/µl in 2X protein sample buffer and 50 or 100 µg were loaded on gels. Antibodies for IPs were used as follows: 25 µl of anti-Flag M2 affinity gel (Sigma-Aldrich) for Flag IPs; 6 µg of anti-GFP mouse monoclonal antibody (WAKO) per GFP IP; 10 µl of affinity purified polyclonal antibody for IP of DRH-3 (Gu *et al.*, 2009).

### Western Blotting

For most experiments, proteins were resolved by SDS-PAGE using precast polyacrylamide gradient 4-15% gels (Bio-Rad Criterion gel system). Optimal resolution of PIR-1 isoforms was achieved using 10% gels. Proteins were transferred to Hybond-C Extra nitrocellulose membranes (GE Healthcare) using a Bio-Rad Trans-Blot SD semi-dry apparatus with protein transfer buffer (20% methanol, 0.04% SDS, 48 mM Tris, 39 mM glycine). Membranes were blocked with PBST with 5% (w/v) of nonfat dried milk for 30 minutes to 1 hour at room temperature. Antibodies were diluted in PBST/milk. Primary antibodies were incubated at 4˚C overnight and secondary antibodies at room temperature

for 1-3 hours. Membranes were washed 3x 5 minutes in PBST after each incubation. Chemiluminescence from the HRP-conjugated secondary antibodies was developed with the Western Lightening ECL kit (Perkin Elmer) and detected with an LAS-3000 Intelligent Dark-Box system (Fujifilm). Primary antibodies were used at a 1:250 or 1:500 dilution if they were antisera (unless otherwise noted, these antibodies were developed in our laboratory). These include anti-DCR-1, anti-ERI-1, anti-RRF-3 (a gift from Thomas Duchaine, McGill University, Canada), anti-RRF-1, anti-EGO-1 and anti-MUT-7 (all rabbit polyclonals). Affinity-purified primary antibodies were generally used at 1:500 to 1:1000 dilutions and include anti-DRH-3, anti-CSR-1, anti-RDE-8, anti-RNA Pol II 8WG16 (ChIP-grade from Abcam), anti-nuclear pore complex proteins mAb414 (Covance), anti-histone H3 (Cell Signalling), HRP-conjugated M2 anti-Flag (Sigma-Aldrich), rabbit anti-GFP (Torrey Pines Biolabs or Invitrogen) and rat anti-tubulin (used at 1:2000, Jackson Laboratories). Anti-rat, mouse or rabbit HRP-conjugated secondaries (Santa Cruz Biotechnologies) were incubated at 1:1000 to 1:2000 dilutions.

### Silver Staining and Mass Spectrometry

Polyacrylamide gels were silver-stained using the SilverQuest Kit (Invitrogen) according to the manufacturer's instructions. Excised bands were sent for analysis by mass spectrometry at the Taplin Mass Spectrometry Facility from Harvard Medical School.

### Immunoprecipitations for MudPIT Proteomics

For comparison of proteins immunoprecipitated by PIR-1::GFP and PIR-1::3xFlag, IPs for each protein were carried out in the presence of an equal amount of N2 (wild-type) worm extract labelled with $^{15}N$ (heavy isotope). During analysis, this allowed to discriminate which proteins bound to the IP matrix nonspecifically by comparing the heavy-to-light ratio of peptides ($^{15}N$ peptides were not expected to bind specifically). The transgenic strains used were the integrated lines obtained by bombardment and were grown to be young gravid adults on HB101 *E. coli*. $^{15}N$-labelled N2 animals were grown for three generations on nitrogen-free plates with HB101 *E. coli* grown on Spectra9-N medium (SpectraGases) and harvested as young gravid adults. Immunoprecipitations were generally carried out as described above. 5 mg of heavy N2 extract was mixed with 5 mg of either of the unlabeled transgenic extracts in 2 ml microtubes at a concentration of 5 mg/ml for a total of 40 mg per strain. After pre-clearing with protein A/G beads, extracts were incubated with either anti-Flag M2 agarose beads (Sigma) or with protein A/G beads

(Santa Cruz Biotechnologies) crosslinked to anti-GFP monoclonal antibody (Wako). The crosslinking was performed with DMP (dimethyl pimelimidate, Sigma) to prevent IgG contamination of eluted material. IPs proceeded at 4˚C for 2 hours and were followed by 5x washes of 5 minutes each with lysis/IP buffer containing protease and phosphatase inhibitors. Bound protein was eluted with 2 bead volumes of 2X protein sample buffer without dye or glycerol (100 mM Tris-HCl pH 6.8, 4% SDS, 200 mM DTT) followed by 2 minutes at 65˚C and 5 minutes at 95˚C with constant shaking. The supernatant was collected and another bead volume of protein sample buffer was added and recovered. The pooled supernatants were run through a 0.45 µm filter to eliminate agarose residue. Protein was precipitated with 4 volumes of cold acetone followed by overnight incubation at -20˚C, and finally collected by centrifugation at 20,000x *g* for 10 minutes at 4˚C. The pellets were washed again with acetone and air-dried. For the $^{15}N$ total protein (required for later analysis), 100 µg of N2 extract were precipitated by adding 250 µl of trichloroacetic acid in 1 ml of the diluted protein, followed by incubation at -20˚C for 10 minutes, centrifugation as above, and two brief washes with acetone. For PIR-1::GFP IP MudPIT the procedure was identical, except that there was no addition of a labelled wild-type extract and worms were non-gravid young adults. For PIR-1::GFP IP in the *drh-3* mutant background, animals were counterselected on ivermectin for 7 days at 20˚C (sterile adults). Lysis, IP and washes were performed under more stringent conditions in order to further minimize recovery of nonspecific interactors. Lysis buffer was 25 mM HEPES-KOH pH 7.5, 150 mM KCl, 0.5% NP-40, 0.5% sodium deoxycholate, with protease and phosphatase inhibitors. Before IP the lysate was ultracentrifuged for 1 hour at 100,000x *g* at 4˚C to further clean up the sample from large mRNPs. The first wash after IP was with lysis buffer, followed by 2x 5 minutes high-stringency washes with 25 mM HEPES-KOH pH 7.5, 1 M KCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, and finally 2x 5 minutes washes with 25 mM HEPES-KOH pH 7.5, 150 mM KCl, 0.5% NP-40 (to wash away salt and ionic detergents).

### MudPIT Sample Preparation and Runs

For the $^{15}N/^{14}N$ experiment, air-dried pellets were dissolved in 60 µl of 0.1% Rapigest SF (Waters Corporations) in 50 mM ammonium bicarbonate. Proteins were reduced with 5 mM Tris(2-carboxyethyl) phosphine hydrochloride (Sigma-Aldrich) and alkylated with 10 mM iodoacetamide (Sigma-Aldrich). Proteins were digested for 18 hours at 37˚C with 0.5 µg trypsin (Promega). The digestion was stopped with formic acid (5%

f.c.). After 1 hour at 37˚C, debris were removed by centrifugation for 30 minutes at 18,000x *g*. For the PIR-1::GFP IP experiment, the air-dried pellets were dissolved in 100 mM Tris-HCl pH 8.5, 8 M urea. Protein reduction and alkylation were performed as above. Proteins were digested for 18 hours at 37˚C in 100 mM Tris pH 8.5, 2 M urea, 1 mM $CaCl_2$ with 2 µg of trypsin (Promega). Digestion was stopped with formic acid (5% f.c.). Debris were removed by centrifugation for 30 minutes at 18,000x *g.* For the PIR-1::GFP IP in the *drh-3* mutant background, treatment was the same as in the $^{15}N/^{14}N$ experiment, except for protein digestion, which occurred for 3 hours at room temperature with 1 µg of chymotrypsin (Princeton Separations). Deionized water (18.2 MΩ $cm^{-1}$, Barnstead) was used for all preparations.

A MudPIT microcolumn (Wolters *et al.*, 2001) was prepared by first creating a Kasil frit at one end of an undeactivated capillary with a 250 µm inner diameter (ID) and a 360 µm outer diameter (OD) (Agilent Technologies). The Kasil frit was prepared by briefly dipping a 20-30 cm glass capillary tube in well-mixed 300 µl Kasil 1624 (PQ Corporation) and 100 µl formamide at 100˚C for 4 hours, and cutting the frit to ~2 mm in length. Strong cation exchange (SCX) particles (SCX Luna, 5 µm diameter, 125 Å pores, Phenomenex) were packed in-house from particle slurries in methanol. An additional 2.5 cm of reversed phase particles (C18 Aqua, 3 µm diameter, 125 Å pores, Phenomenex) were then packed into the capillary using the same method as SCX loading, to create a biphasic column. An analytical RPLC column was generated by pulling a 100 µm ID/360 µm OD capillary (Polymicro Technologies) to generate a 5 µm ID tip. Reversed phase particles (Aqua C18, 3 µm diameter, 125 Å pores, Phenomenex) were packed directly into the pulled column at 800 psi until reaching 12 cm. The MudPIT microcolumn was connected to an analytical column using a zero-dead volume union (Upchurch Scientific; IDEX Health & Science).

For the $^{15}N/^{14}N$ experiment, liquid chromatography (LC)-MS/MS analysis was performed using an Eksigent nano liquid chromatography pump and a Thermo LTQ-Orbitrap using an in-house built electrospray stage. Samples were run in a stepwise fashion, where each step corresponded to 0, 30 and 100% buffer C (500 mM ammonium acetate and 5% acetonitrile) being run for 5 minutes at the beginning of each gradient of buffer B (80% acetonitrile 0.1% formic acid). Electrospray was performed directly from the analytical column by applying the electrospray ionization voltage at a tee (150 mm ID, Upchurch Scientific). Electrospray directly from the LC column was executed at 2.5 kV with an inlet capillary temperature of 250˚C. Data-dependent acquisition of MS/MS spectra

with the LTQ-Orbitrap was performed with the following settings: MS/MS on the 5 most intense ions per precursor scan, 1 microscan, charge state 1 and >3 rejected; dynamic exclusion repeat count 2, repeat duration of 15 seconds; exclusion list size 50; and exclusion duration of 120 seconds. The same procedure was followed for the PIR-1::GFP IP in the *drh-3* mutant background, except for the following LTQ-Orbitrap settings: dynamic exclusion repeat count 1, repeat duration of 30 seconds; exclusion list size 150; and exclusion duration of 180 seconds. For the PIR-1::GFP IP experiment, LC-MS/MS analysis was performed using an Agilent 1100 pump and a Thermo LTQ using an in-house built electrospray stage. Samples were run with incremental steps consisting of 0, 20, 30, 40, 70, 100% buffer C being run for 5 minutes at the beginning of each gradient of buffer B. Data-dependent acquisition of MS/MS spectra with the LTQ was performed with the following settings: MS/MS on the 8 most intense ions per precursor scan, 1 microscan, dynamic exclusion repeat count 2, repeat duration of 30 seconds; exclusion list size 50; and exclusion duration of 60 seconds.

### MudPIT Data Analysis

Protein and peptide identification and protein quantitation were performed with Integrated Proteomics Pipeline - IP2 (Integrated Proteomics Applications http://www.integratedproteomics.com). Tandem mass spectra were extracted from raw files using RawExtract 1.9.9.2 (McDonald *et al.*, 2004) and were searched (both light and heavy) against the WormBase database (WP236) with reversed sequences using the ProLuCID program (Peng *et al.*, 2003; Xu *et al.*, 2006). The search space included all fully tryptic peptide candidates. Carbamidomethylation (+57.02146) of cysteine was considered a static modification. Peptide candidates were filtered using the DTASelect program, with the parameters -p 2 -y 1 --trypstat --pfp .01 -DM 10 --DB --dm -in (McDonald *et al.*, 2004; Tabb *et al.*, 2002). Quantitation was performed using the Census program (Park *et al.*, 2008).

### Gel Filtration Analysis

For gel filtration, 20 mg of young adult total protein extract in lysis/IP buffer were run in a Superose 6HR 10/30 column (Amersham Biosciences) using a Bio-Rad Biologic FPLC system. 500 µl fractions were collected and 25 µl aliquots were mixed with 5 µl of 6X protein sample buffer and denatured at 95˚C for 5 minutes. Every other fraction was

run on a Bio-Rad precast 4-15% SDS-polyacrylamide gel. Western blotting was performed as described above.

### RNase Treatments of IP Complexes

PIR-1::3xFlag was immunoprecipitated from 6.25 mg of total protein extract from mixed-stage worm populations. The washed beads were split into 4 tubes: one for immediate elution (no treatment control) and three for equilibration in PBS with 2 mM $MgCl_2$. Beads were then resuspended in 100 µl of PBS, 2 mM $MgCl_2$. Treatments were performed with 2 µl each of RNase T1 and RNase A, or 2 µl of RNase V1 (all from Ambion, Biochemistry Grade at 1 U/ml, 1 µg/ml and 0.1 U/ml, respectively). Reactions were incubated at 37˚C for 1 hour, after which beads were washed 2x 5 minutes with PBS, and 2x 5 minutes with PBST, all at room temperature. All samples were eluted with 2X protein sample buffer for SDS-PAGE and western blot analysis.

### SL1/SL2 PCR of *pir-1* and *sec-5* mRNAs

For each 25 µl PCR reaction, 0.5 µl of cDNA made from total RNA of mixed-stage wild-type animals were used. ExTaq DNA polymerase (Takara) was used with two different annealing temperatures (55 and 58˚C). PCRs were run on a 1.2% agarose/1X TAE gel and products that seemed specific were gel-extracted with the Qiagen Minelute kit. 4 µl of eluted material was used for TOPO cloning into pCR2.1-TOPO (Invitrogen) according to the manufacturer's instructions. Ligations were transformed into E. cloni 10G chemically competent *E. coli* cells (Lucigen) and grown on ampicillin/X-gal agar plates for blue/white selection. A few colonies of each product were selected for plasmid minipreparation and digested with *EcoR*I to confirm the presence of the insert. Positive clones were sequenced with M13 forward and reverse primers. Primers used in this experiment were O72-O75 (Appendix B).

### Generation of Start Site Mutant Constructs and Strains

The PIR-1(M1A), PIR-1(M29L) and PIR-1(M89L) mutants (assuming a 261 aa-long PIR-1 sequence) were generated by site-directed mutagenesis of pDONR-201-*pir-1::gfp* with primers O66-O71 as described in Chapter II Materials and Methods. *pir-1/mnC1*; unc-119; cxTi10882;* avr3x animals were injected with wild-type and mutant vectors, but no integrated lines were obtained. All experiments were thus performed with non-integrated lines exhibiting a high rate of transmission.

### Phosphatase Treatments of IP Complexes

PIR-1::3xFlag was immunoprecipitated from 2x 5 mg of total protein extract from young adult worms using buffer without EDTA. For each treatment condition, the equivalent of a 1 mg IP was used. Reaction volumes were of 100 µl. For CIP (NEB) treatment, beads were incubated with 1X NEB Buffer 3 (50 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT) and 2 µl of CIP with or without 1 M urea for 1 hour at 37˚C with constant shaking. For Lambda Protein Phosphatase (NEB) treatment, beads were incubated with 1X NEB Buffer for PMP (50 mM HEPES pH 7.5, 10 mM NaCl, 2 mM DTT, 0.01% Brij 35), 1 mM $MnCl_2$, 2 µl of enzyme with or without 1 M urea for 1 hour at 30˚C with constant shaking. Beads were then eluted at 95˚C for 5 minutes using 2X protein sample buffer and analyzed by SDS-PAGE/western blotting.

### Isolation of Nuclei and Chromatin

From frozen worm pellets, 2 volumes of Buffer A (10 mM Tris-HCl pH 8.0, 250 mM sucrose, 10 mM $MgCl_2$, 1 mM EGTA) without detergents and with protease and phosphatase inhibitors (Roche protease and phosphatase inhibitor cocktail tablets and 1 mM PMSF) were added and pellets were allowed to thaw. All steps were carried out on ice or 4˚C. For inputs, a fraction of this worm suspension was briefly spun down and Buffer A was replaced with 3 volumes of lysis buffer (25 mM HEPES-KOH pH 7.5, 150 mM KCl, 0.5% NP-40, 0.1% Triton X-100, protease and phosphatase inhibitors) and crushed with a metal dounce on ice with 100 strokes. Lysates were cleared by centrifugation at 10,000x *g* for 15 minutes and set aside for quantification and later use. For nuclear isolation worms were crushed with 30 strokes on ice, and debris were cleared briefly by centrifugation in 2 ml tubes at 500x *g* for 1 minute. This step was repeated with the resulting supernatant, after which it was again transferred to fresh 1.5 ml tubes to spin down nuclei at 4,000x *g* for 5 minutes. Nuclear pellets were gently resuspended in at least 10 volumes of Buffer A with inhibitors. 1 ml of nuclear suspension was gently overlayed over 10 ml of sucrose cushion solution (10 mM Tris-HCl pH 8.0, 1 M sucrose, 10 mM EDTA) in a 15 ml conical tube and the interphase was gently disrupted by swirling a pipette tip to create a gradient (no more than 1 cm into the sucrose solution). This was repeated for the entire volume of each lysate. These tubes were centrifuged at 3,200x *g* for 1 to 1.5 hours, until a uniform white pellet corresponding to nuclei accumulated at the bottom of the tubes. The supernatant was aspirated and nuclei were carefully resuspended in Buffer A with inhibitors (about 10 volumes). A second sucrose flotation step was carried out in 1.5 ml tubes by overlaying

150 µl of resuspended nuclei onto 1 ml of sucrose solution, again gently disrupting the interphase. The preparations were spun at 20,000x *g* for 10 minutes, and the resulting pellets were resuspended in 10 volumes of lysis buffer with detergents and inhibitors (as above for inputs) and crushed with a dounce with 150 strokes to completely disrupt nuclei. This suspension was cleared at 8,000x *g* for 5 minutes and the supernatant was kept for protein quantification and use as a nuclear extract for IP. Pure chromatin is gray when pelleted. The presence of white patches indicates that intact nuclei remain. When this occurred, the pellet was resuspended in lysis buffer with detergent and crushed a further 100 times until the pellet was uniformly gray. The pellet was then washed once with 10 mM Tris-HCl pH 7.5 with 0.1% Triton X-100 and another two times with 10 mM Tris-Cl pH 7.5 (with 8,000x *g* centrifugations in between and the last one at 20,000x *g* for 5 minutes). Pellets were weighed and resuspended in 9 volumes (for a 1:10 dilution) of protein sample buffer for a 1X final concentration and such that the chromatin concentration was between approximately 50-100 µg/µl. Before loading onto a denaturing protein gel, the chromatin was heated to 95˚C for 5 minutes.

### RNAi of *ama-1*

Silencing of *ama-1* was achieved by feeding animals with dsRNA-expressing HT115 *E. coli* from the Ahringer feeding RNAi library. Cultures were prepared and induced as described in Chapter II. Animals were grown on OP50 *E. coli* until they reached the L2 stage, after which they were collected, washed and transferred to large plates with *ama-1* HT115 at a density of ~100,000 worms per plate. After ~2 days at 20˚C, all animals were arrested as sterile adults with no embryos. One day later they were collected and frozen for protein and chromatin extraction. Placing animals on RNAi food as L1s gave rise to an L3 arrest. When they were placed as L3s, they arrested at a range of sterile to gravid adults that laid some dead eggs, and when RNAi started at the L4 stage, all animals arrested as gravid adults that laid many eggs, most of which did not hatch (the ones that did hatch exhibited an L2-L3 arrest).

### Cloning of Human PIR1 and Immunoprecipitations from Transfected Human Cells

The human *PIR1* (*DUSP11*) cDNA sequence was amplified from the pOTB7 IMAGE cDNA clone construct with primers carrying terminal *Xba*I and *Not*I sites (O76 and O77) using iProof Taq DNA polymerase (Bio-Rad). The product was digested with

*Xba*I and *Not*I and column-purified (Qiagen gel purification kit). The p3xFLAG-Myc-CMV-26 vector (Sigma) was digested with the same enzymes, dephosphorylated with CIP and gel purified. After ligation with T4 DNA ligase (NEB), the reaction was transformed and the plasmids from a few positive clones were digested with *Xba*I and *Not*I to check for the release of the insert. Two clones were sent out for sequencing of the *PIR1* cDNA. The full-length GST-Dicer fusion construct is cloned into the pDEST27 vector (Life Technologies) and was a kind gift from the Filipowicz laboratory at the Friedrich Miescher Institute in Switzerland (described in Tahbaz *et al.*, 2004).

HEK293T cells in 6 cm plates at 80-90% confluency were transfected with 4-8 µg of each vector using Lipofectamine 2000 (Invitrogen) according to the manufacturer's specifications. Transfected cells were collected after 24 hours for protein extraction. All of the following steps were done on ice or at 4˚C. After cells were washed with 1X PBS, 1 ml of lysis buffer was added (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.5% NP-40, 2.5 $MgCl_2$, 2.5% glycerol, supplemented with 1 mM DTT, protease and phosphatase inhibitor cocktail tablets (Roche, 1 tablet of each per 10 ml)). Cells were scraped of the plates, transferred to 1.5 ml microtubes and lysed by passage through a 20 gauge needle 15 times followed by rotation for 10 minutes to completely release cell contents. Extracts were cleared by centrifugation at 10,000x *g* and quantified. For inputs, 250 µg of protein were acetone-precipitated and resuspended in 1X protein sample buffer. Per IP, 1 mg of extract was used in 1 ml of lysis buffer with inhibitors. For Flag IPs, 25 µl of M2 anti-Flag agarose beads (Sigma) were used. For Dicer IPs, 5 µg of anti-Dicer [13D6] monoclonal antibody were used (Abcam, ChIP-grade) and captured with 20 µl of Protein A/G PLUS-Agarose beads (Santa Cruz Biotechnologies). After incubating for 2 hours, beads were washed 4x 5 minutes with IP buffer and eluted with 25 µl of 2X protein sample buffer at 95˚C for 5 minutes. 50 µg of inputs and half of each 1 mg IP were loaded onto a 4-15% gradient polyacrylamide gel (Bio-Rad). Western blot analysis was performed as described above. Dicer was detected with the same antibody used for IP at a 1:250 dilution; PIR1 was detected with a chicken affinity purified anti-DUSP11 at a 1:250 dilution (USBiological); actin was detected with a mouse anti-actin at 1:2500. Secondary antibodies were HRP-conjugated anti-mouse (1:5000, Santa Cruz Biotechnologies) and HRP-conjugated rabbit anti-chicken IgY (1:5000, Jackson ImmunoResearch).

# CHAPTER IV

Small RNA and mRNA Profiling

of *pir-1* Mutant Animals

## INTRODUCTION

At the beginning of this project, the known repertoire and genetic requirements of endo-sRNAs was very limited and the ERI pathway was just beginning to be defined (Duchaine *et al.*, 2006). We therefore started exploring the possibility of sRNA defects in *pir-1* mutants through a northern blotting-based candidate approach using a limited number of endo-sRNAs. However, it was not until high-throughput sequencing (deep-sequencing) technologies became available that we were able to address this problem at a genome-wide level. We thus set out to conduct a thorough comparative analysis between the sRNA populations of wild-type and *pir-1* mutant animals.

In our initial analyses of *pir-1* mutants, we did not observe depletion of sRNA species such as miRNAs, 21U-RNAs (piRNAs), ERGO-1-dependent 26G-RNAs, and ERI-independent 22G-RNAs (refer to Table 1.1, page 29 for an overview of sRNA pathways). Unexpectedly, we did find that both *pir-1(tm1496)* and *pir-1(tm3198)* deletion mutants could still produce functional siRNAs in response to exogenous dsRNA triggers, challenging the initial model for PIR-1 function in RNAi. Through deep-sequencing we found that the sRNAs most clearly affected by the loss of PIR-1 belonged to the ALG-3/4-dependent 26G-RNA class. These endo-sRNAs regulate sperm differentiation in both the hermaphrodite germline (preceding female gamete production) and in the male germline (Han *et al.*, 2009; Conine *et al.*, 2010; Conine *et al.*, 2013). This result is in agreement with the spermatogenesis defect of the *pir-1* mutant (Chapter II) and the interaction of the protein with the ERI complex (Chapter III), which generates all 26G-RNAs. Additionally we showed that recombinant PIR-1 is able to convert 5'-triphosphorylated RNAs to 5'-monophosphorylated RNAs, and that this property requires the active-site cysteine. Since the molecular events that lead to the production of mature 26G-RNAs are still largely undefined, these discoveries allowed us to refine the current working model of how this process occurs. Most importantly, this was the first concrete biological function attributed to PIR-1 in any organism.

## RESULTS

### Developmentally Arrested *pir-1* Animals Are Competent for Exogenously Triggered RNAi

In its initial characterization, *pir-1(tm1496)* mutants were reported to be resistant to exogenously triggered RNAi due to a lack of accumulation of siRNAs (Duchaine *et al.*, 2006). In trying to confirm and expand these observations, we discovered that *pir-1(tm1496)* animals were actually able to accumulate exogenously triggered siRNAs. In these experiments, we induced RNAi by feeding animals with bacteria expressing dsRNA targeting the *sel-1* (*S*uppressor/*E*nhancer of *L*IN-12 1) mRNA, expressed in both somatic and germline cells, and from which exceptionally abundant RNAi-induced 22G-RNAs are produced. We used *pir-1(tm1496)* homozygote animals that had been counter-selected on ivermectin for seven days to avoid any maternally loaded *pir-1* product. After 24 hours of feeding, not only were we able to detect *sel-1* siRNAs but also showed that their levels increased with longer exposure to the trigger dsRNA (Fig. 4.1A). The RNAi-defective mutant *rde-1(ne300)* was used as a negative control, as it is unable to accumulate Dicer-dependent primary siRNAs and resulting 22G-RNA RdRP products. Additionally, in a *pir-1(1496); rrf-1* double mutant, we could still detect a time-dependent accumulation of *sel-1* siRNAs, albeit at lower levels than in the *pir-1* single mutant (Fig. 4.1B). RRF-1 is the main 22G-RNA RdRP operating in somatic cells, whereas in the germline it acts redundantly with the EGO-1 RdRP. The fact that we could still detect *sel-1* siRNAs in the absence of RRF-1 means that *pir-1* mutant animals are able to mount an RNAi response in both the soma and the germline. It should be noted that the ivermectin resistant strain (avr3x DA1316) we used as a positive control along with N2, exhibited a defect in siRNA accumulation which we could not explain at the time. We describe the resolution of this issue later in this section.

We then targeted the body wall muscle-specific *unc-22* transcript. Expression of *unc-22* increases throughout development, and RNAi against it leads to a progressively stronger twitching phenotype that culminates in almost complete paralysis of the animals. In contrast to wild-type N2 animals, when arrested *pir-1(tm1496)* homozygotes were exposed to *unc-22* dsRNA-expressing bacteria for 48 hours they did not manifest any of the *unc-22*-associated phenotypes (not shown), despite the accumulation of *unc-22* siRNAs (Fig. 4.1C). Avr3x worms twitched with the same treatment, although the onset of the phenotype was delayed relative to N2, and, paradoxically, siRNAs were not detected. From

the same samples we measured *unc-22* mRNA levels by qRT-PCR and found that, similar to avr3x alone, *pir-1* mutants had a ~50% reduction in mRNA, not far from the 75% reduction in N2. This indicated that, under our experimental conditions, the *pir-1(tm1496)* mutant is able to produce functional siRNAs.

We have identified reasons for the discrepancy between the results reported in Duchaine *et al.* (2006) and the aforementioned results, regarding the production of exo-siRNAs. First, northern analysis of *sel-1* siRNAs was conducted in animals that had been exposed to *sel-1* dsRNA bacteria for 12 hours, which we found to not always be sufficient to generate robust amounts of siRNAs relative to wild-type N2 animals. Second, the inability of *pir-1(tm1496)* animals to twitch upon *unc-22* RNAi was erroneously interpreted as an inability to produce *unc-22* siRNAs. UNC-22 protein associates with myosin and is required for regulation of actomyosin contraction-relaxation and normal muscle morphology (Moerman *et al.*, 1988). We speculate that the inability of *pir-1* mutants to acquire the twitching and paralysis phenotype may reflect a lowered need for UNC-22, allowing them to maintain proper muscle function with existing UNC-22 in the absence of further growth (in other words the turnover of this protein at the arrested stage is low).

Assuming that PIR-1 functions as an RNA 5' phosphatase, we next asked whether different types of endogenous sRNAs exhibited changes in abundance or 5'-phosphorylation state in the *pir-1(tm1496)* mutant background (Fig. 4.1D). For this, we treated total RNA depleted of high MW RNAs with Terminator 5' phosphate-dependent exonuclease, and analyzed it by northern blotting (Fig. 4.1D). This enzyme destroys single-stranded RNA (ssRNA) bearing 5'-hydroxyl (5'-OH) or 5'-monophosphate (5'-P) ends, but not 5'-PPP ends. 5'-PPP 22G-RNAs were unaffected by Terminator digestion, regardless of the source of trigger (exogenously by *sel-1* dsRNA or endogenously by abundant ERI/ERGO-1-dependent 26G-RNA-1 and 26G-RNA-263). miRNAs, 26G-RNAs and 21U-RNAs, all normally carrying a 5'-P, were destroyed by the treatment in both wild-type and *pir-1* backgrounds, suggesting that these sRNA species do not require PIR-1 for proper 5'-end processing. Additionally, non-ERI endogenous 22G-RNA species such as those targeting *F37D6.3* and *bath-20* in the germline or *Y47H10A.5* in the soma (Gu *et al.*, 2009), also kept intact 5'-PPP ends in *pir-1* animals. Importantly, the levels of most sRNAs tested did not change considerably in *pir-1* mutant animals relative to wild-type N2. The only exceptions were 26G-RNA-1 and 263, which were reduced, and *Y47H10A.5* 22G-RNAs, which were dramatically increased. The lowered accumulation of 26G-RNAs in

*pir-1* is explained by the fact that ERGO-1 26G-RNAs are synthesized predominately during oogenesis and embryogenesis (Han *et al.*, 2009; Pavelec *et al.*, 2009; Gent *et al.*, 2010; Vasale *et al.*, 2010), processes that do not occur in *pir-1* animals. *Y47H10A.5* 22G-RNAs are known to steadily accumulate with age (Weifeng Gu, unpublished observation), so we attribute the discrepancy in siRNA levels to the age difference between N2 control animals (2.5-day old L4s) and *pir-1* (seven days old). The strong accumulation of *Y47H10A.5* siRNAs in *pir-1* arrested animals further implies that PIR-1 is not required for 22G-RNA synthesis.

After obtaining the new *pir-1(tm3198)* allele, we proceeded to validate the results we obtained with *pir-1(tm1496)* (Fig. 4.2). Although arrested *pir-1(tm3198)* mutants never twitched (either through feeding or injection of *unc-22* dsRNA), they still produced abundant *unc-22* 22G-RNAs, which led to an almost wild-type decrease in *unc-22* mRNA (Fig. 4.2A). In this experiment we included seven-day counter-selected *eft-3(q145); avr3x* homozygote animals derived from genetically-balanced wild-type mothers. *eft-3* encodes an essential translation elongation factor, and homozygote animals grow to become adults without germline tissue (accordingly, they do not express *F37D6.3* 22G-RNAs). These animals served as a control for aging and exposure to ivermectin to more fairly compare the accumulation of *unc-22* siRNAs. In comparison to the level of *unc-22* siRNAs in the *eft-3* mutant, *pir-1(tm3198)* animals showed only a modest reduction in the accumulation of siRNAs.

Since the use of an isogenic avr3x control for further experiments using *pir-1(tm3198);* avr3x animals was imperative, we had to address the defects in exo- and endo-siRNA accumulation of the original avr3x triple mutant DA1316 (*avr-14(ad1305); glc-1(pk54::Tc1) avr-15(ad1051)*). The fact that in this strain dsRNA feeding led to target downregulation, and that in some northern blots we were able to detect low levels of *sel-1* siRNAs (not shown) told us that the RNAi pathway was not completely compromised. We outcrossed the original strain against N2 animals six times and resegregated the avr3x mutant combination from the *pir-1(tm1496)/mnC1*;* avr3x strain. Additionally, we obtained control strains which included the *avr-14(ad1305); glc-1(pk54::Tc1)* double mutant, the *avr-15(ad1051)* mutant and the *eat-2(ad1113)* mutant. The *eat-2* and *avr-15* genes encode ligand-gated ion channel subunits which are required for normal rates of pharyngeal pumping, and when defective interfere with feeding (Dent *et al.*, 1997; Raizen *et al.*, 1995). Since avr3x mutants exhibit a two to three hour delay in development relative to N2 at 20˚C likely due to slower feeding, we wondered whether this could also contribute

to decreased RNAi efficiency. After feeding all of these strains *unc-22* dsRNA bacteria, only the original avr3x DA1316 retained the exo- and endo-RNAi defects (Fig. 4.2B). Even though the feeding mutants *eat-2* and *avr-15* still showed a delayed onset of the *unc-22* RNAi phenotypes compared to N2 or *avr-14; glc-1* doubles, they produced abundant *unc-22* siRNAs. We concluded that the original strain must carry one or more unidentified mutations that affect the accumulation of some types of sRNA. Henceforth, we used the outcrossed avr3x strain as the wild-type control for all experiments involving ivermectin counter-selection.

With this new control in hand, we performed the *unc-22* RNAi experiment in triplicate with *pir-1(tm3198)* animals. We measured the relative amount of *unc-22* siRNAs and the relative amount of *unc-22* mRNA in order to correlate them (Fig. 4.2C). We concluded that *pir-1(tm3198)* generated wild-type levels of *unc-22* siRNAs that led to silencing of the cognate mRNA, although at a lower efficiency than in N2 and avr3x (~50% *versus* ~80%). Moreover, we targeted the germline mRNA *pos-1* by feeding RNAi and registered a dramatic drop in the relative level of this transcript in the *pir-1* mutant (Fig. 4.2D). Lastly, we crossed a nuclear GFP reporter expressed in the lateral seam cells (16 across each side of L4s and adults) into *the pir-1(tm3198)* background (non-avr3x). Although aged control *pir-1* worms did not express GFP as brightly as wild-type, both strains exhibited marked silencing of GFP after being fed bacteria expressing *gfp* dsRNA (Fig. 4.3). Taken together, these results suggest that PIR-1 is not required for exogenously triggered RNAi, nor for the production of certain endogenous 22G-RNAs.

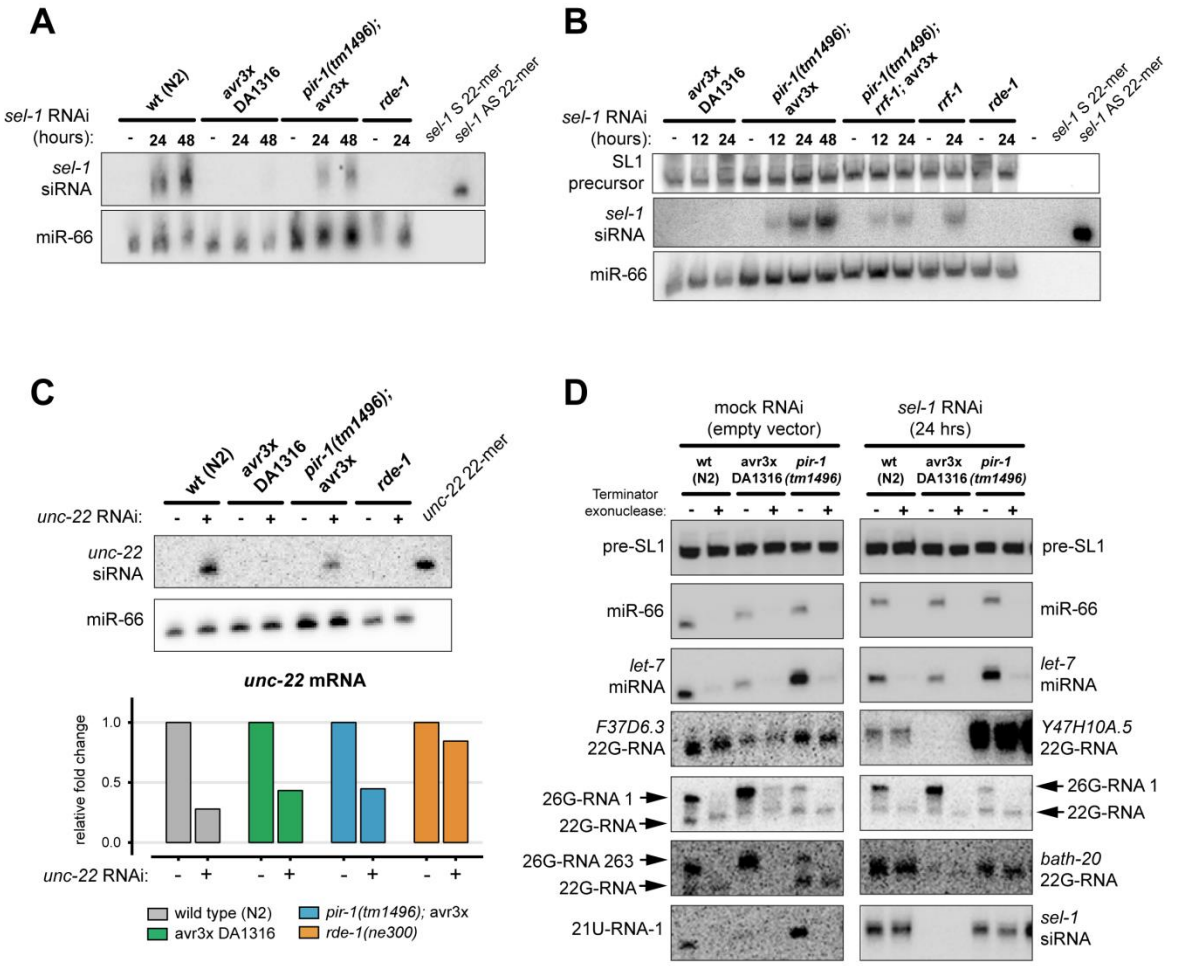



**Figure 4.1. *pir-1(tm1496)* arrested animals can mount an RNAi response and are not generally deficient in endogenous small RNAs. (A)** Northern blot of *sel-1* siRNAs after being fed bacteria expressing *sel-1* dsRNA. The miR-66 miRNA, expressed at a constant level throughout development, was used as a loading control. Sense ('S') or antisense ('AS') *sel-1* RNA oligos were run as probe specificity controls. **(B)** Northern blot of *sel-1* siRNAs in the *rrf-1(pk1417)* mutant background to determine germline RNAi contribution. The SL1 trans-splicing precursor (~100 nt) was used as an additional loading control. **(C)** Northern blot of *unc-22* siRNAs and corresponding *unc-22* mRNA levels measured by qRT-PCR and normalized to *GAPDH* (*gpd-2*). **(D)** Northern blots of sRNAs before and after treatment with Terminator 5' phosphate-dependent exonuclease which cannot digest 5'-capped, 5'-di- or 5'-triphosphorylated RNA. *pir-1* animals were counter-selected for seven days on ivermectin prior to feeding with dsRNA-expressing bacteria. The remaining strains were grown for 10-12 hours, then transferred to RNAi plates so as to be harvested as L4s (after 12 hours of RNAi), or grown since L1s on RNAi plates to be collected as L4s (after 24 hr) or young adults (after 48 hr). In (A) and (B) total RNA was used. In the remaining blots sRNA fractions were used. All no-RNAi controls used induced bacteria carrying the empty double T7-promoter vector.

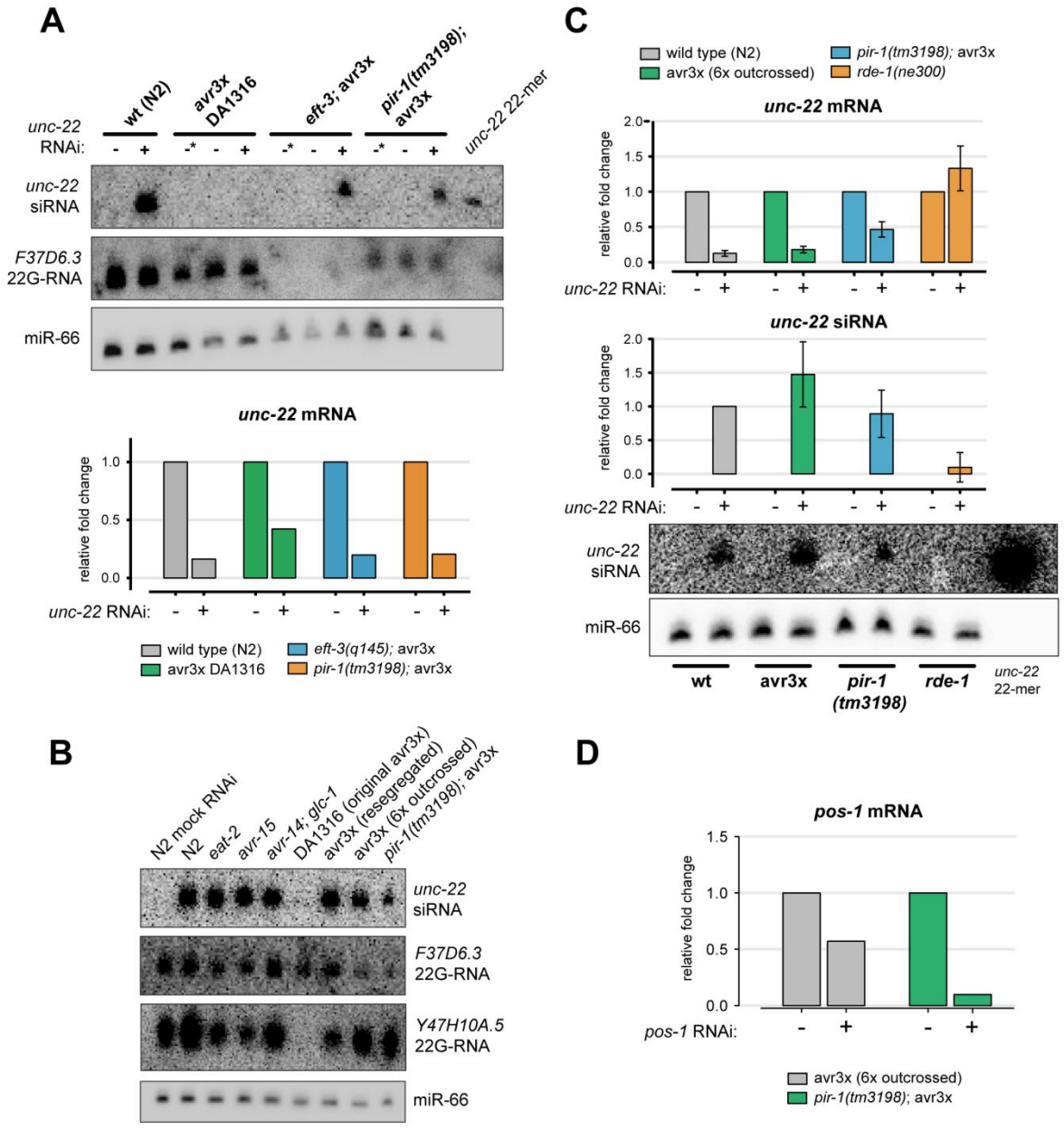



**Figure 4.2. *pir-1(tm3198)* arrested animals also mount an effective RNAi response. (A)** Northern blot of *unc-22* siRNAs and corresponding *unc-22* mRNA levels measured by qRT-PCR and normalized to *GAPDH* (*gpd-2*). miR-66 served as the sRNA loading control. Asterisks indicate growth on OP50 *E. coli*. In the other conditions animals were transferred to either mock RNAi or *unc-22* RNAi food plates and grown for an additional 24 hours. N2 and avr3x animals were harvested as non-gravid young adults, ivermectin counter-selected *eft-3(q145)* animals were harvested as five-day old sterile adults, and *pir-1* animals were harvested as five-day old arrested L4s and young adults. **(B)** Northern blot of sRNAs from outcrossed and resegregated avr3x strains to restore wild-type sRNA accumulation. *pir-1* animals were counter-selected for seven-days. The other strains were transferred to RNAi food after 24 hours on OP50 and harvested as non-gravid young adults. **(C)** qRT-PCR measurements of *unc-22* mRNA normalized to *act-3* (*Act*in 3) mRNA levels and siRNA levels normalized to miR-66. Experiments were performed in triplicate. Error bars show the standard error of the mean (SEM). siRNA northern intensities were calculated using ImageQuant software (GE healthcare). **(D)** qRT-PCR of germline *pos-1* mRNA normalized to actin (*act-3*) mRNA levels after *pos-1* RNAi in seven-day counter-selected *pir-1* animals. Small RNA

fractions were used for all northern blots. All mock RNAi samples used induced bacteria carrying the empty double T7-promoter vector. RNAi for all experiments occurred over a 24-hour period.

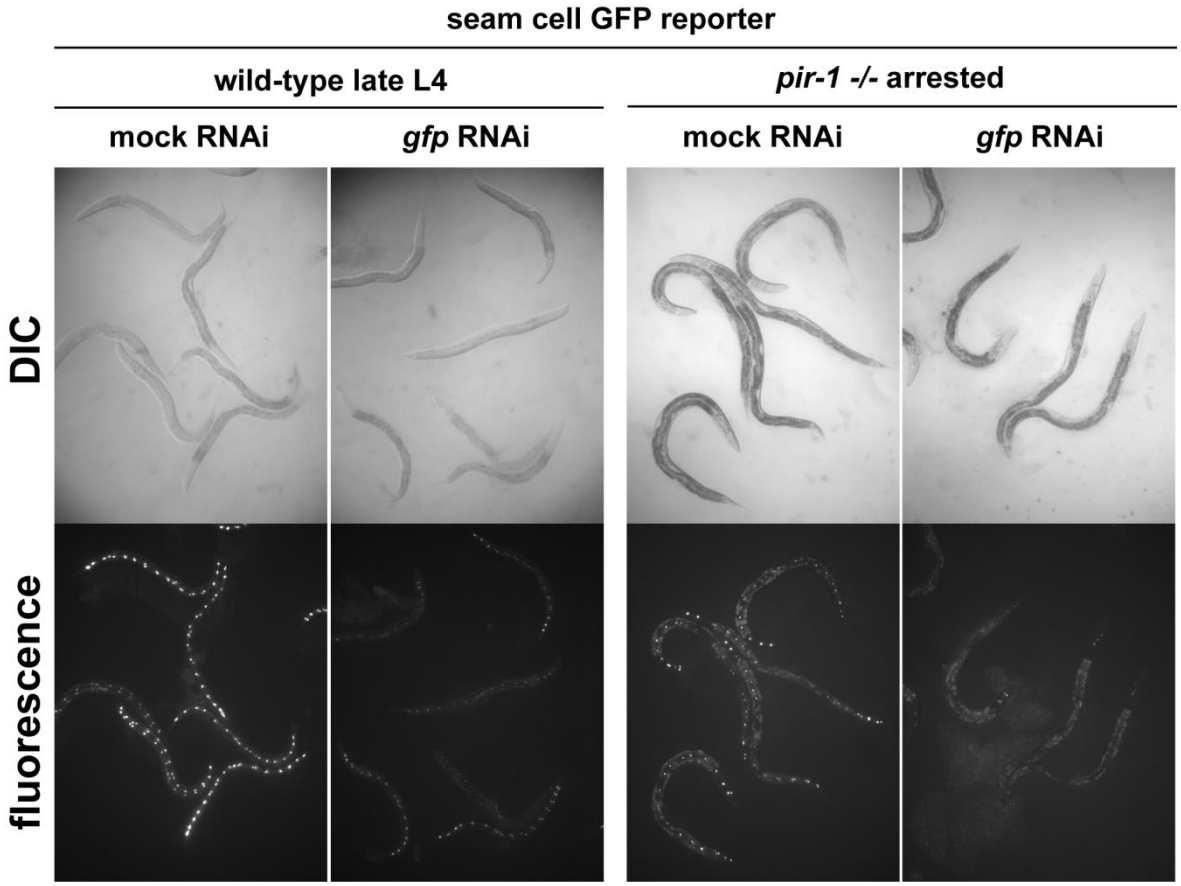



**Figure 4.3. *pir-1(tm3198)* arrested animals can silence a GFP reporter upon *gfp* RNAi.** Wild-type or *pir-1* mutant animals expressing nuclear GFP from a seam cell promoter were fed *gfp* dsRNA-expressing bacteria for 24 hours, followed by fluorescence microscopy to assess GFP protein knockdown. All images were acquired with identical exposure settings. *pir-1* animals were at least seven days old when placed on RNAi plates.

## *pir-1* Mutant Animals Are Not Deficient in miRNAs or 21U-RNAs but Express Reduced Levels of Germline 22G-RNAs

In order to globally study the effect that loss of PIR-1 could have on endogenous sRNA populations, we cloned and deep-sequenced 18-30 nt RNAs from ivermectin counter-selected *pir-1* animals. We adopted a cloning strategy that allows the cloning of most sRNA classes, regardless of their 5'-phosphorylation state. For this, RNA is treated with Tobacco Acid Pyrophosphatase (TAP) prior to RNA size selection and 5' linker ligation. This enzyme converts 5'-PPP and capped 5' ends to 5'-P, and does not alter

molecules that were originally 5'-P. We refer to this method as 5'-independent cloning or TAP cloning. To exclude differences due to age and ivermectin exposure period, we sequenced sRNAs from *pir-1* arrested animals counter-selected for three, seven and ten days. As a wild-type control we used avr3x animals grown on ivermectin until the population consisted of a range of L4 larvae to young adults (ivermectin affects the population's synchronicity).

Taking the ten-day counter-selection animals as an example, we compared levels of the main categories of endo-sRNAs with those of the wild-type control (Fig. 4.4), using the total number of genome-matching sRNA reads as a normalization standard. These reads excluded structural reads, which we assume to be degradation fragments of ncRNAs (non-coding), tRNAs, rRNAs, snoRNAs (small nucleolar), snRNAs (small nuclear) and scRNAs (small cytoplasmic). The scatter plots provide information about the relative abundance of each type of sRNA by displaying the number of normalized reads in reads per million (rpm; Fig. 4.4A). miRNAs, which are predominately expressed in the soma and represent the most abundant sRNA class, were slightly upregulated in *pir-1* animals regardless of age (Fig. 4.4B and C, orange bars). The upregulation likely arises from a lower germline-to-soma ratio compared to wild type. Germline-expressed 21U-RNAs decreased slightly in *pir-1* animals, consistent with the small germlines of the mutants. Moreover, western blots of DCR-1 (which generates mature miRNAs) and PRG-1 (the PIWI Argonaute that interacts with 21U-RNAs) revealed that these proteins are expressed in *pir-1* at wild-type levels (Fig. 4.4D). From these results we conclude that PIR-1 is not required for the accumulation of miRNAs or 21U-RNAs.

Regarding all sRNAs antisense to protein-coding sequences (targeting over 9,000 genes, nearly 50% of all *C. elegans* genes), we observed a general decrease in *pir-1* animals (Fig. 4.4B and C, gray bars). In order to better understand this decrease, we considered 22G-RNA subpopulations that belong to the two major pathways according to which Argonautes they interact with. The CSR-1 pathway is required for accurate germline gene expression, while the WAGO pathway, which includes 12 different Argonautes, is involved in silencing repetitive elements, transposons as well as various somatic- and germline-expressed mRNAs (refer to Chapter I for details). These 22G-RNA populations were experimentally defined in our laboratory, as explained in the Materials and Methods section (page 215). WAGO-class 22G-RNAs were reduced in the *pir-1* mutant, with at least 50% of targets exhibiting a two-fold or higher downregulation of 22G-RNAs compared to wild type (Fig. 4.4, dark blue bars). CSR-1-class 22G-RNAs were reduced

more pronouncedly, with at least 75% of targets exhibiting a downregulation upwards of two-fold (Fig. 4.4, red bars). As most 22G-RNAs derive from the germline, we attribute this difference at least in part to the lower germline-to-soma ratio of the mutant animals. However, because germline 21U-RNA levels are not significantly affected in *pir-1* mutants, we cannot exclude that there are additional reasons for WAGO- and CSR-1-22G-RNA depletion.

In agreement with loss of 22G-RNAs in *pir-1* animals due to germline immaturity, ivermectin counter-selected *dcr-1* deletion mutants, which arrest as adults with a large germline, do not exhibit reductions in these WAGO- or CSR-1-dependent 22G-RNAs (Fig. 4.5B). In contrast, these animals exhibited a marked decrease in miRNAs, in agreement with the role of DCR-1 in miRNA processing, and also reinforcing the idea that PIR-1 is not required for miRNA biogenesis. Lastly, we analyzed sRNA populations in arrested *pir-1; dcr-1* double mutants (Fig. 4.5C), which are phenotypically indistinguishable from *pir-1* single mutants. Not surprisingly, the most noticeable difference relative to the *pir-1* single mutant, was a downregulation of miRNAs, albeit less pronounced than in *dcr-1* single mutants.

The WAGO-class 22G-RNA population is very complex in terms of the variety of functions performed along development, in distinct tissues and cell compartments. This makes it difficult to find direct causes for the deregulation of 22G-RNAs as a whole. However, when we focused on WAGO-dependent 22G-RNAs targeting transposons only, the differences in their levels in *pir-1* compared to wild type were less pronounced than when considering the expanded WAGO 22G-RNA population (Fig. 4.5D and E). This led us to conclude that the silencing of transposons by the WAGO pathway in the germline of *pir-1* animals is not compromised. We extended our analysis to other populations of sRNAs targeting specific types of transcripts, namely non-coding RNA, pseudogenes, repetitive elements and introns (Fig. 4.5F). In order to provide an idea of the proportion of each sRNA population in the entire set of cloned RNAs, we displayed non-normalized reads from each category as a percentage of the total reads matching the genome (minus nonstructural reads). miRNAs, by far the most abundant class, were cloned in *pir-1* at a higher proportion as a result of the decrease in sRNAs antisense to coding RNA (including WAGO, CSR-1, and ALG-3/4 pathway sRNAs; gray line) as reflected in the analysis of normalized reads in the box-and-whisker plots. All of the remaining classes, including 21U-RNAs (in green), exhibited no change in cloning frequency between wild-type and mutant animals.

Importantly, since CSR-1-targeted mRNA (enriched in germline transcripts) or WAGO mRNA-targets are not generally downregulated in *pir-1* mutant animals, we cannot attribute the 22G-RNA reduction to a decrease in the amount of template molecules (mRNA-sequencing results, Fig. 4.7A and B, page 179). Furthermore, CSR-1 and some WAGO Argonautes concentrate in germline P granules, and loss of CSR-1 leads to a dramatic disassociation of P granules from germline nuclei (Gu *et al.*, 2009; Claycomb *et al.*, 2009). Since *pir-1* mutants have intact P granules, as shown in Chapter II (Fig. 2.3C, page 77), we cannot justify the reduction of these 22G-RNA classes with the disassembly of P granules. While we cannot completely rule out that the downregulation of WAGO- and CSR-1-pathway 22G-RNAs indirectly arise from the germline immaturity of the mutants, our results suggest that PIR-1 plays a role in promoting the accumulation of these two classes of 22G-RNAs. Unfortunately, the inability to silence *pir-1* by RNAi confined us to a developmental stage where phenotypes associated with defects in the CSR-1 and WAGO pathways cannot be adequately studied.

### *pir-1* Mutant Animals Fail to Accumulate Spermatogenesis-Associated ALG-3/4-Class 26G-RNAs

We next considered ERI-class sRNAs, which we subdivided into ERGO-1 and ALG-3/4 classes and further into primary 26G-RNAs and corresponding secondary 22-RNAs. Regarding ERGO-1-class 22G-RNAs, we did not observe dramatic changes in their levels between *pir-1* and wild-type animals (Fig. 4.4 and 4.5, magenta bars). These 22G-RNAs are synthesized primarily during postembryonic development in somatic tissues, triggered by 26G-RNAs produced during oogenesis and embryogenesis (Vasale *et al.*, 2010; Fischer *et al.*, 2011). Consistent with ERGO-1 26G-RNA production starting at oogenesis, *pir-1* animals, which almost never reach oogenesis, as well as *dcr-1* animals, which have severe oogenesis and embryogenesis defects, express lower levels of ERGO-1 26G-RNAs (Fig. 4.4 and 4.5 rightmost magenta bars). We infer that the ERGO-1 26G-RNAs in the mutant homozygotes derive from synthesis during oogenesis in the heterozygous mother, and later during embryogenesis using maternally-derived PIR-1 or DCR-1. Interestingly, in the *pir-1; dcr-1* double mutant, ERGO-1 26G-RNAs exhibited the strongest depletion of any of the samples analyzed (Fig. 4.5C). Since these sRNAs require Dicer, it is tempting to interpret this change as a synergy between the absence of both PIR-1 and DCR-1, perhaps hinting that some ERGO-1 26G-RNAs may be synthesized prior to oogenesis. Regardless of this, we cannot make a definitive conclusion regarding a role for

PIR-1 in the production of ERGO-1 26G-RNAs. Addressing this issue will require RNA analysis at the peak developmental stages for ERGO-1 26G-RNA biogenesis. When we looked at ALG-3/4-class 22G-RNAs, however, we observed a decrease similar to what we saw for CSR-1 22G-RNAs in *pir-1* arrested mutants (Fig. 4.4 and 4.5A, left light blue bars). Interestingly, ALG-3/4 26G-RNAs were almost completely depleted in *pir-1* mutant animals regardless of their age (Fig. 4.4 and 4.5A, rightmost bars), strongly suggesting that the reduction in 22G-RNAs could be a direct consequence of the absence of trigger sRNAs and that PIR-1 could promote the accumulation of spermatogenesis-associated 26G-RNAs.

In contrast to *pir-1* mutant animals, the *dcr-1* mutant expressed wild-type levels of ALG-3/4 22G-RNAs (Fig. 4.5B). The lack of ALG-3/4 26G-RNAs in *dcr-1* mutant animals is explained by the fact that the animals were collected as mature adults where spermatogenesis was no longer occurring. Since *dcr-1* homozygotes reach adulthood due to a strong maternal load of DCR-1 protein (Duchaine *et al.*, 2006), we presume that ALG-3/4 22G-RNAs were abundantly triggered during spermatogenesis by ALG-3/4 26G-RNAs produced by inherited DCR-1. In *pir-1* arrested animals, because the amount of 26G-RNAs may have never reached wild-type levels, the triggering of 22G-RNAs was consequently much weaker.

In contrast to spermatogenesis in hermaphrodites, which is confined to a relatively short time frame preceding female gamete production, males continuously undergo spermatogenesis. ALG-3/4 26G-RNAs against a greater variety of transcripts can therefore be detected in adult males, at all stages of spermatogenesis – and associated sRNAs – are simultaneously captured. We found that males also exhibit a strong depletion of spermatogenesis-associated 26G-RNAs in the absence of PIR-1 (Fig. 4.6A and B). The reduction was not as pronounced as in hermaphrodites, possibly because the mRNA templates are expressed at higher levels, which in turn leads to a more abundant synthesis of 26G-RNAs while maternal PIR-1 is still available. We obtained this result using either the standard TAP-cloning method (5'-independent) or a 5'-dependent direct cloning (no enzymatic treatment), which excludes all 5'-PPP species (*i.e.*, 22G-RNAs). The fact that we could still abundantly clone ALG-3/4 26G-RNAs with the direct method implied that they are still predominately 5'-monophosphorylated in the *pir-1* mutant germline (Fig. 4.6C). Unexpectedly, 21U-RNAs became enriched in *pir-1* males regardless of cloning method and normalization standard (total nonstructural reads for TAP cloning and miRNAs for direct cloning). Since male germlines only produce sperm, the reduction in ALG-3/4 26G-RNAs may have a more widespread effect on deregulation of germline gene

expression, and the rise in 21U-RNA expression may be a consequence of that. The fact that 21U-RNAs are actually enriched, however, does increase our confidence that PIR-1 is not required for their biogenesis. Additionally, 22G-RNAs targeting protein-coding transcripts are also significantly reduced in *pir-1* males (Fig. 4.6B, gray bar), once again suggesting a role for PIR-1 in promoting their synthesis.

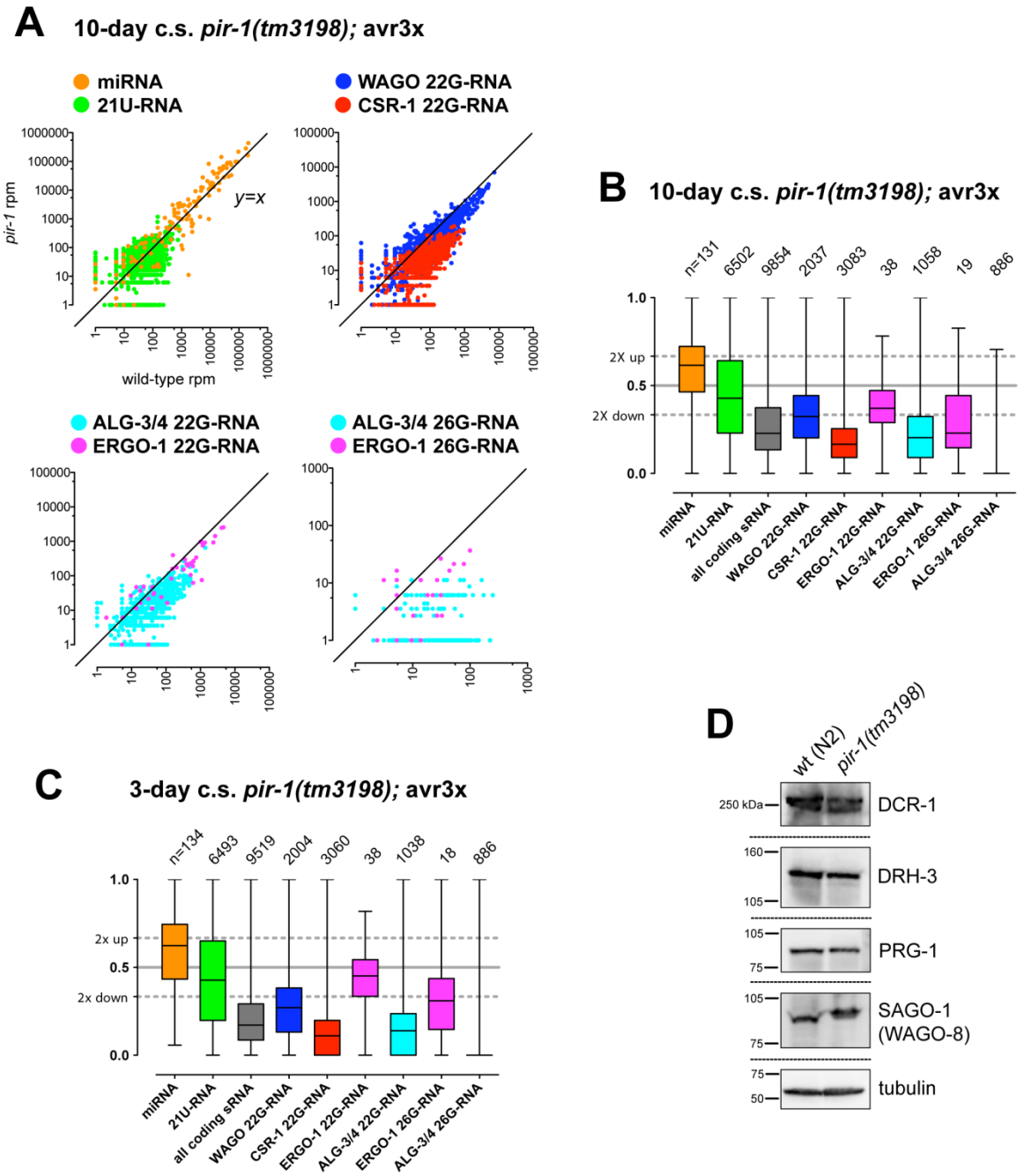



**Figure 4.4. Deep-sequencing analysis reveals that specific classes of germline endogenous small RNAs are downregulated in *pir-1(tm3198)* mutant animals.** Small RNA libraries for deep

sequencing were constructed by the 5'-independent TAP cloning method, which captures all known species of sRNA. Wild-type hermaphrodite avr3x animals were grown on ivermectin and collected as a range of L4 to young adults, serving as the control for all comparisons. **(A)** Scatter plots of sRNA reads in ten-day counterselected *pir-1* arrested hermaphrodites (*y* axis) *versus* avr3x wild-type hermaphrodites (*x* axis). Each dot represents a unique miRNA and 21U-RNA, or in the case of 22G-RNAs and 26G-RNAs, represents all the sequences mapping to an individual gene, measured as the number of reads per million (rpm). Normalization was performed with total nonstructural reads. No read number cutoff was applied to the graphed data, and all data points were transformed by adding one read (such that one rpm in reality corresponds to zero reads). **(B)** Box-and-whiskers plot of the ratio of reads per million *pir-1*/(*pir-1* + wild type) from data in (A), such that a ratio of 0.5 corresponds to no change in read number between the two samples, and ratios above or below correspond to an enrichment or depletion of reads in the *pir-1* mutant compared to wild-type, respectively. (0.3(3) represents a two-fold downregulation in *pir-1*, and 0.6(6) represents a two-fold upregulation in *pir-1*). Boxes contain 50% of the sRNA loci (between the $25^{th}$ and $75^{th}$ percentiles). The line inside the box represents the median value and the ends of the whiskers represent the loci with the most upregulated or downregulated sRNAs. All box-and-whiskers plots consider genes with sRNAs displaying ≥10 rpm in either mutant or wild-type. Above each bar is the number (n) of genes that met the cutoff. The 'all coding sRNA' category in gray includes all sRNA reads mapping to protein-coding genes, including, but not restricted to the specific 22G-RNA and 26G-RNA categories shown. **(C)** Box-and-whiskers analysis of sRNAs from three-day counter-selected *pir-1* mutants. **(D)** Western blot on seven-day arrested *pir-1* total protein extracts, revealing wild-type expression levels of DCR-1, DRH-3, the 21U-RNA Argonaute PRG-1 and the WAGO Argonaute SAGO-1(WAGO-8).

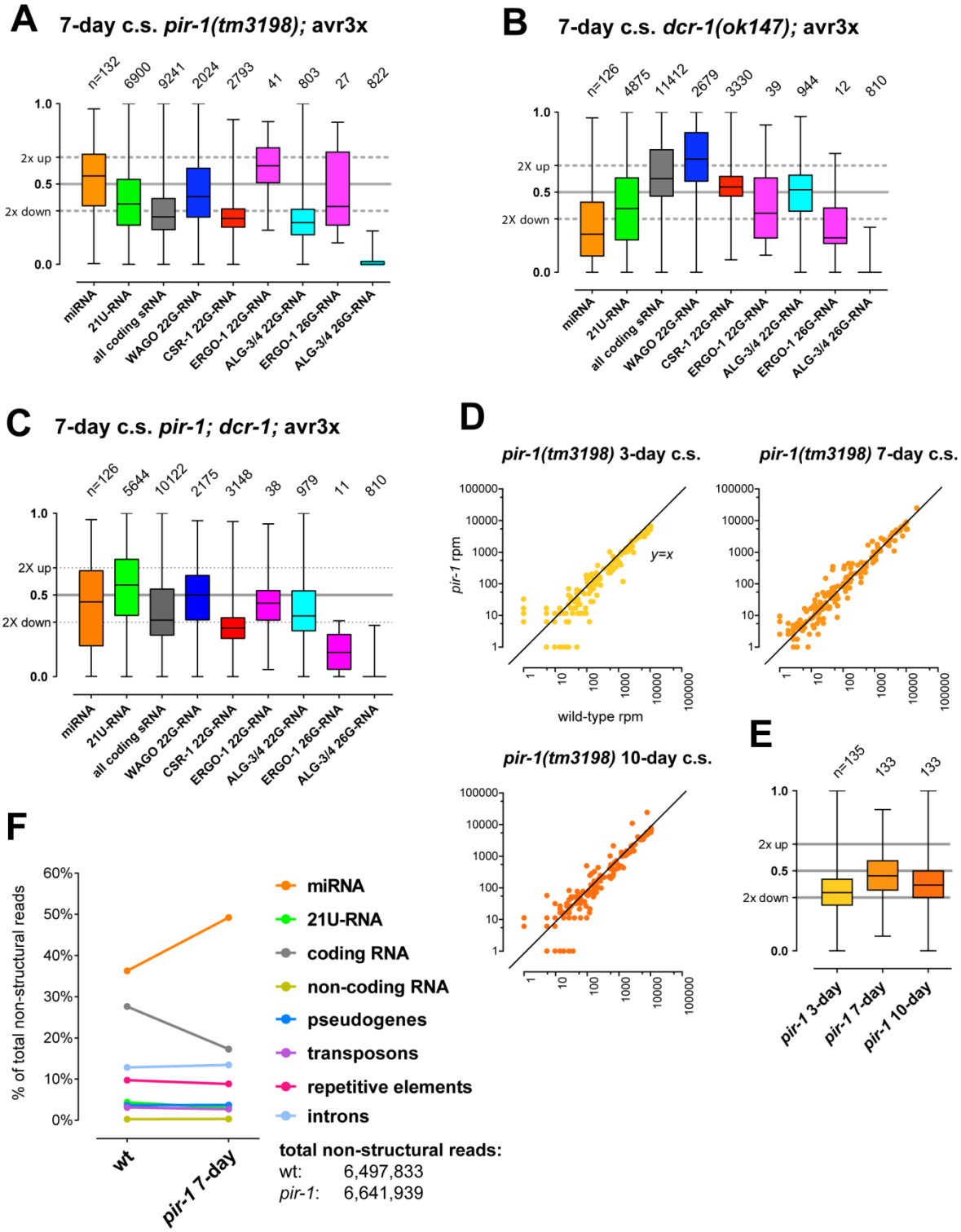



**Figure 4.5. Comparison of small RNA profiles from *pir-1* and *dcr-1* loss-of-function mutants and analysis of transposon and repeat-associated small RNAs in *pir-1* mutant animals. (A)** Box-and-whiskers analysis of sRNAs from seven-day counter-selected *pir-1* animals. **(B)** Box-and-whiskers analysis of sRNAs from seven-day counter-selected *dcr-1* loss-of-function mutant animals arrested as sterile adults. **(C)** Box-and-whiskers analysis of sRNAs from seven-day counter-selected *pir-1; dcr-1* double mutant animals, which are phenotypically identical to the *pir-1* single mutant. **(D)** Scatter plots of reads for sRNAs targeting all expressed transposons at different times of counter-selection. The represented data is from the same sRNA libraries analyzed

in Figure 4.4. and 4.5A **(E)** Box-and-whisker plot representation of the ratio of reads *pir-1*/(*pir-1* + wild type) after applying the ≥10 rpm cutoff. Normalization was performed with total nonstructural reads. **(F)** Representation of the percentage of different sRNA populations in wild-type and seven-day old *pir-1* animals when considering all genome-matching reads minus sense structural reads. The absolute number of these reads in the libraries is indicated in the figure. All classes consider antisense reads, except miRNAs and 21U-RNAs (only sense reads) and reads mapping to repetitive elements (sense and antisense reads).

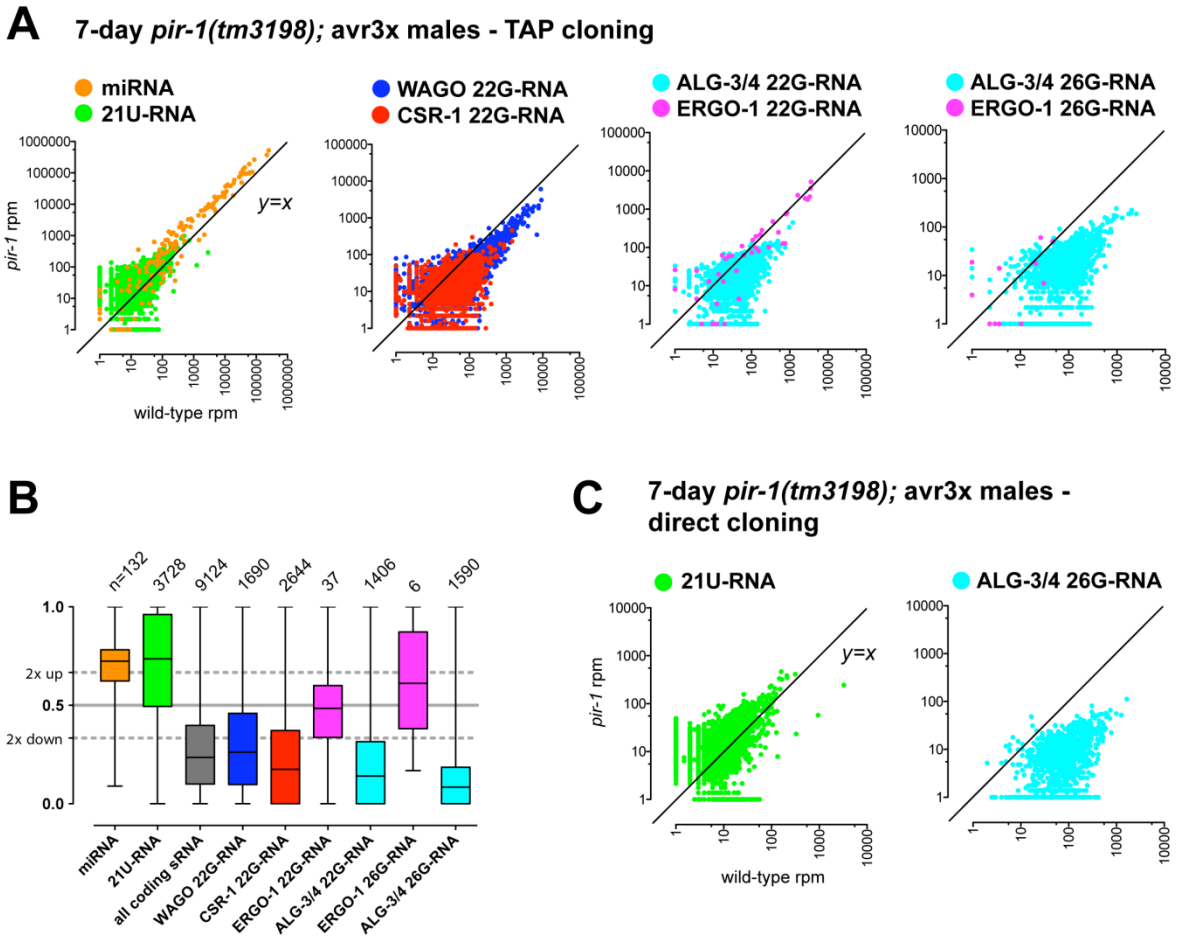



**Figure 4.6. Deep-sequencing analysis of *pir-1* mutant males confirms the downregulation of spermatogenesis-associated 26G-RNAs. (A)** Scatter plots of sRNAs from a population of single-picked worms compared to wild-type avr3x males of the same age, and not grown in the presence of ivermectin. Data was normalized with total nonstructural reads. **(B)** Box-and-whisker representation of the same data in terms of the ratio of reads *pir-1*/(*pir-1* + wild type) after applying a ≥10 rpm cutoff. **(C)** Analysis of the same samples cloned by a direct cloning (5'-independent) approach, which excludes all 5'-triphosphorylated 22G-RNAs. Reads were normalized to miRNAs.

## The ALG-3/4 26G-RNA Defect Is Not Due to Lack of Expression of Target mRNAs

Conceivably, the downregulation of ALG-3/4 26G-RNAs in *pir-1* mutant animals could be due to a lack of expression of template/target mRNA. We therefore deep-sequenced mRNA from three- and ten-day counter-selected *pir-1* mutant populations. Normalization was performed using total coding sense reads and some targets were validated by qRT-PCR (Fig. 4.7). Despite considerable variation in expression levels of numerous individual mRNAs between *pir-1* and wild type, the majority of genes expressed in the control are also expressed in *pir-1* mutant animals (*i.e.*, there appears to be no global decrease in gene expression in *pir-1*; Fig. 4.7A and B, gray bars). When subdividing genes into different Argonaute-target classes, this trend held true for WAGO, CSR-1 and ALG-3/4 classes. Curiously, for the eight ERGO-1 targets that were sequenced above the 50 read-per-million (rpm) cutoff used to build the box-and-whiskers plot there was a marked increase in the *pir-1* mutant, suggesting some deregulation, likely due to decreased levels of ERGO-1-class 22G-RNAs (Fig. 4.4, page 173). In three-day animals, the general mRNA abundance was closer to wild type, while in ten-day animals mRNAs were overall slightly upregulated. We speculate that this may reflect a gradual loss of mRNA silencing due to persistent lowered abundance of germline 22G-RNAs (Fig. 4.4).

Concerning the use of mRNAs as templates for 26G-RNA biogenesis, we needed to consider that transcripts are often targeted by multiple sRNA pathways, making it difficult to distinguish between Dicer-independent 22G-RNAs and Dicer-dependent 22G-RNAs (*i.e.*, triggered by 26G-RNAs generated by Dicer cleavage). For instance, according to the data sets used in our study, out of a total of 3,659 CSR-1 target mRNAs, 424 are also targeted by ALG-3/4 26G-RNAs (from a total of 1,684 ALG-3/4 mRNA targets). In turn, the 2,911 WAGO target mRNAs overlap with 130 ALG-3/4-class targets. As an example, the 26G-RNAs produced from the *ZK973.8* gene are not detected in *pir-1*, yet WAGO-dependent 22G-RNAs still accumulate significantly (all sRNAs, Fig. 4.8A). In cases such as this we cannot determine which fraction of 22G-RNAs arises from Dicer-independent triggering or through Dicer-synthesized 26G-RNAs. We therefore focused on a subset of ALG-3/4 targets that were not only missing 26G-RNAs, but also exhibited a severe depletion of 22G-RNAs (*K03H1.12* as an example in Fig. 4.8A). In other words, we wanted to analyze a representative set of mRNAs from which all 22G-RNAs were triggered by ALG-3/4 26G-RNAs, and not from other pathways (such as CSR-1 or WAGO). The criteria used to create this set were that the ALG-3/4-class mRNAs should

not be targeted by WAGO or CSR-1-class 22G-RNAs; that both 26G-RNAs and 22G-RNAs should be expressed at at least ten rpm in wild type and should exhibit a ten-fold or higher depletion in *pir-1*; and that the mRNAs should be expressed at at least 50 rpm in either wild type or mutant. After ten days of counter-selection, we found 101 genes that met these sRNA criteria. For 99 of these we had corresponding mRNA-seq data that met our 50 rpm cutoff. For this set of targets the average read number was higher in *pir-1* (359.4 rpm) than in wild type (318.3 rpm) (Fig. 4.8D). At three days we found a comparable set of 129 genes and found that the average mRNA-seq reads in *pir-1* (515.5 rpm) was only slightly lower than in wild-type (673.6 rpm). Only 57 mRNAs overlapped between the two sets, providing added confidence that PIR-1 is crucial for the accumulation of 26G-RNAs. Since *pir-1* mutants express mRNAs that are exclusively regulated by ALG-3/4 26G-RNAs at levels close to or higher than those of wild-type animals we conclude that the absence of 26G-RNAs is not attributed to a lack of templates.

When we compared mRNA levels of these exclusive ALG-3/4 26G-RNA targets measured by mRNA-seq in *pir-1* mutant animals with levels obtained by tiling microarray analysis of *alg-3; alg-4* double mutant animals (from Conine *et al.*, 2009) we observed that after ten days these mRNAs were upregulated almost as much as in the *alg-3; alg-4* mutant (Fig. 4.8B). This result indicated that these mRNAs are also desilenced in the *pir-1* mutant, as one would expect to observe if the targets are being actively expressed. Consistent with this, a previous study found that sperm genes were upregulated in *eri-1* and *rrf-3* mutant hermaphrodites, both of which lack 26G-RNAs (Asikainen *et al.*, 2007). At three days, the mRNAs were more upregulated in the *alg-3; alg-4* mutant than in *pir-1*, perhaps due to the fact that in *pir-1* mutant animals the expression of target mRNAs had still not recovered from a more recent loss of silencing 22G-RNAs due to maternal load.

One last possible cause for the absence of ALG-3/4 26G-RNAs could be the lack of expression of the Argonautes themselves, which would lead to destabilization of the sRNAs. Two observations rule out this possibility. First, mRNA-seq showed that in both three- and ten-day old *pir-1* animals the mRNAs of *alg-3* and *alg-4* are expressed at similar or higher levels than in the L4-young adult wild-type control animals (Fig. 4.9A). Second, when examining arrested *pir-1* hermaphrodites expressing a rescuing *gpf::alg-3* transgene (Conine *et al.*, 2010), GFP::ALG-3 could be detected in the spermatogenic area of the germline, in a pattern identical to that of the wild-type background (Fig. 4.9B). Taken together, all the aforementioned results indicate that the lack of 26G-RNAs in *pir-1*

arrested animals is due to a problem in their synthesis or accumulation, and not from a lack of transcripts to be used as templates for the 26G-RNA-specific RdRP RRF-3.

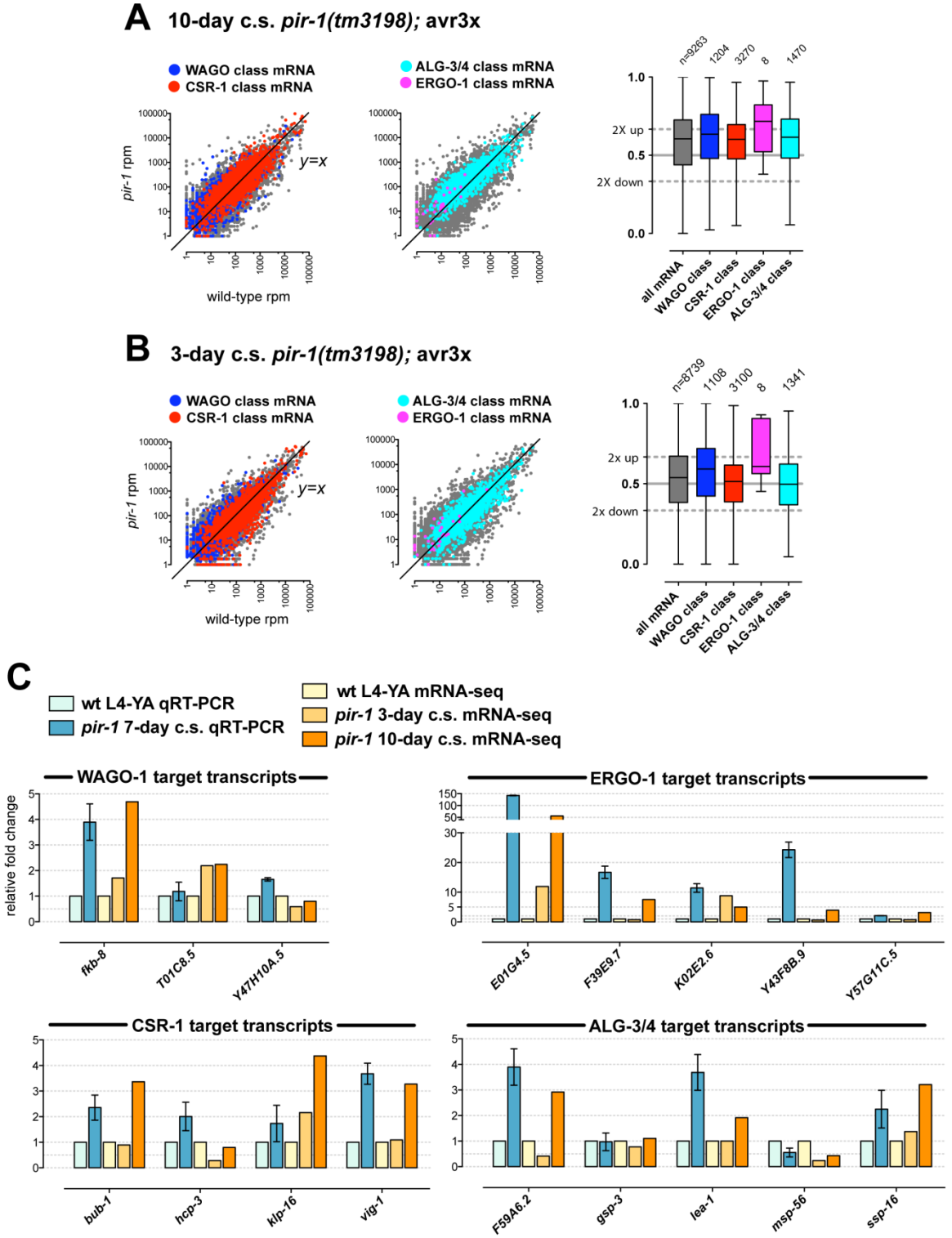



**Figure 4.7. Deep-sequencing analysis of poly(A)-selected RNA in *pir-1* mutant animals.** The RNA samples used in this experiment were the same used for sRNA analysis. **(A)** Scatter plots of sRNA reads in ten-day counter-selected *pir-1* arrested hermaphrodites *versus* avr3x wild-type

animals obtained by TAP cloning. Each dot represents the number of reads per million (rpm) that map to a unique gene. Normalization was performed with total sense reads derived from protein coding genes. No read number per gene cutoff was applied to the graphed data, and all data points were transformed by adding one read (a value of one rpm in corresponds to zero reads). The gray data points correspond to all sequences not belonging to the four categories considered. The accompanying box-and-whiskers plot represents the ratio of reads *pir-1*/(*pir-1* + wild type) after applying a ≥50 rpm cutoff for either *pir-1* or wild-type reads. The number of genes that met this cutoff value is depicted above the graphed bars. **(B)** Identical analysis applied to three-day counter-selected *pir-1* animals. **(C)** qRT-PCR validation of some mRNAs from each of the different sRNA-target categories considered (blue bars). Measurements were performed on three independent replicates for wild-type and *pir-1* animals counter-selected for seven days. Transcript levels were normalized to actin mRNA (*act-3*) and are relative to wild-type. Error bars report the SEM. Relative fold changes from RNA-sequencing were calculated by normalizing transcript rpm to total actin rpm, relative to normalized rpm in avr3x wild type (orange bars). Note that qRT-PCR fold changes best match the mRNA levels from the ten-day counter-selected sequencing data, as expected from a more similar developmental state and age.

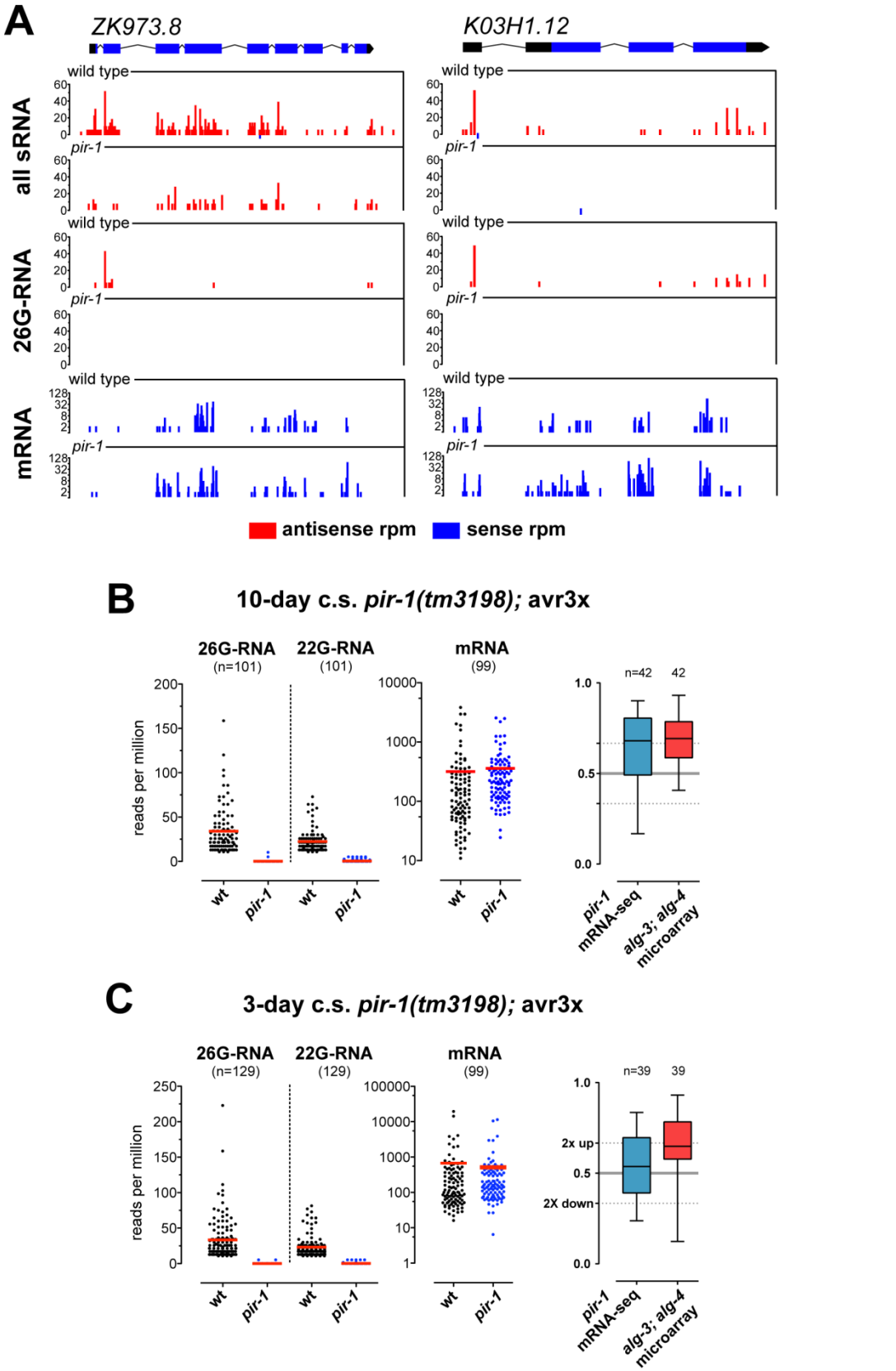



**Figure 4.8. *pir-1* mutant animals express ALG-3/4 mRNA templates from which downregulated 26G-RNAs are derived. (A)** Genome browser representation of two ALG-3/4 26G-RNA target mRNAs. Each bar maps the first nucleotide of each read and the height indicates the associated rpm value on the scale. In both cases the transcripts are expressed in both wild-type

and ten-day *pir-1* animals. sRNAs (mostly 22G-RNAs) are expressed in *pir-1* from *ZK973.8* in the absence of ALG-3/4 26G-RNAs, because it is also targeted by a 26G-RNA-independent WAGO pathway. In wild type most of these WAGO-dependent sRNAs are distributed along the exons of the entire gene, in contrast with 26G-RNAs which are located primarily at the 5'-end. Conversely, sRNAs targeting *K03H1.12* are exclusively triggered by 26G-RNAs as no sRNAs are present in *pir-1* animals and all sRNAs map around the sites of 26G-RNA accumulation. **(B)** Comparison in ten-day *pir-1* animals of 22G-RNA, 26G-RNA and mRNA reads per million for a set of genes whose 22G-RNAs are entirely dependent on ALG-3/4 26G-RNAs, of which *K03H1.12* is an example. The criteria for building this gene set were the following: (1) ALG-3/4-class mRNAs not targeted by WAGO or CSR-1-class 22G-RNAs, (2) expression at ≥10 rpm for both 26G-RNAs and 22G-RNAs in wild-type, (3) ≥10-fold 26G-RNA and 22G-RNA depletion in *pir-1* relative to wild type, and (4) ≥50 rpm for the corresponding mRNAs in *pir-1* or in wild type. 'n' indicates the number of genes that met the criteria and the red bar marks the mean rpm value for each group. The box-and-whisker plot compares the ratio of mRNA measurements mutant/(mutant + wild type) obtained by mRNA-seq comprised in the analyzed sets to the same RNAs measured in a tiling microarray of *alg-3; alg-4* mutant young adults. **(C)** Same analysis applied to three-day *pir-1* animals.

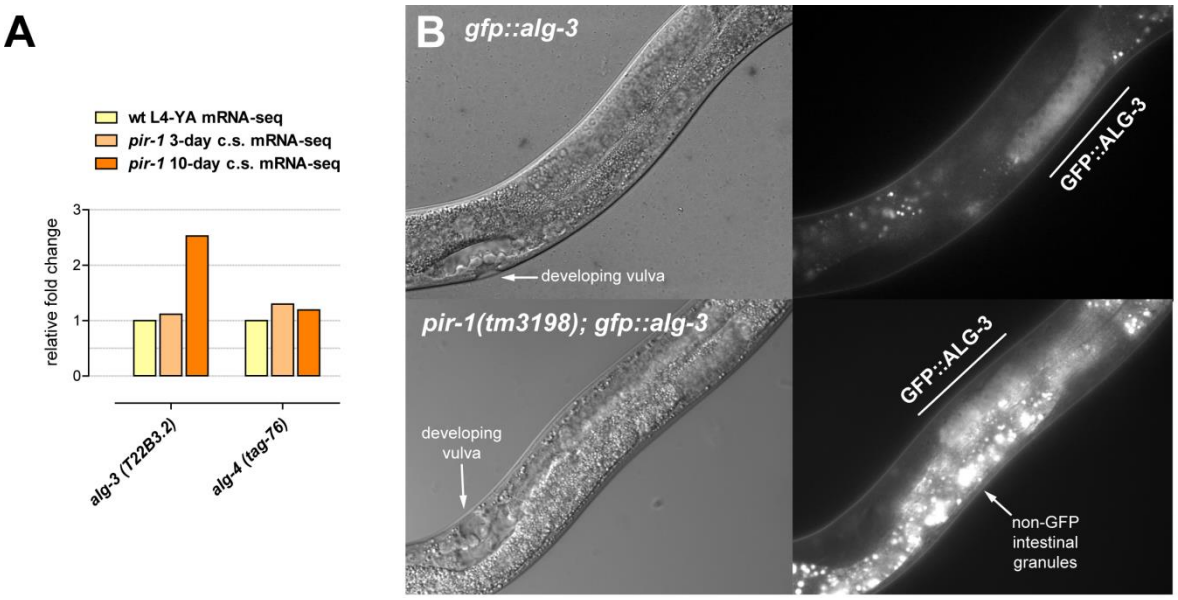



**Figure 4.9. *pir-1* mutants express *alg-3* and *alg-4* mRNAs, and ALG-3 protein. (A)** *alg-3* and *alg-4* mRNAs are expressed in *pir-1* animals at a wild-type or higher level. Fold-change was calculated based on total coding sense RNA-normalized reads. **(B)** Expression of GFP::ALG-3 from an integrated transgene with native regulatory elements in a live wild-type L4 hermaphrodite or in a live arrested *pir-1* homozygote hermaphrodite. GFP signal can be seen near the vulva in the proximal germline where spermatogenesis occurs, localizing to the cytoplasm but also surrounding the nuclei in P granules. In *pir-1* animals the signal is partially masked by the presence of an abnormal amount of fluorescent intestinal granules frequently observed in older arrested animals. It is still clear, however, that GFP::ALG-3 exhibits a wild-type localization.

## *pir-1* Mutants Exhibit an Upregulation of Innate Immunity Pathway Gene Expression

Continuing the characterization of *pir-1*-associated phenotypes, we analyzed our mRNA-seq libraries for changes in the expression of specific mRNAs. In order to create a stringent, high-confidence list of misregulated mRNAs, we only considered genes that were at least two-fold down- or upregulated in both three- and ten-day counter-selected *pir-1* animals. The inclusion of the three-day old *pir-1* animals provided a control for age-dependent effects, since the wild-type avr3x worms were grown for the same period of time. Additionally we set a cutoff of at least 50 rpm for each mRNA in any of the three samples considered. With these criteria, we obtained a list of 813 downregulated and 1,393 upregulated genes. We organized the genes for which functions are known or predicted into informative biological categories. These genes, along with their descriptions, fold de-regulation relative to wild-type, and read numbers are compiled in electronic Appendix D. The top ten categories for up- and downregulated mRNAs are depicted in Figure 4.10.

Generally, downregulated genes are associated with processes that are expected to function during active growth. Examples include a variety of collagens, which are heavily expressed at every larval molt to build successively larger cuticles (outer extracellular matrix); proteins related to protein metabolism, including several ribosomal proteins; regulators of cell division, including the essential cyclin-dependent kinase *cdk-1*, required for cell cycle progression in both meiosis and mitosis; and histones. Consistent with the absence of oocytes and fewer mature spermatids in *pir-1* mutant animals, we also observed a heavy downregulation of mRNAs for oocyte vitellogenins and other yolk proteins, as well as numerous mRNAs associated with mature sperm, especially of the major sperm protein (MSP) family. Additionally, we registered a downregulation of several genes encoding hedgehog-like intracellular signaling proteins that perform essential roles in growth and morphogenesis. This gene list is in agreement with the developmental arrest that characterizes *pir-1* mutants.

Unexpectedly, when we analyzed the list of upregulated mRNAs, we found an enrichment for genes involved in innate immunity and response to various stresses, along with genes for the upstream signaling pathways that activate those responses. Specifically, central components of the TGF-β-like pathway (*T*ransforming *G*rowth *F*actor β), p38/MAPK pathway (*M*itogen-*A*ctivated *P*rotein *K*inase), DAF-2/DAF-16 insulin-like receptor pathway (Abnormal *Da*uer *F*ormation), and apoptosis pathway were upregulated (pathways reviewed in Ermolaeva and Schumacher, 2014). Downstream effectors included

factors required to combat oxidative stress, heat-shock, and bacterial infection: carbohydrate-binding C-type lectins, a variety of antimicrobial peptides, lysozymes, proteins associated with vesicles (including lysosomes), autophagy and endocytosis, ubiquitination proteins, and factors required for the misfolded protein response. Interestingly, our results match those reported in a microarray-based study of gene expression in *C. elegans dcr-1*, *rde-4* and *rde-1* loss-of-function mutants (Welker *et al.*, 2007). All mutants showed enrichment for innate immunity mRNAs, uncovering an important role for RNAi factors in the suppression of stress and immunity pathways in *C. elegans*. The addition of PIR-1 to this repertoire further underscores the possibility that these important pathways can be specifically downregulated via endogenous siRNAs. However, we do not have enough knowledge to exclude the possibility that this response is an indirect effect from a global shift in sRNA populations and consequent gene expression changes. In spite of being beyond the scope of our study, this finding deserves further investigation.

Importantly, *pir-1* mutant animals exhibited higher levels of mRNAs for several sRNA factors, consistent with the absence of defects in exo- and non-ERI endo-RNAi, miRNA and 21U-RNA pathways. For RNAi, these included mRNAs encoding the primary Argonaute RDE-1, the secondary Argonaute WAGO-8(SAGO-1), RDE-2 (a germline RNAi factor found in complex with the 3'-5' exonuclease MUT-7; Tops *et al.*, 2005), the RdRP RRF-1, and CDE-1 (also known as CID-1 or PUP-1), a poly(U) polymerase required for regulation of CSR-1-associated 22G-RNAs by 3' uridylation (van Wolfswinkel *et al.*, 2009). Regarding miRNAs, these included mRNAs for the Argonautes ALG-1, ALG-2 and for miRISC-associated factors AIN-1 and NHL-2, which promote miRNA-mediated silencing (Ding *et al.*, 2005; Hammell *et al.*, 2009). Lastly, the 21U-RNA Argonaute PRG-1 was also found to be upregulated in *pir-1* mutant animals.

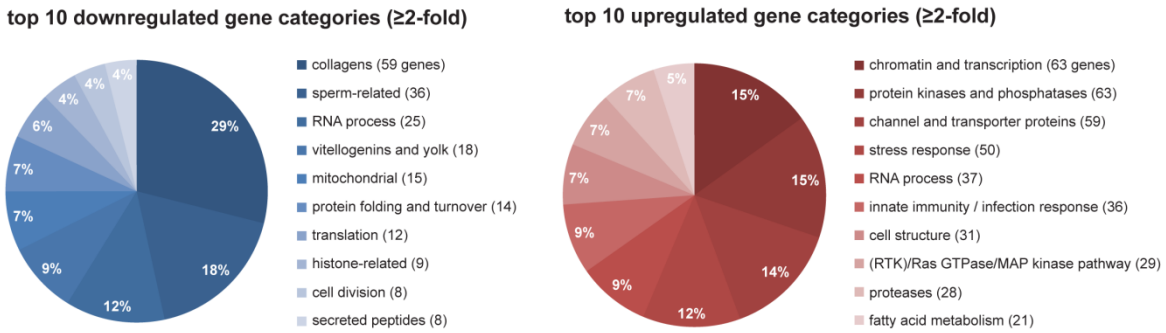



**Figure 4.10. Analysis of mRNAs depleted or enriched in *pir-1* mutant animals according to biological category.** From our mRNA-sequencing data we identified mRNAs that were consistently down- or upregulated in both three- and ten-day *pir-1* animals relative to wild-type. We established a cutoff of ≥50 rpm for at least one of the samples and a ≥2-fold misregulation. The categories were manually curated following gene ontology analysis from the WormBase Ontology Browser (WormBase.org; Harris *et al.*, 2014). Only the top ten categories are represented, with the number of genes for each category in parenthesis. The complete list of genes accompanied by the corresponding read counts, relative fold misregulation, and gene description can be found in electronic Appendix D.

## ALG-3/4 26G-RNAs Are Enriched in Nuclear Extracts

Due to the primarily nuclear localization of PIR-1, we wanted to examine the sRNA populations in nuclei isolated from whole wild-type animals. Since we were unable to isolate pure cytoplasmic sRNAs, we used total sRNAs as a control. Consistent with studies that compared miRNA abundance in cytoplasm *versus* nucleus of human cells concluding that the concentrations in the two compartments were similar (Liao *et al.*, 2010; Gagnon *et al.*, 2014), the ratio of nuclear to total miRNAs in *C. elegans* indicated no significant change in their abundance (Fig. 4.11A and B).

Regarding 21U-RNAs, we were surprised not to observe an appreciable change, given the fact that PRG-1 predominately localizes to P granules, on the outer nuclear periphery. However, since no studies have directly addressed 21U-RNA localization, it is plausible that a sizeable population of 21U-RNAs may reside inside the nucleus. Unlike mouse and flies, where piRNAs exert their function directly in the nucleus to deposit repressive chromatin marks, in *C. elegans* 21U-RNAs are thought to promote these silencing events by triggering 22G-RNA synthesis in P granules (reviewed in Cecere and Grishok, 2014). These 22G-RNAs associate with the nuclear RNAi machinery to induce chromatin silencing, but this does not conflict with the possibility that many 21U-RNAs may enter the nucleus to exert this and other functions.

Analysis of WAGO 22G-RNAs revealed that they were slightly depleted from the nuclear extract, consistent with the majority of WAGO Argonautes exerting post-transcriptional silencing in the cytoplasm and in P granules. When we considered CSR-1 we were surprised to find that they were also slightly depleted. Since CSR-1 heavily localizes to P granules throughout the germline and is also known to interact extensively with chromatin, we expected instead to see an enrichment of CSR-1 22G-RNAs in the nuclear fraction. The fact that that P-granule associated 22G-RNAs were depleted from our nuclear extracts, indicated that our nuclear isolation method likely disrupts P granules from the outer surface of nuclei.

Since the ALG-3 Argonaute also accumulates in P granules (Conine *et al.*, 2010), we would expect ALG-3/4 26G-RNAs to behave like CSR-1 22G-RNAs regarding their relative nucleo-cytoplasmic abundances. Instead, ALG-3/4 26G-RNAs were the most enriched sRNAs in nuclear fractions, contrasting with ERGO-1 26G-RNAs which were depleted. Additionally, the fact that (1) only a small proportion of germline cells undergo spermatogenesis, that (2) we used total instead of cytoplasmic extracts in our comparison, and that (3) there was a slight cytoplasmic contamination in our nuclear extract (Fig. 4.11D), implies that the nuclear enrichment for ALG-3/4 26G-RNAs measured in our experiment is likely an underestimation of the actual enrichment of 26G-RNAs in the nuclei of spermatogenic cells. Paradoxically, the P-granule localization of ALG-3 suggests that its 26G-RNAs function mainly at these sites. Nonetheless, it is conceivable that a large fraction of Argonautes in P granules are not loaded with sRNAs and that enough nuclear ALG-3 and ALG-4 exist to stabilize the 26G-RNAs in that compartment. In contrast, and consistent with activity outside the nucleus, stands our result that the 26G-RNA-dependent ALG-3/4 22G-RNAs are not enriched in the nucleus. Since we demonstrated that the ERI complex assembles in the nucleus with PIR-1 (Fig. 3.8, page 136), and that PIR-1 is present in the nuclei of germ cells undergoing spermatogenesis (L4 larvae and male germlines in Fig. 2.4A, page 81, Fig. 2.5C, page 82 and Fig. 2.7, page 85), it is plausible that ALG-3/4 26G-RNAs may be synthesized within the nucleus, and subsequently exported into P granules perhaps already in association with their target mRNAs. However, the synthesis of some ALG-3/4 26G-RNAs may also take place in the common cytoplasm of the spermatogenic syncitial germline, where both PIR-1::GFP and GFP::ALG-3 are also found to be diffusely distributed (compare Chapter II figures with Fig. 4.9, page 182).

We extended the analysis to other sRNA categories that we would predict to be enriched in nuclei, measured as fractions of total nonstructural reads (Fig. 4.11C). In

comparison to miRNAs or 21U-RNAs, which were cloned at similar frequencies in total and nuclear extracts, and to antisense sRNAs derived from protein-coding RNAs, which were cloned less frequently in nuclear extracts, only transposon and repetitive-element-associated sRNAs were cloned more frequently in nuclear extracts. The enrichment of these categories likely reflects the need to control transposition and expansion of repeat elements to limit their deleterious effects on genome integrity. This result also reinforced our confidence that the nuclear enrichment we observed for ALG-3/4 26G-RNAs is not artifactual.

In a prior study from our laboratory it was observed that, for specific targets, ALG-3/4 26G-RNAs tend to accumulate at the 5' or at the 3' regions of mRNAs (Conine *et al.*, 2010). Considering this information, we split all mRNA templates into 20 separate bins of equal length within each gene, and added all 26G-RNAs reads falling into each bin to interrogate their global distribution (Fig. 4.12A). When examining the first four and last four bins, which concentrate the bulk of 26G-RNA reads, we saw that there were twice as many reads derived from 5' ends than from the 3' ends (1.9-fold for total extracts and 2.3-fold for nuclear extracts from wild-type animals). We wondered whether this unusual preference for the 5' regions of mRNAs could be related to the nuclear enrichment of ALG-3/4 26G-RNAs we observed. Interestingly, in *pir-1* mutants this bias disappeared completely, although the 26G-RNA read number may be too low to confidently infer biological relevance. In contrast, ERGO-1 26G-RNAs accumulated predominately along the 3' half of its targets. 22G-RNAs, on the other hand, are abundantly distributed across the entire mRNAs, but tend to show an increased accumulation at the 3' ends (Gu *et al.*, 2009). We confirmed this for 22G-RNAs against ALG-3/4, CSR-1 and WAGO targets. Curiously, ERGO-1 22G-RNA distribution profiles more closely resembled the corresponding 26G-RNA distribution, suggesting that the majority of 22G-RNAs are triggered directly by 26G-RNAs within the ERGO-1 mRNA target set. Collectively, all of the aforementioned findings are highly suggestive that the nucleus is a site of synthesis for spermatogenesis-associated 26G-RNAs.

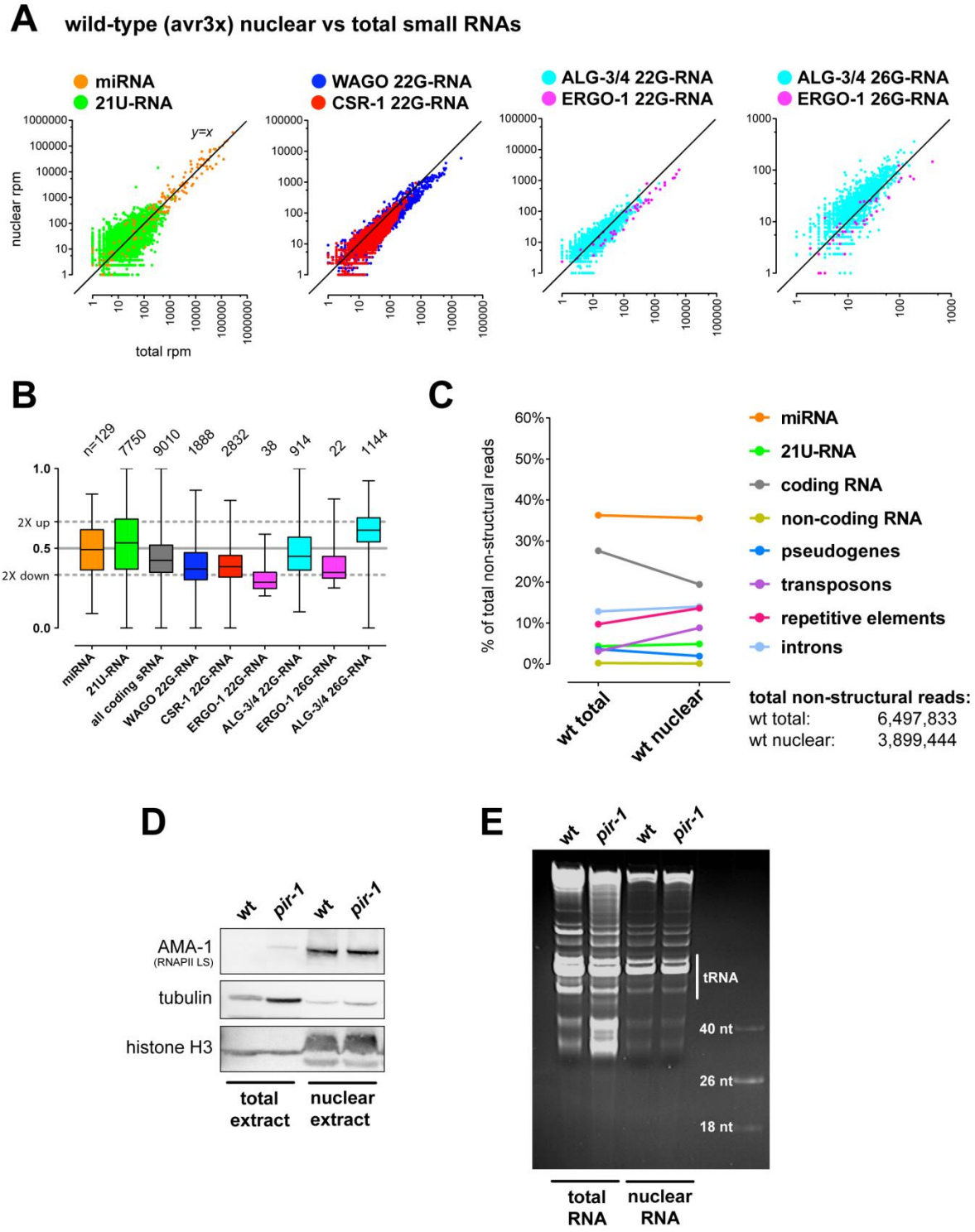



**Figure 4.11. ALG-3/4-dependent 26G-RNAs are enriched in nuclear extracts of wild-type animals relative to other small RNA species. (A)** Scatter plots of sRNA rpm in nuclear enriched RNA (*y* axis) compared to total RNA (*x* axis). Normalization was performed with total nonstructural reads. **(B)** Box-and-whisker plots of relative sRNA abundances from cloning of total RNA or RNA isolated from purified nuclei of avr3x wild-type L4 to young adult hermaphrodites. The vertical axis represents the ratio of nuclear reads/(nuclear reads + wild-type reads) such that a value above 0.5 is interpreted as sRNAs that are enriched in the nucleus in comparison to the total pool of sRNAs. **(C)** Representation of the percentage of different sRNA populations cloned from total and nuclear RNA, considering all genome-matching nonstructural reads. The absolute number

of these reads in the libraries is indicated in the figure. **(D)** Western analysis of nuclear extracts used for RNA purification, showing a depletion of cytoplasmic tubulin and an enrichment for the large subunit of Pol II (AMA-1) and histone H3. **(E)** Ethidium-bromide stained denaturing 15% polyacrylamide gel of RNAs to assess the integrity of the samples used for the nuclear sRNA deep-sequencing experiments.

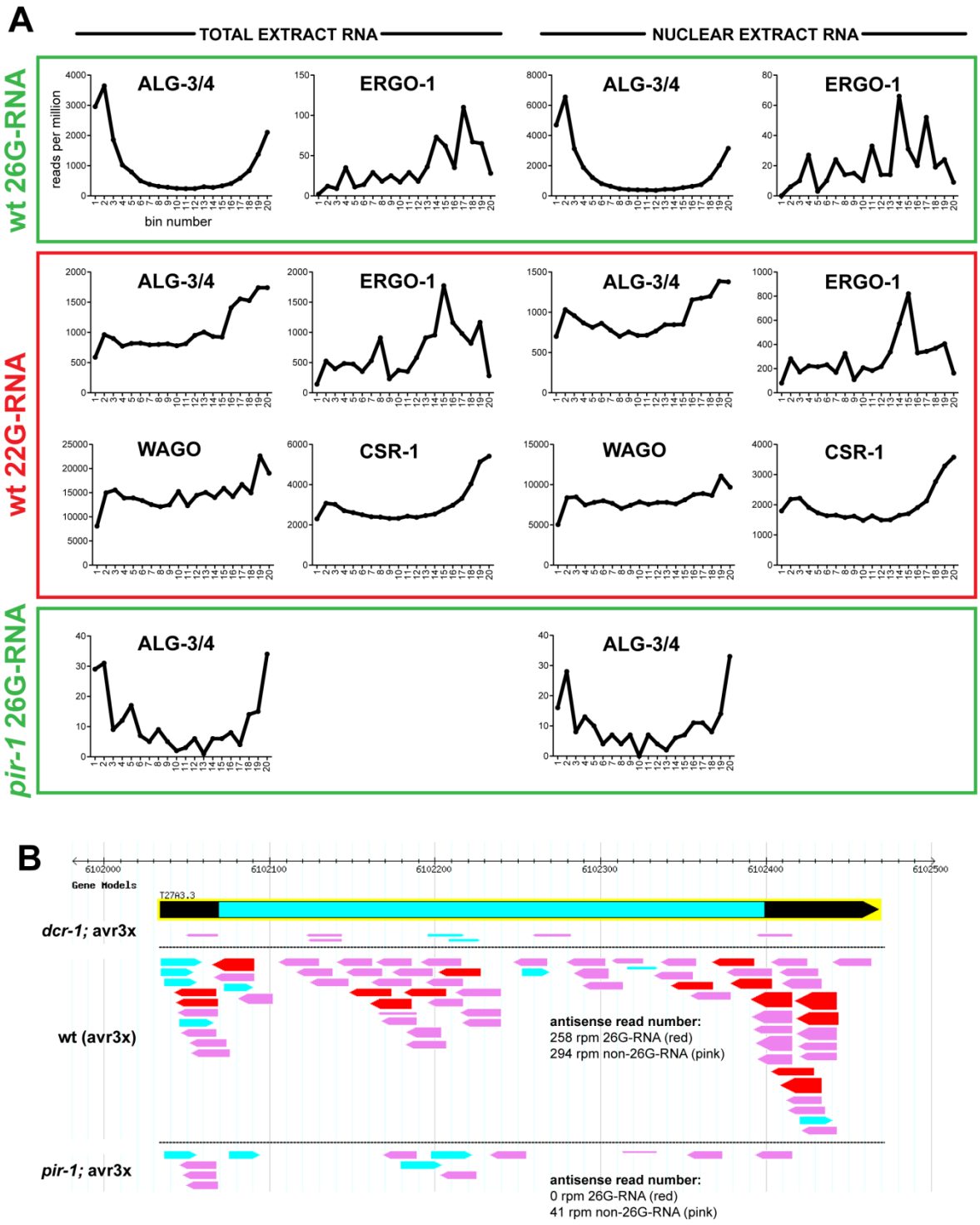



**Figure 4.12. ALG-3/4 26G-RNAs map predominately to the 5' regions of template/target mRNAs. (A)** The distribution of ALG-3/4 26G-RNAs is expressed as the total normalized number of reads derived from one of 20 bins into which all matching mRNA sequences were divided.

Distributions were compared with ERGO-1 26G-RNA and with 22G-RNA distributions from the four Argonaute classes we considered (inside red rectangle). We also considered total and nuclear sRNAs (notice differences in read numbers), as well as the distribution of ALG-3/4 26G-RNAs in seven-day counter-selected *pir-1* mutant animals. **(B)** Genome browser representation of reads targeting the gene *T27A3.3*. This gene expresses a short (330 nt) intronless transcript targeted by ALG-3/4 26G-RNAs, which are distributed in a relatively uniform manner along its length (in red). All other antisense sRNAs (pink) map close to 26G-RNAs and are downregulated in *dcr-1* and *pir-1* mutants, implying that they are triggered by the DCR-1- and PIR-1-dependent 26G-RNAs. Sense sequences in blue are mostly pieces of degraded transcript with no 5'G bias.

## PIR-1 Removes the Two Terminal Phosphates from 5'-Triphosphorylated RNA *in Vitro*

Similarly to what has been shown for the baculoviral and human homologs (Takagi *et al*., 1998; Deshpande *et al.*, 1999), we wanted to test whether *C. elegans* PIR-1 also possesses RNA phosphatase activity. We therefore cloned wild-type and catalytically inactive mutant (negative control) *pir-1* cDNA sequences into an expression vector allowing C-terminal fusion with a 6x-histidine tag. Proteins were expressed in *E. coli*, purified and incubated with *in vitro* synthesized single-stranded RNAs of about 26 nt. As the substrate RNAs are 5'-triphosphorylated, we incorporated Terminator 5' phosphate-dependent exonuclease in the reaction to degrade RNAs that would become monophosphorylated if PIR-1 functioned as predicted. PIR-1-untreated RNA was resistant to Terminator relative to the no-treatment control (Fig. 4.12A). RNA that was treated with wild-type PIR-1 was almost entirely degraded, indicating that it became monophosphorylated and therefore susceptible to Terminator digestion. Incubation with catalytically inactive PIR-1(C150S) protein did not allow Terminator-mediated degradation, showing that the cysteine-based catalytic activity of PIR-1 is required for 5' dephosphorylation. Since, under certain reaction conditions, Terminator exonuclease can also degrade 5'-hydroxyl RNAs, we repeated the experiment with an added control where the RNA was pre-treated with CIP to completely dephosphorylate its 5' ends (Fig. 4.12B). This control was necessary to discard the possibility of contaminating phosphatases in the purified PIR-1 fractions used in the assays. In this experiment, the CIP-treated RNA was still resistant to Terminator, while the PIR-1-treated RNA suffered degradation, proving that indeed the activity of *C. elegans* PIR-1 *in vitro* is to remove the first two phosphates from the 5' ends of RNA molecules.

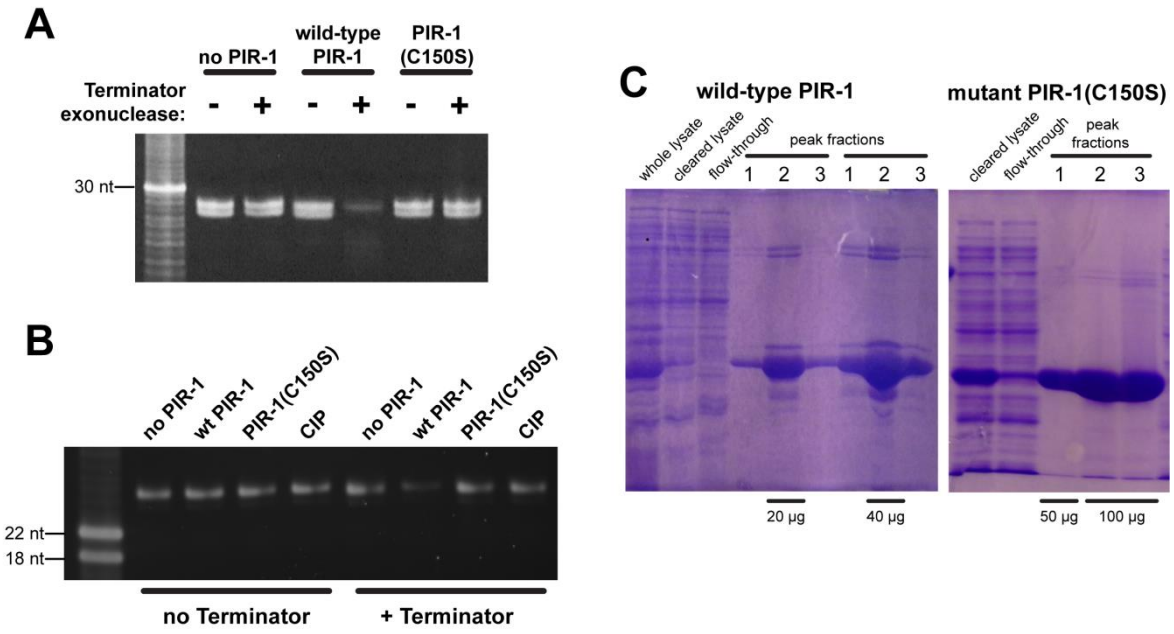



**Figure 4.13. PIR-1 catalyzes the dephosphorylation of 5'-triphosphorylated RNA, leading to a 5'-monophosphorylated product. (A)** Denaturing PAGE analysis of digestion of an *in vitro*-transcribed 5'-triphophorylated 26 nt RNA with recombinant wild-type or catalytically inactive (C150S) PIR-1. Terminator exonuclease was included in the same reaction for simultaneous digestion. The depletion of substrate indicates that phosphate groups were removed. The double bands are two sizes from the same transcription reaction, which were co-recovered following PAGE purification. **(B)** Independent experiment where the same 26-nt RNA (purified independently from the one in (A)) was digested first with recombinant PIR-1 (wt and mutant) or CIP and subsequently with Terminator exonuclease. The CIP control showed that the digestion by wild-type PIR-1 resulted in a 5'-monophosphate and not a 5'-hydroxyl RNA, both of which can be digested by Terminator under certain reaction conditions. **(C)** SDS-PAGE/Coomassie staining analysis of the purified histidine-tagged PIR-1 proteins used in the assays before dialysis.

## Contributions

Daniel Chaves constructed and grew all *C. elegans* strains, extracted RNA and performed RNAi experiments, northern blots, qRT-PCRs and western blots. Some cloning libraries were prepared by Daniel Chaves, but most, including mRNA libraries, were prepared in our laboratory by Weifeng Gu. Mapping of reads and preliminary analyses of deep-sequencing data were performed by Weifeng Gu, and subsequent analyses were performed by Daniel Chaves. Recombinant PIR-1 *in vitro* dephosphorylation assays were performed by Weifeng Gu in his own new laboratory at the University of California Riverside, with contributions from his students Ruidong Li, who purified wild-type PIR-1, and Lichao Li, who purified the catalytically inactive PIR-1.

## DISCUSSION

In this chapter we described results that conclusively implicate PIR-1 in sRNA biogenesis, particularly of the ALG-3/4 26G-RNA class, which promotes sperm development (Gent *et al.*, 2009; Han *et al.*, 2009; Pavelec *et al.*, 2009; Conine *et al.*, 2010). Furthermore, we have shown that, similarly to its human and baculoviral orthologs, *C. elegans* PIR-1 catalyzes the removal of the two terminal phosphates from 5'-triphosphorylated RNAs. Below we present and discuss a model that incorporates these findings to explain why PIR-1 activity is required for 26G-RNA accumulation. Additionally, we explore the intriguing possibility that ALG-3/4 26G-RNAs may be synthesized in the nucleus.

### Enzymatic Activity of PIR-1 in the Context of the ERI Complex: a New Step in 26G-RNA Biogenesis

After having determined that PIR-1 interacts with proteins of the ERI complex, analysis of the sRNAs expressed in the *pir-1* loss-of-function mutant uncovered a defect in the accumulation of sperm-specific 26G-RNAs that associate with the Argonautes ALG-3 and ALG-4. This molecular phenotype is in line with the spermatogenesis defect and the expression pattern of PIR-1::GFP in the male and hermaphrodite germline described in Chapter II. While many of the molecular events and respective order that lead to a mature 26G-RNA are still not defined, we propose a model based on the activities of some of the proteins that compose the ERI complex (Figure 4.14).

Synthesis of dsRNA is initiated by the RdRP RRF-3 on the mRNA template. Similarly to RRF-1 and EGO-1, RRF-3 preferentially starts synthesis on a cytidine residue of the template strand, originating a 5'-triphosphorylated guanosine as the first nucleotide of its product. This makes it unlikely that the synthesis is primer-dependent, raising the long-standing question of how 26G-RNA synthesis is triggered. Findings from a recent study in *C. elegans* showed that Dicer binds to thousands of regions of secondary structure predominately within the coding and 3' UTR regions of mRNAs, without, however, generating a detectable amount of sRNAs (Rybak-Wolf *et al.*, 2014). In light of this finding, it is plausible that the ERI complex binds its targets via the interaction of Dicer with such dsRNA structures, perhaps in cooperation with partners capable of binding dsRNA, such as RDE-4 and DRH-3. Studies of DRH-3 activity *in vitro* have shown that even though its preferred substrate is dsRNA, and that it is activated upon binding dsRNA

(as judged from the induction of strong ATP-hydrolyzing activity), this enzyme is not capable of unwinding annealed RNA strands (Matranga and Pyle, 2010). Additionally, DRH-3 is also capable of binding ssRNA at an affinity four-fold lower than dsRNA. Since DRH-3 is an ortholog of the cytoplasmic viral sensor RIG-I (*R*etinoic Acid-*I*ncucible *G*ene I), which has been shown to translocate along dsRNA in an ATP-dependent manner, it has been proposed that DRH-3 could possess a similar translocase activity (Myong *et al.*, 2009). With this possibility in mind, we can imagine a scenario wherein the binding of the ERI complex to an internal mRNA stem, possibly accompanied by its cleavage or unwinding, enables DRH-3 to scan the mRNA in a 3'-5' direction until a cytidine nucleotide is encountered and RRF-3 can start synthesizing a complementary strand.

Since Dicer requires dsRNA termini to cleave, it is not likely fully engaged by the newly synthesized dsRNA substrate until the template strand has been cleaved close to the site of RdRP initiation. We speculate that, upon binding of the ERI complex to the target RNA, such a step could be mediated through endonucleolytic cleavage by one of the proteins in the complex. A putative candidate is RDE-8, which our laboratory recently implicated as having endonucleolytic activity on mRNAs targeted by RNAi (Tsai *et al.*, 2015). The mRNA cleavage event mediated by RDE-8 is proposed to promote the engagement of secondary RdRPs, such as RRF-1, to produce silencing secondary 22G-RNAs. Importantly, consistent with our finding that RDE-8 is pulled down with PIR-1 complexes, RDE-8 was found to be required for the accumulation of 26G-RNAs. Another potential candidate that has also been pulled down with PIR-1 is the protein ERI-9, which, despite sharing significant homology with RDE-8, has thus far only been implicated in ERGO-1 26G-RNA biogenesis (Pavelec *et al.*, 2009; Zhuang and Hunter, 2012). However, because catalytically inactive RDE-8 is able to rescue the 26G-RNA defect of an *rde-8* loss-of-function mutant, and because ERI-9 lacks the conserved catalytic residues that characterize the Zc3h12a family of nucleases, it is likely that such an mRNA cleavage step is not executed by either of these factors. We therefore cannot exclude that other unidentified nucleases may perform this role, by either stable or substrate-dependent transient interactions with the ERI complex.

According to a study using embryo extracts to process dsRNA with different types of termini, it was demonstrated that *C. elegans* Dicer cleavage can result in a 26-nt 5'-3' sRNA and a shorter antisense 22-23-nt strand, provided the dsRNA substrate has a blunt terminus (Welker *et al.*, 2011). Additionally, under these conditions and following the first cleavage, Dicer adopts an ATP-dependent processive cleavage mode to generate phased

23-nt duplexes with 2-nt 3' overhangs from the remainder of the dsRNA molecule. This activity implies usage of Dicer's helicase domain, which is strictly required for 26G-RNA accumulation (Gent *et al.*, 2010; Pavelec *et al.*, 2009; Welker *et al.*, 2010). In contrast, dsRNAs with 3'-overhanging termini can be cleaved in the absence of ATP to produce 21-23 nt duplexes. This scenario recapitulates miRNA cleavage from pre-miRNAs and primary siRNA cleavage from long dsRNA, which have been shown not to require a functional Dicer helicase domain in the studies referenced above. Bearing these properties in mind, one could imagine that following an endonucleolytic cut of the template RNA, 3'-5' digestion could occur concurrently with scanning of the target mRNA for cytidines by the ERI complex. ERI-1b, which was shown to have 3'-5' exonuclease activity by degrading the 3'-overhangs of siRNA duplexes (Kennedy *et al.*, 2004) could provide such an activity. Digestion would stop at the initial site of RRF-3 synthesis, as ERI-1b is unable to hydrolyze dsRNA, thus generating a blunt end for Dicer to act on.

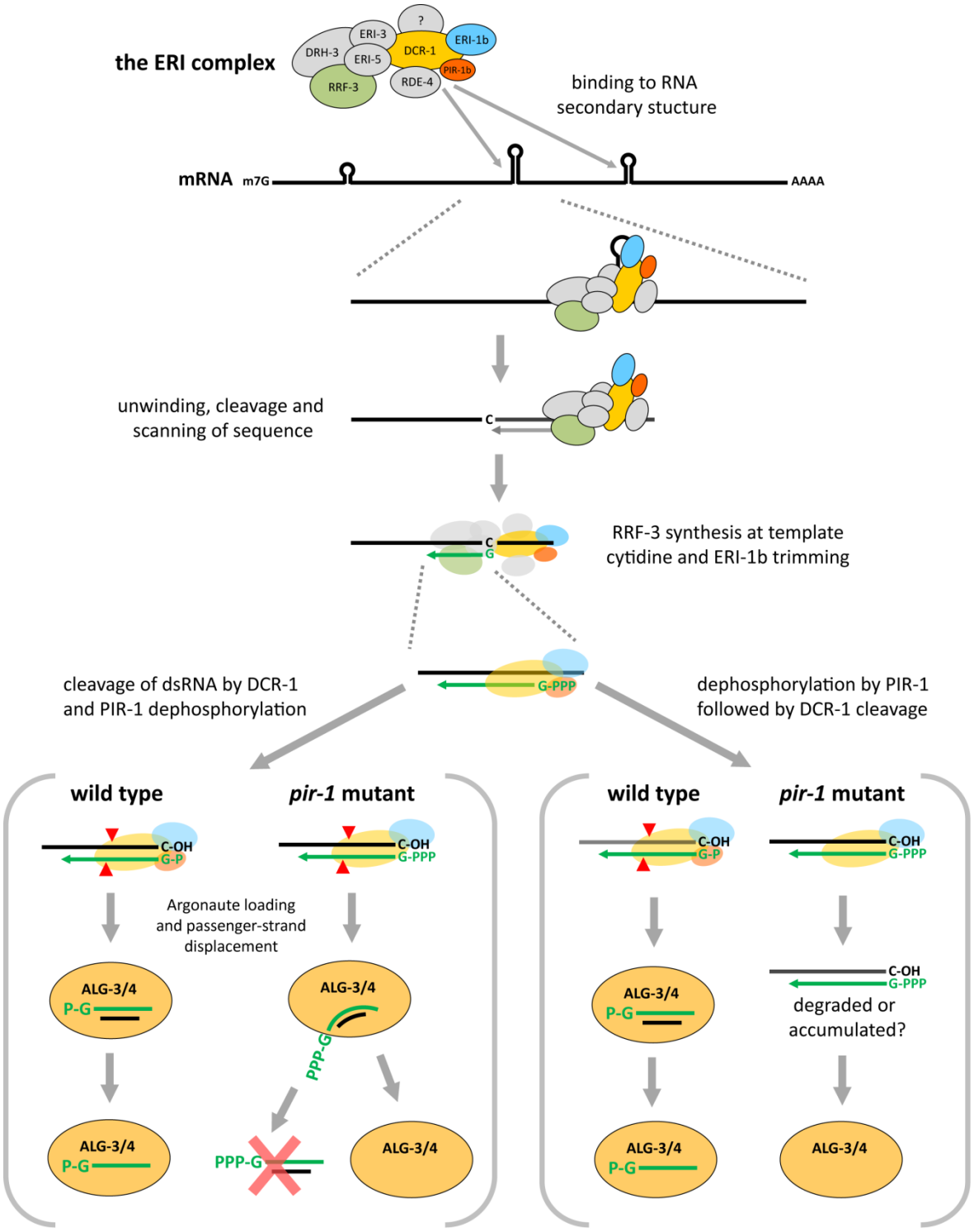
But where does PIR-1 fit in this model? Unlike mature 22G-RNAs, which have 5' triphosphates, 26G-RNAs have monophosphate 5' ends. Since RNA polymerases typically initiate synthesis with a triphosphorylated nucleotide, we assume that all 26G-RNAs are originally triphosphorylated. Human Dicer is known to cleave dsRNA substrates with different 5' phosphorylation states with the same efficiency *in vitro* (Zhang *et al.*, 2002). Should this also be true for *C. elegans* Dicer, it is unlikely that PIR-1-mediated dephosphorylation is required for Dicer cleavage of the 26G-RNA precursor dsRNA. Instead, PIR-1 could remove phosphates to ensure productive loading into the correct Argonautes. In support of this hypothesis, a variety of studies, particularly in plants (which encode ten distinct Argonautes), have shown that 5'-end nucleotide identity of the sRNA and its physical interactions with the MID domain of Argonautes are crucial to determine which Argonaute it gets loaded into (Mi *et al.*, 2008; Takeda *et al*., 2008; Frank *et al.*, 2010; Frank *et al.*, 2012). One report describing the structure of an Archeal PIWI Argonaute complexed with a 5'-phosphorylated sRNA demonstrated that the MID domain makes contacts with the 5'-phosphate and that eliminating this phosphate lowers the binding affinity of the sRNA to the PIWI protein by about an order of magnitude (Ma *et al.*, 2005). A similar observation was made regarding the 5'-phosphates of sRNAs interacting with human Argonaute-2 (Rivas *et al.*, 2005). It is therefore conceivable that the phosphorylation state of sRNA 5' ends is a major determinant in sorting sRNAs into the appropriate Argonautes. This is particularly important in *C. elegans*, which encodes 26

different Argonautes, in contrast to humans expressing only eight Argonautes and (to date) a less diverse repertoire of endogenous sRNAs.

Therefore, in our model, we consider that in addition to size, 5'-nucleotide identity and 3' methylation, the 5' phosphorylation status is a crucial signature that allows the distinction between different sRNA classes in *C. elegans*, effectively separating 5'-PPP 22G-RNAs secondary siRNAs from 5'-P primary endo- and exo-siRNAs. Lack of dephosphorylation by PIR-1 would thus preclude binding of 26G-RNAs to ALG-3 and ALG-4, resulting in their degradation. This dephosphorylation step could occur at any point after RRF-3 synthesis. However, since we cannot disprove that *in vivo* DCR-1 activity may be influenced by the 5' phosphorylation of its substrates, we also present an alternative model where *pir-1* mutant animals fail to accumulate 26G-RNAs due to the inhibition of Dicer cleavage by the 5'-PPP nucleotide. Although the stability of intact RRF-3 dsRNA precursors in the absence of cleavage is probably low, it remains possible that if PIR-1 dephosphorylation is indeed required for Dicer cleavage, these species may accumulate to high enough levels in the *pir-1* mutant background to allow detection. In the future, this possibility may be worth exploring via RNA-seq strategies that favor the cloning of dsRNA molecules (such as those described in Saldi *et al.*, 2014 and Whipple *et al.*, 2015). The characterization of such precursors could provide us with key insights into the mechanism of action of the ERI complex.

One last consideration pertains to the passenger strand that arises as a byproduct of Dicer cleavage. Normally, the removal of the passenger strand occurs via the slicer activity of the Argonaute during loading (Matranga *et al.*, 2005; Leuschner *et al.*, 2006). Both ERGO-1 and ALG-3/4 Argonautes have intact catalytic residues that presumably remove the passenger strand, as demonstrated previously for RDE-1 (Yigit *et al.*, 2006; Steiner *et al.*, 2009). In *ergo-1* deletion mutant animals, a two-fold accumulation of putative 26G-RNA passenger strands has been observed relative to wild-type animals for the most heavily targeted ERGO-1 locus, *E01G4.5* (Fischer *et al.*, 2011). Intriguingly, this strand does not have the 22-23-nt length with a 3'-receded end predicted by the results from Welker *et al.* (2011), but is rather predominately 18-19 nt, with both a 5'- and a 3'-receded end relative to the 26G-RNA. Although the authors speculate that ERI-1b may be responsible for generating the 5' three-nucleotide receded end, there is no proof that this is the case. In both the *pir-1* and the *alg-3; alg-4* mutant backgrounds we do observe rare instances of sequences antisense to ALG-3/4-dependent 26G-RNAs that match these characteristics, although we do not have evidence that they accumulate (Colin Conine,

unpublished observation and this work; see Fig. 4.12B, page 189 for two 5'-end examples of antisense passenger strands in blue). At the very least, our observations hint that ALG-3/4 passenger strands are identical to those of ERGO-1 in that they are not generated by a canonical Dicer processing mechanism.




**Model A.** PIR-1 is required for productive loading of 26G-RNAs into Argonautes.

**Model B.** PIR-1 is required for cleavage by Dicer to occur.

**Figure 4.14. Models for the involvement of PIR-1 in 26G-RNA biogenesis.** The assembled ERI complex could interact with template RNA in a variety of ways. One interesting possibility is that this happens through Dicer recognition (without cleavage) of regions of secondary structure, perhaps aided by other RNA-binding proteins of the complex (RDE-4, DRH-3, RRF-3, PIR-1 and ERI-1b). Endonucleolytic cleavage 3' of the binding site could occur through the unidentified activity of one of the complex components. Unwinding of the duplex structure would allow translocation of the complex along the mRNA, perhaps through DRH-3 activity. The complex would only stop upon recognition of a cytidine by the RRF-3 polymerase, prompting it to initiate RNA synthesis. In the meantime, ERI-1b exonuclease activity could digest the 3' tail of the cleaved template mRNA. Digestion would halt at the beginning of the new region of dsRNA to give rise to a blunt end consisting of a 3'C-hydroxyl paired with a 5'G-triphosphate (at this point most of the complex proteins were omitted from the figure for clarity). In **Model A**, PIR-1 dephosphorylates the 5'G during or following DCR-1 cleavage (the sites for which are marked by the red arrowheads) to generate a 26G-RNA (green) and a passenger strand with a 2-4 nt 5'-receded end (according to Welker *et al.*, 2011). The duplex is then loaded into the Argonaute and the passenger strand is removed via its slicer activity. In the absence of PIR-1, Dicer would still cleave the dsRNA precursor, but the resulting 5'-PPP 26G-RNA/passenger strand duplex would be unable to stably interact with the Argonaute and would be subsequently degraded. In **Model B**, Dicer cleavage only occurs after PIR-1-mediated dephosphorylation of the 5'G. In the *pir-1* mutant background, 26G-RNAs are never produced and dsRNA precursors are degraded or perhaps accumulate.

## PIR-1 Is Not Required for Secondary Exogenously Triggered siRNA Synthesis

We found that *pir-1* mutants are able to silence mRNAs in the germline and soma in response to dsRNA triggers. Since primary siRNAs are the direct cleavage products of Dicer, they are already monophosphorylated and would therefore not be predicted to require additional modification as 26G-RNAs do. One can argue, however, that when injecting or feeding dsRNA to worms, the 5'-PPP termini are degraded by pervasive phosphatases present in the intestinal lumen and in the animal's body cavity, bypassing any mechanism that would rely on 5'-PPP recognition. Therefore, while we can say with confidence that PIR-1 is not required for the amplification stage of RNAi, in which 22G-RNA effector siRNAs are generated in a Dicer-independent manner, we cannot exclude that the presence of 5' triphosphates in the trigger dsRNA could affect its processing in the *pir-1* background (if their stability could be guaranteed). Additionally, since PIR-1 interacts stably with DCR-1 and RDE-4, and both are required for primary siRNA generation, we cannot entirely discard a role for PIR-1 in the process. Ideally, this problem should be addressed biochemically by studying the processing of different dsRNA substrates with *in vitro* reconstituted DCR-1 primary processing complexes, with and without PIR-1.

### Is PIR-1 Required for Synthesis of ERGO-1 26G-RNAs?

Due to the inability to silence *pir-1* by RNAi, we could not determine whether we could arrest animals at the oogenesis and embryogenesis stages, during which ERGO-1 26G-RNAs are produced. Therefore, we could not make any conclusions about the involvement of PIR-1 in the pathway. The ERGO-1 26G-RNAs we do detect in our experiments are likely of maternal origin and persist in somatic tissues to induce 22G-RNAs against about 80 genes (detailed in Chapter I, page 46; Vasale *et al.*, 2010; Fischer *et al.*, 2011). ALG-3/4- and ERGO-1-dependent 26G-RNAs are differentiated by other features and requirements besides the targets that they regulate and the developmental stage at which they are produced. First, ERGO-1 26G-RNAs, like 21U-RNAs, are methylated at their 3' ends by the conserved HENN-1 methylase, which primarily resides in P granules (Kamminga *et al.*, 2012). Generally, 2'-*O*-methylation has a stabilizing effect on sRNAs by preventing them from being uridylated and targeted for degradation (Kamminga *et al.*, 2010; Ameres *et al.*, 2010). This is consistent with their persistence throughout postembryonic development. Additionally, because the 3'-interacting PAZ domains of some Argonautes have been shown to have higher affinity to 3'-methylated sRNAs (Tian *et al.*, 2011), this molecular signature could have evolved to ensure specificity of interaction with ERGO-1 and exclusion from the ALG-3 or ALG-4 Argonautes. Second, only ERGO-1 26G-RNAs require ERI-9, a putative RNA transferase (Pavelec *et al.*, 2009; Zhuang and Hunter, 2012), and the helicase ERI-6/7 (Fischer *et al.*, 2011). The former interacts with the ERI complex (Thivierge *et al.*, 2012), while the latter does not and is located primarily in the cytoplasm (Fischer *et al.*, 2008). Third, ERGO-1 26G-RNAs specifically require a variety of Mutator proteins, including MUT-2/RDE-3, MUT-15 and MUT-16, which concentrate in Mutator foci adjacent to P granules, where the bulk of WAGO-dependent 22G-RNA amplification occurs (Phillips *et al.*, 2012). Finally, our own results indicate that ERGO-1 26G-RNAs are not enriched in nuclei and have a more uniform distribution along their targets than ALG-3/4 26G-RNAs.

While the above differences highlight important distinctions in how these two classes of primary endo-siRNAs function, they do still retain the same requirement for the ERI complex and a need for dephosphorylation of the 5' end. Since PIR-1 seems to be expressed in the cytoplasm of oocytes and early embryos, it may function during the peak of ERGO-1 expression. However, as PIR-1 also accumulates heavily in nuclei of embryos after the eight-cell stage, it should not be entirely ruled out that ERGO-1 26G-RNA synthesis could also take place in the nucleus. Furthermore, we found that the PIR-1/ERI

complex also assembles during developmental stages that do not yet have a germline (Fig. 3.4, page 124), suggesting that ERGO-1 26G-RNA synthesis may extend beyond embryogenesis. Lastly, we have pulled down ERI-9 in PIR-1 immunoprecipitates, implying that at least a fraction of the complexes incorporate this ERGO-1 26G-RNA-specific factor, the function of which remains to be elucidated. This collection of results strongly implies that ERGO-1 26G-RNAs do require PIR-1 activity.

### The Case for Biogenesis of ALG-3/4 26G-RNAs in the Nucleus

To our knowledge, this study is the first to characterize a nucleus-enriched population of sRNAs in *C. elegans*, despite prior studies demonstrating that 22G-RNA pathways operate in the nucleus. The clearest example pertains to the 22G-RNAs binding to the WAGO-family nuclear Argonautes NRDE-3 (in somatic cells) and HRDE-1 (in the germline). These Argonautes lead to the establishsment of repressive chromatin marks with diverse functional outcomes such as the silencing of ERGO-1 targets, behavioral adaptation to odor, maintenance of heritable gene silencing, and transgenerational maintenance of germ-cell totipotency (Juang *et al.*, 2013; Buckley *et al.*, 2012; Burton *et al.*, 2011; Burkhart *et al.*, 2011; Guang *et al.*, 2010; Guang *et al.*, 2008). The predominately germline-expressed CSR-1 Argonaute also localizes to nuclei where it promotes proper chromosome segregation in the early embryo, represses antisense transcription and ectopic transcription from silent chromatin regions, and assists processing of histone mRNAs (Claycomb *et al.*, 2009; Cecere *et al.*, 2014, Avgousti *et al.*, 2012). Within P granules, CSR-1 can also silence the expression of sperm-specific genes in the adult hermaphrodite germline, while in the exclusively spermatogenic male germline it can activate the expression of genes required for sperm differentiation at the chromatin level (Conine *et al.*, 2013; Campbell and Updike, 2015). Nuclear-acting 22G-RNAs are thought to be produced predominately in the cytoplasm of somatic cells or in germline P granules. It is only upon association with specific Argonautes that 22G-RNAs travel to the nucleus. This was strikingly demonstrated for NRDE-3, which only localizes to nuclei when ERGO-1-dependent 22G-RNAs are produced (Guang *et al.*, 2008). As discussed in Chapter III, however, several examples from other organisms prove or strongly suggest nuclear synthesis of Dicer-dependent sRNAs.

Our results have shown that in contrast to the expected depletion of 22G-RNAs in nuclear extracts, ALG-3/4 26G-RNAs were enriched. Since we know that ALG-3 is mainly localized to the P granules and cytoplasm of cells committed to sperm differentiation, this

unexpectedly indicated that a substantial population of 26G-RNAs resides in the nucleus. Two possibilities thus arise: these RNAs are synthesized in P granules and are subsequently imported into the nucleus, or their biogenesis is entirely nuclear and is followed by export to P granules, possibly in concert with nascent transcripts. In Chapter III we have demonstrated that PIR-1/ERI complexes assemble in the nucleus, and that most of its components also stably interact with chromatin (Fig. 3.8, page 136). Additionally, other studies have shown that ALG-3/4 26G-RNAs can trigger CSR-1-dependent 22G-RNAs, and that both the EGO-1 RdRP (which predominately makes CSR-1 22G-RNAs) and the helicase DRH-3 also localize to the nucleus and chromatin (Maine *et al.*, 2005; Claycomb *et al.*, 2009; Conine *et al.*, 2013). It is therefore plausible that at least a small population of ALG-3/4-bound 26G-RNAs could lead to 22G-RNA production without ever leaving the nucleus. However, given that ALG-3 concentrates in P granules, most 26G-RNA targeting of mRNAs probably takes place in those regions. In contrast, our experiment showed that ERGO-1 26G-RNAs were not enriched in nuclei. This could result from a combination of two factors: (1) the accumulation of the 22G-RNAs that they trigger require proteins located in cytoplasmic Mutator foci (Zhang *et al.*, 2011; Phillips *et al.*, 2012); and (2) the animals used for sequencing were not in the developmental stage during which most ERGO-1 26G-RNA synthesis occurs.

When we analyzed the distribution of 26G-RNAs along mRNAs, we observed that the majority of ALG-3/4 26G-RNAs are derived from the 5' regions of mRNAs (Fig. 4.12, page 189). Although we cannot yet explain the reason for this asymmetry in sRNA distribution along targets, it is tempting to speculate that the preferential accumulation of ALG-3/4 26G-RNAs at mRNA 5' ends is related to a predominant synthesis in the nucleus. Perhaps as a way of ensuring tight regulation of spermatogenesis mRNAs, 26G-RNAs are synthesized co-transcriptionally, as nascent transcripts emerge from Pol II complexes. Since there is a strong depletion of ALG-3/4 26G-RNAs in the middle of genes relative to the ends, it is plausible that mRNA-processing RNPs such as the spliceosome physically block access of the ERI complex to these regions. Consistent with this idea is the fact that shorter genes with no introns, such as the abundant mRNAs for major sperm proteins, have a more uniform distribution of ALG-3/4 26G-RNAs along their sequences. In contrast, 22G-RNAs, which are primarily made in the cytoplasmic P granules and Mutator foci, abundantly cover the entire length of their target mRNAs. In support of PIR-1 operating in the nucleus, the 5' bias disappears completely in the few 26G-RNAs that remain in *pir-1* arrested animals, which likely reside in P granules, bound to ALG-3 and ALG-4 (Fig. 4.12,

page 189). This suggests that 26G-RNAs only accumulate in the nucleus while they are being synthesized, to be subsequently exported to P granules for triggering of 22G-RNA production. Under this scenario, in wild-type hermaphrodites undergoing active spermatogenesis, nuclear synthesis of 26G-RNAs would outpace the rate of export and accumulation in P granules, leading to a temporary nuclear retention of 26G-RNAs. In Figure 4.15 we present a model illustrating this process.

As an important note, it must not be forgotten that in these experiments we are not capturing the sRNA distribution along individual transcripts. Rather, for each mRNA, we are considering sRNAs derived from a population of transcripts expressed in a variety of cells with distinct gene expression outputs. Each type of transcript is likely subjected to very different regulatory schemes depending on which point along germline development it is expressed. In other words, the same gene in early spermatogenesis may be regulated by a set of sRNAs completely distinct from the set of sRNAs present in late spermatogenesis. It is therefore plausible that 5'- and 3'-26G-RNAs derived from the same sequence may actually have been produced in different compartments and/or distinct differentiation stages, imparting very different fates on the mRNAs that they target. Since PIR-1 and ALG-3 also localize to the syncytial cytoplasm of early spermatogenesis, the synthesis of 26G-RNAs is likely to also take place there, perhaps to synthesize predominately 3' species. Further experimental exploration of these differences could potentially help us gain a more detailed understanding of how sRNAs contribute to the precise orchestration of highly complex and dynamic biological processes such as sperm cell differentiation.

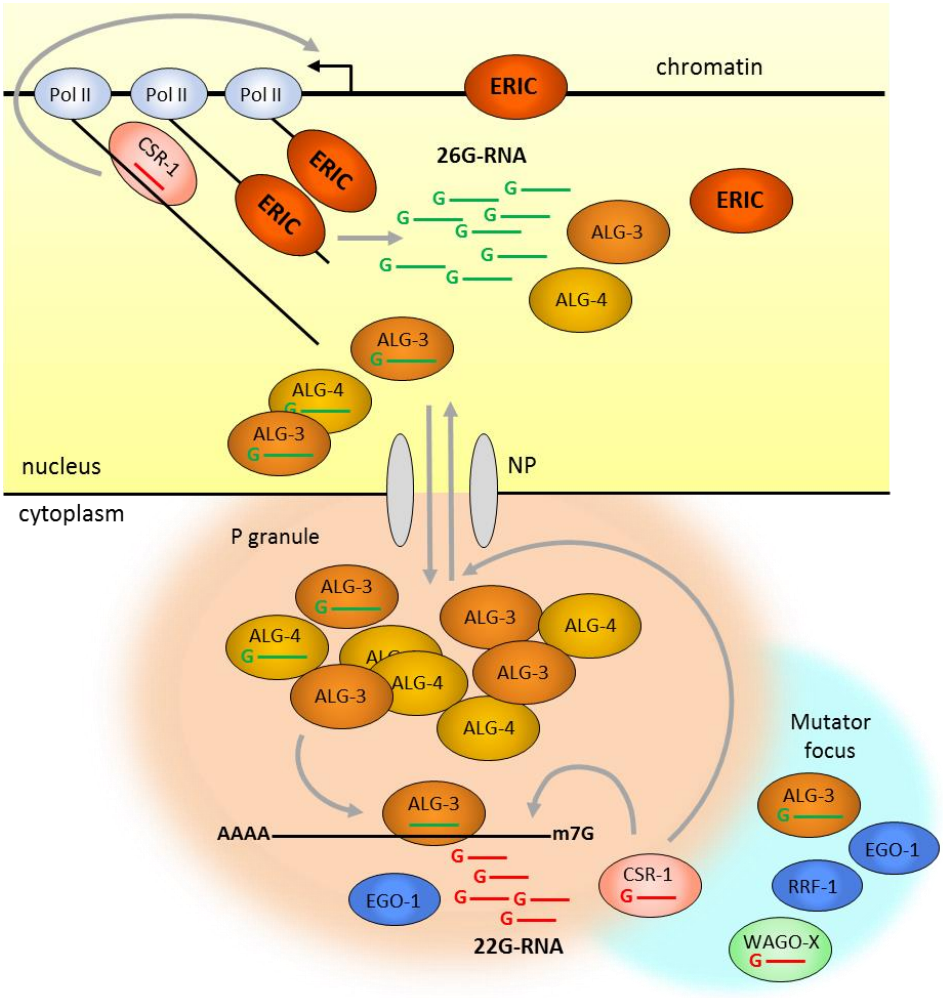



**Figure 4.15. Model for ALG-3/4 26G-RNA biogenesis in the nuclei of germ cells.** Our data are compatible with a model where the ERI complex (ERIC) synthesizes 26G-RNAs in the nucleus. Nuclear and chromatin-associated ERI complexes engage nascent RNA Pol II transcripts in the vicinity of nuclear pores (NP) to use them as RdRP templates, in competition with other nuclear mRNA processing factors for binding. ERIC activity leads to both cleavage and production of 26G-RNAs from the 5' ends of the nascent transcripts. A small population of empty ALG-3 and ALG-4 Argonautes localizes to the intranuclear perifery of the NP to allow loading and stabilization of 26G-RNAs. These complexes are subsequently exported to NP-associated P granules in the cytoplasm, where they recognize target RNAs to promote synthesis of effector 22G-RNAs. Alternatively, loaded ALG-3/4 Argonautes could exit through the nuclear pore in association with intact mRNAs. The enrichment of 26G-RNAs in the nucleus is due to an elevated rate of synthesis during spermatogenesis which outpaces the rate of nuclear export of loaded Argonautes. The majority of ALG-3/4 proteins, however, still reside at their site of action in cytoplasmic P granules. There, loaded ALG-3/4 can trigger the synthesis of 22G-RNAs by the EGO-1 RdRP, which in turn associate with the CSR-1 Argonaute to regulate spermatogenic gene expression at the post-transcriptional and transcriptional levels. Additionally, ALG-3/4 may also trigger the production of WAGO-class 22G-RNAs in Mutator foci, closely associated with P granules.

## PIR-1 May Promote the Synthesis of Endogenous 22G-RNAs Targeting Protein-Coding Transcripts

We observed a significant decrease in the accumulation of 22G-RNAs antisense to protein-coding genes in *pir-1* arrested animals (Fig. 4.4, page 173 and Fig. 4.6, page 176 gray bars and gray line). We initially attributed this effect to a lower germline-to-soma ratio compared to wild-type control animals. However, as the levels of germline 21U-RNAs do not significantly change in *pir-1* mutant animals, the aforementioned argument is a weak one. Rather, we justify this result with the possibility that PIR-1 promotes the accumulation of 22G-RNAs in the germline. Within the WAGO-pathway, we demonstrated that 22G-RNAs targeting transposons do not exhibit a decreased abundance in *pir-1* mutant animals (Fig. 4.5, page 175), although WAGO 22G-RNAs are globally downregulated. This downregulation therefore occurs mainly at the expense of 22G-RNAs targeting protein-coding transcripts in the context of the WAGO pathway. Regarding CSR-1 22G-RNAs, they exhibit a more pronounced reduction in *pir-1* mutant animals, consistent with the fact that all identified CSR-1 targets are protein-coding transcripts. Since *csr-1* mRNA levels remained at wild-type levels in both three- and ten-day arrested *pir-1* animals, the 22G-RNA reduction is unlikely to be caused by decreased CSR-1 protein (mRNA-seq data; not shown).

The work of Conine *et al.* (2013) demonstrated that in male animals a subset of ALG-3/4 target genes required for spermatogenesis and spermiogenesis is directly targeted by CSR-1 22G-RNAs to promote their expression. This implied that ALG-3/4-bound 26G-RNAs can act as primary triggers for the synthesis of a subset of CSR-1 22G-RNAs. Since *pir-1* arrested animals exhibit a marked decrease in ALG-3/4 26G-RNAs, it is also plausible that the extra downregulation of CSR-1 22G-RNAs in the *pir-1* mutant reflects the reduction of this specific set of 26G-RNAs. We found that from the 1,852 genes for which CSR-1 22G-RNAs are downregulated more than two-fold in *pir-1* males (with a 10 rpm cutoff), merely 183 overlap with the 1,156 genes targeted by ALG-3/4-dependent CSR-1 22G-RNAs (from Conine *et al.*, 2013). It is therefore unlikely that the heavy reduction of CSR-1 22G-RNAs in *pir-1* arrested animals is due to the interaction between the CSR-1 and the ALG-3/4 pathways.

The inability to study the effects of PIR-1 loss in more advanced developmental stages (*e.g.*, by RNAi of *pir-1*) prevented us from adequately addressing this interesting result with further experiments. It is difficult to envision how PIR-1 would be able to promote 22G-RNA accumulation, as these RNAs are 5'-triphosphorylated in their active

form. Perhaps PIR-1 functions upstream, by regulating the production of primary sRNAs to trigger the production of all 22G-RNAs targeting protein-coding sequences. Since we still do not understand how most endogenous 22G-RNAs are triggered, the possibility that PIR-1 may play a role in the process should be seriously considered in future studies. Finally, RNAi of *csr-1* has long been known to impair the proliferation of germ cells (Maine *et al.*, 2005; Vought *et al.*, 2005; Duchaine *et al.*, 2006; She *et al.*, 2009). It is thus tempting to speculate that a possible cause for the underproliferation we observe in the germlines of *pir-1* mutants could be directly attributed to defects in processes regulated by CSR-1. Should this be true, it would still fall short of explaining why PIR-1 is also required for somatic cell division and growth.

## Potential for Conservation of PIR-1 Small RNA-Related Functions in Other Animals

Given that PIR-1 is highly conserved among metazoans, it is pertinent to ask whether other animals could require this enzyme for endo-sRNA production. In *Drosophila* and mice, endo-sRNA studies have been typically restricted to miRNAs and piRNAs, the latter with a prevalent role in the suppression of transposition in the germline. In recent years, however, several groups have reported the discovery of many other types of Dicer-dependent endo-sRNAs processed from a variety of dsRNA substrates, such as transcripts with inverted repeats or hairpins, convergent transcription, and pairing of transcripts expressed in *trans* (reviewed in Piatek and Werner, 2014; Svoboda , 2014). In the mouse, endo-siRNAs have been cloned from embryonic stem cells, oocytes and male germ cells where they were shown to regulate gene expression (Watanabe *et al.*, 2008; Tam *et al.*, 2008; Babiarz *et al.*, 2008; Song *et al.*, 2011). Whether silencing by such endo-sRNAs occurs transcriptionally, post-transcriptionally, or at both levels, however, is still an open question. Interestingly, all of the aforementioned cell types are similar in their insensitivity to the dsRNA-induced interferon response that often results in cell death (Wianny and Zernicka-Goetz, 2000; Yang *et al.*, 2001; Stein *et al.*, 2003). These cells are therefore able to accumulate sufficient dsRNA to be processed by Dicer, and thus add another layer of regulation to coordinate their complex differentiation gene expression programs, concurrently with the miRNA and piRNA networks. Vertebrate PIR1 could conceivably play a regulatory part in the processing of endogenous dsRNA via its association with Dicer, analogous to the activity we report for *C. elegans* germline endo-sRNA biogenesis.

Fungi and plants lack PIR-1 orthologs but, similarly to *C. elegans*, generate endo-sRNAs through Dicer cleavage of RdRP-generated dsRNA intermediates. With the exception of fungi, in which Dicer products are the effector sRNAs (Dang *et al.*, 2011), plant and nematode Dicer-dependent sRNAs trigger an RdRP-mediated secondary amplification step that produces effector sRNAs (reviewed in Ghildiyal and Zamore, 2009). In *C. elegans,* effector 22G-RNAs can be triggered by a variety of sRNA species, including Dicer-dependent 26G-RNAs, Dicer-dependent RDE-1-associated sRNAs, and Dicer-independent 21U-RNAs/piRNAs (Table 1.1, page 29; Pak and Fire, 2007; Sijen *et al.*, 2007; Das *et al.*, 2008; Gent *et al.*, 2010; Vasale *et al.*, 2010; Lee *et al.*, 2012; Ashe *et al.*, 2013). The effector endo-sRNAs described for *Drosophila* and mice (both of which not appearing to encode RdRPs) are similar to those of yeast, in the sense that Dicer generates them directly by cleavage of dsRNA precursors. By bypassing the requirement for an amplification step, these organisms obviate the need for RNA species equivalent to the nematode-specific monophosphorylated 26G-RNAs. Nonetheless, future investigations of PIR-1 homologs in *Drosophila* and vertebrates should not disregard potential changes in the composition of endogenous sRNA populations in the absence of this RNA phosphatase.

## MATERIALS AND METHODS

### RNAi of *sel-1*, *unc-22*, *gfp,* and *pos-1*

The HT115 *E. coli* strains used for RNAi silencing in this chapter carried plasmids with the sequences of interest flanked by T7 promoters. The *sel-1* construct consisted of the *Bgl*II restriction fragment of the respective gene (from Duchaine *et al.,* 2006). The *unc-22* construct carried a 40-nt sequence corresponding to a coding portion of the *unc-22* gene repeated five times in tandem (from Yigit *et al*, 2006). The *gfp* construct also expressed dsRNA from a tandem repeat of five short identical sequences from the *gfp* open reading frame (Pedro Batista, unpublished). The *pos-1* RNAi feeding strain was taken from the Ahringer Lab RNAi library (Kamath *et al.,* 2003). Bacterial culture, induction and preparation of RNAi plates was performed as described in Chapter II.

### Extraction of RNA

For large worm populations, RNA extraction was performed as described in the Materials and Methods section of Chapter II. For extraction of RNA from a single animal or a small number (10-20) of animals (as performed for deep-sequencing of male sRNA), live worms were washed several times with M9 buffer containing 0.5% Tween-20 using a handmade glass capillary mouth pipette. Alternatively, 100-200 worms were washed 3-4x in PCR tubes by successive suspension in M9 buffer and pelleting using a mini centrifuge. Clean individual worms were transferred to 0.25 ml PCR tubes, centrifuged, and all but 2-3 µl of M9 was removed, making sure that the worms were left in the tube. Per tube, 10 µl of lysis buffer were added (40 mM Tris-HCl pH 7.5, 10 mM EDTA, 300 mM NaCl, 0.5% SDS, 0.4 µg/µl proteinase K) and incubated at 45-65˚C for 5-10 minutes. After checking that the worms had been lysed under a dissecting microscope, 100 µl of TE buffer and 20 µg of glycoblue (Ambion) were added. For a larger quantity of worms (100-200), 50 µl of lysis buffer and 50 µl of TE buffer were used. One volume of phenol:chloroform was then added and mixed. Adapted phase-lock tubes were made by transferring gel from Phase-Lock tubes (Eppendorf) to a PCR tube with a pipette tip (~20-30 µl) and spun at 10,000x *g* for 2 minutes in adaptor tubes. The lysis/phenol mixture was transferred to the tube, remixed by tapping, and centrifuged at 10,000x *g* for 4 minutes at room temperature. The aqueous phase was transferred to a fresh tube with 10 µl of 3 M sodium acetate pH 5.2, followed by addition of 130 µl of isopropanol and precipitation at -20˚C for at least 30 minutes. RNA was pelleted at 15,000x *g* for 15 minutes at 4˚C, washed with cold 75%

ethanol and resuspended in 5-10 µl of ultrapure water. For sRNA cloning, this RNA was run on a 15% polyacrylamide/0.5X TBE gel and stained with SYBR Gold dye for increased sensitivity. At this scale, sRNAs are recovered in the picogram (femtomole) range and are therefore difficult to visualize under UV light.

## qRT-PCR and Calculation of Standard Error of the Mean

cDNA synthesis and qRT-PCR were performed as described in Chapter II. Additional primers for *pos-1* mRNA and for validation of mRNA-sequencing target mRNAs can be found in Appendix B (O95-O114). For experiments performed in triplicate, we reported the standard error of the mean (SEM), instead of the standard deviation (SD), because it provides a better indication of the uncertainty associated with the estimate of the mean, namely by taking into account the sample size (*i.e.*, number of replicates). The SEM was calculated as follows: the quantity of mRNA in each sample, as estimated from a standard curve with serial dilutions of cDNA, was normalized by dividing it to the quantity of actin or *gapdh* mRNA. The mean of the three normalized quantities was used to calculate the SD associated with each group of replicates. The relative fold change was calculated by dividing the mean quantities of mutant or treated samples to the mean quantity of the wild-type or untreated sample (reference). The reference quantity is divided by itself to become one, making it meaningless to report an associated SEM, as all other samples are now compared to a standard value which we assume not to vary. For the remaining samples the SD was calculated by the using the formula for propagation of SD for multiplication or division $SD_{div} = x\sqrt{((SD_{rf}/mean_{rf})^2 + (SD_{sp}/mean_{sp})^2)}$, where *x* is the calculated relative fold value of the sample, *rf* pertains to the reference and *sp* to the sample for which the SD is being calculated. The SEM was then calculated using the formula $SE = SD_{div}/\sqrt{(\text{sample size})}$.

## Northern Blotting of Small RNAs

Small RNAs were enriched by using the MirVana sRNA isolation kit (Ambion) skipping the binding column step. For this, up to 1 mg of total RNA in an 80 µl volume was mixed with 400 µl of MirVana lysis/binding buffer and 48 µl of MirVana homogenate buffer, mixed and incubated for 5 minutes at room temperature to denature RNA. A 1/3 volume (176 µl) of ethanol was added, mixed and centrifuged at 2,500x *g* for 4 minutes at room temperature to pellet large (>200 nt) RNA. The supernatant was transferred to a new tube, mixed with 1 volume of isopropanol and incubated at -20˚C for at least 30 minutes.

RNA was pelleted by centrifugation at 20,000x *g* at 4˚C for 15 minutes and washed once with cold 75% ethanol. RNA was dissolved in either 50 µl TE buffer or ultrapure water depending on downstream treatment. Quantification was performed using a NanoDrop spectrophotometer. The sRNA yield using this protocol was typically ~8% of total RNA.

Between 5 and 20 µg of sRNA-enriched RNA was denatured in formamide-containing RNA gel loading buffer II (Ambion) in a maximum volume of 20 µl to ensure good electrophoretic resolution. In cases where RNA was too dilute, the appropriate amount of RNA was mixed with loading buffer and excess volume was evaporated in a SpeedVac instrument at a low temperature setting. Samples were separated on a denaturing 15% polyacrylamide/7 M urea/0.5X TBE buffer gel (17 cm wide and 15 cm long) at 15 W (to reach a maximum of 700 V) until the xylene cyanol dye reached approximately half the length of the gel. The gel was then stained with ethidium bromide in 0.5X TBE buffer for 10 minutes and destained for 10-15 minutes before visualization with UV light to assess integrity of the RNA. For samples with low degradation, a sRNA band corresponding mostly to 22G-RNAs could be easily distinguished. Large, in-sample endogenous size markers include 5.8S rRNA (~160 nt), 5S rRNA (~120 nt) and tRNA (~70-80 nt). RNA was transferred to a Hybond N+ charged nitrocellulose membrane (Amersham) with 0.5X TBE buffer using a semi-dry TransBlot apparatus (Bio-Rad) for 1 hour at 25V, and crosslinked with UV in a Stratalinker instrument (Stratagene). In order to minimize non-specific probe binding, membranes were pre-treated with 10 ml of Ultrahyb Oligo buffer (Ambion) for 30 minutes at 42˚C in hybridization flasks undergoing constant rotation in a hybridization oven. Probes against different targets consisted of labelled custom Starfire oligos (IDT) labelled with [α-32P]-dATP according to the manufacturer's instructions and purified with Micro Bio-Spin P-30 chromatography columns (Bio-Rad). Probes were added to the buffer used for pre-hybridization blocking, and incubated overnight at 42˚C. Washes were 2x 5 minutes with 2X SSC/0.1% SDS, followed by 2x 15 minutes with 0.1X SSC/0.1% SDS, all at 42˚C (20X SSC buffer is 3 M NaCl and 300 mM sodium citrate dehydrate). Membranes were wrapped in plastic cling wrap and exposed to a Phosphor screen (GE) for a few hours to 2 days (depending on the RNA target) and scanned with a Phosphorimager instrument (GE). For multiple hybridizations, membranes were stripped with 0.1% SDS at 85˚C with constant shaking for 45 minutes. After cooling, membranes were re-blocked and hybridized with a new probe. The sequences of the Starfire probes used can be found in Appendix B (O86-O94).

### Treatment of RNA with Terminator 5' Phosphate-Dependent Exonuclease

For Terminator 5' phosphate-dependent exonuclease (Epicentre) treatment of RNAs for northern analysis, 6-10 µg of MirVana-purified RNA resuspended in ultrapure water was digested with 2 Units of Terminator and 1X Terminator Buffer A in a total volume of 15 µl. In the negative controls, ultrapure water was used instead of enzyme. Reactions were incubated at 30˚C for 5 hours, and stopped by adding 10 µl of formamide RNA loading buffer.

### Small RNA Cloning and Deep-Sequencing

Small RNAs in this study were cloned by one of two methods: 5'-independent or 5'-dependent. In the 5'-independent method, RNA was treated with Tobacco Acid Pyrophosphatase (or TAP; Epicentre), which removes 5' cap moieties and terminal phosphates, leaving only one 5'-terminal phosphate. Since it does not alter sRNAs that were originally 5'-monophosphorylated, this approach clones the widest range of sRNA types. This method constitutes an improvement from an earlier method which completely dephosphorylated all RNAs using CIP, followed by the addition of one 5' phosphate by T4 Polynucleotide Kinase (PNK). This had the disadvantage of allowing the cloning of unphosphorylated RNA degradation fragments, lowering the representation of sequences of interest in the sRNA libraries. The use of TAP minimizes this problem. In the second, 5'-dependent approach, also known as direct cloning, sRNAs are not subjected to any enzymatic treatment prior to 5' adaptor ligation leading to preferential cloning of 5'-monophosphorylated sRNAs. The direct cloning method is therefore suited for analysis of Dicer products such as miRNAs, primary sRNAs (including 26G-RNAs and viral-23-mers), and 21U-RNAs.

Generally, 10 µg of total RNA (the sRNA fraction in *C. elegans* total RNA is ~0.01%) in TE buffer (10 mM Tris-Cl pH 7.5, 1 mM EDTA) was denatured in formamide loading buffer II (Ambion) at 65˚C for 5 minutes and run on an 18-well (~0.5 cm wide) denaturing gel (15% polyacrylamide/7 M urea/0.5X TBE buffer) poured into a Bio-Rad Criterion disposable cassette. The loaded volume was kept to a minimum (10 µl or less, ideally). For size markers, three RNA oligos (18, 26 and 40 nt; 1 µl of 1 µM) were used. Samples were run at 20W until the bromophenol blue dye migrated two-thirds the length of the gel (7 cm long). The gel was then stained with SYBR Gold dye (Invitrogen) in 0.5X TBE buffer for 10 minutes and destained in 0.5X TBE buffer for 5-15 minutes. The gel

was visualized with UV light and each lane was sliced just below the 18 nt position and between the 26 and 40 nt positions to include the desired 18-30 nt range. Gel fragments were transferred to 1.5 ml siliconized tubes (used throughout the rest of the procedure) and crushed to fine pieces with a 1 ml pipette tip, to which 750 µl of elution buffer (0.3 M NaCl in 1X TE buffer) was subsequently added. Elution occurred overnight at room temperature with constant rotation or shaking in a thermomixer. Eluates were filtered using Spin-X 0.45 µm filters (Costar) for 1 minute at 10,000x *g*. After transfer to fresh tubes, 20 µg of glycogen and 1 volume of isopropanol were added and precipitation was allowed to occur at -20˚C for at least 30 minutes. RNA was pelleted by centrifugation at 20,000x *g* at 4˚C for 15 minutes and washed once with cold 75% ethanol. All ethanol was thoroughly removed and each pellet was allowed do dry for about 1 minute (residual ethanol interferes with the TAP reaction). 10 µl of TAP reaction mix (1X TAP buffer, 1 U/µl Superasin RNAse inhibitor (Ambion), 0.25 U/µl TAP) was added per pellet, making sure the RNA completely went into solution, followed with mixing by pipetting and transfer to a PCR tube. Reactions were incubated for 1 hour at 37˚C, after which the volume was brought to 100 µl with ultrapure water and extracted with phenol:chloroform at room temperature. The aqueous phase was recovered by centrifugation in Phase-lock tubes (Eppendorf) and RNA was precipitated by the addition of a $1/10^{th}$ volume of 3 M sodium acetate, pH 5.2, under the conditions described above (glycogen added previously remained in the aqueous phase). For direct cloning, dried RNA pellets after the first precipitation were used for 3' ligation straight away, as explained below.

The washed, dried pellets were then subjected to the ligation of the 18-nt 3' linker. This RNA oligo is “activated” through adenylation of the 5' end (rApp), which results in extremely efficient ligation and precludes the use of ATP. At the 3' end it possesses a dideoxy cytidine modification (ddC) in order to prevent ligation between linkers. For the reaction, a T4 RNA ligase buffer without ATP was prepared (10X buffer is 0.5 M Tris-HCl pH 7.5, 0.1 M $MgCl_2$, 0.1 M DTT). To each pellet, 10 µl of the reaction was added consisting of 1X ligation buffer, 1U/µl Superasin, 0.1 mg/ml BSA, 2 U/µl T4 RNA ligase (Takara), 5 µM 3' linker, and 10% DMSO. A no-reaction control was performed in parallel to monitor the extent of ligation between linkers, which can occur to some extent because the 3' ddC modification is not present in 100% of the molecules. After mixing, the reaction proceeded in a PCR machine at 15˚C for 2 hours and then 4˚C overnight. The ligation mixtures were then mixed with RNA loading buffer (with no need for heating) and run on a denaturing 15% polyacrylamide gel as before, but letting the bromophenol blue dye reach

the bottom. As size markers, 1 µl of a 1 µM mixture of DNA oligos with sizes representing different ligation product sizes (33, 48, 50, 54, 65, and 69 nt for both 3' and 5' ligations) were run in parallel. The gel was stained with SYBR Gold as above. Ligated products were recovered from gel slices cut between the 33 and 54 nt markers as described above, except that 20 µg of glycogen were added before precipitation.

For 5' ligation a 21-nt linker was used. It consists of a hybrid oligo, in which the first 5 nucleotides are DNA and the remaining sequence is RNA. The last 4 nucleotides comprise the variable barcode used to multiplex several samples in one deep-sequencing run. Again, each dried pellet was resuspended in 10 µl of ligation mixture containing 1X T4 ligation buffer with ATP, 1 U/µl Superasin, 0.1 mg/ml BSA, 2 U/µl T4 RNA ligase, 5 µM 3' linker, and 10% DMSO. Incubation occurred at 15˚C for 6 hours and then at 4˚C overnight. Ligation products were separated in a gel as described for the 3' ligation procedure, and recovered from slices cut just below the 54 nt marker and ~0.5 cm above the 69 nt marker. RNA was recovered as in the 3' ligation. cDNA synthesis followed, using a DNA oligo complementary to the 3' linker. For this, pellets were resuspended in 13 µl of 0.5 mM dNTPs and 2 µM of RT (reverse transcription) oligo, and incubated 5 minutes at 65˚C, followed immediately by incubation on ice for 2 minutes. The remaining components were then added on ice for a final volume of 20 µl with the following final concentrations: 1X first-strand buffer, 10 mM DTT, 0.5 U/µl Superasin, 0.5 U/µl Superscript III reverse transcriptase (Life Technologies). The reactions were incubated at 50˚C for 1 hour, 85˚C for 5 minutes, followed by 20 minutes at 37˚C after adding 1 µl (2 U) of RNase H (Life Technologies).

To produce the PCR products for sequencing, two rounds of PCR were performed: a first round with short oligos complementary to the linkers, and a second round with longer oligos containing the linker sequences plus the Illumina P5 sequence at the 5' end or the Illumina P7 sequence at the 3' end, required for attachment of the molecules to the flow cells and sequencing by synthesis in the Illumina instrument. The shorter primers are used because the longer primers tend to generate a significantly higher amount of primer dimers if used during the whole amplification reaction. For the first PCR, a 50 µl reaction per sample was assembled with 2 µl of cDNA, 1X ExTaq buffer, 0.5 µM of each short oligo, 0.25 µM dNTPs, and 0.025 U/µl ExTaq polymerase (Takara). Cycling conditions were as follows: 94˚C/30 seconds, at least 5 cycles of 94˚C/20 seconds, 55˚C/20 seconds, 72˚C/20 seconds, and stopped at 4˚C. For the second PCR the following components were added to the first: 4 µl of ultrapure water, 1 µl of 10X buffer, and 2.5 µl of 10 µM solutions of each

long oligo (2.5 µM f.c.). The PCR was then resumed with the same cycling conditions as above, taking 3 µl every two cycles (2, 4, 6, and 8 or higher cycle numbers depending on the amount of starting material). These samples were mixed with DNA loading dye and loaded onto a non-denaturing 8% polyacrylamide/0.5X TBE buffer gel along with a 10 bp DNA ladder (5 µl of 0.1 µg/µl; Invitrogen). The xylene cyanol dye was allowed to reach 2.5 cm from the bottom of the gel. The gel was stained with ethidium bromide in 0.5X TBE for 10 minutes and destained for 10-20 minutes. The optimal cycle number for each sample was chosen based on the highest amount of 110 nt amplicons produced and the least amount of larger, bulged products. The latter products form when primers begin to be depleted and full-length sequences denature and anneal at the linkers to generate larger products that are not fully hybridized. Once the best PCR conditions were found, 3x 50 µl reactions for each sample were prepared, pooled and extracted with phenol:chloroform. The aqueous phase was recovered using phase-lock tubes and concentrated to a ~30 µl volume by centrifugation in Amicon Ultra 0.5 ml filters (EMD Millipore) at 13,500x *g*, 4˚C for 20 minutes, but checking the sample every 3 minutes after 12 minutes of centrifugation. The concentrated products were then mixed with 6X DNA loading buffer and loaded onto two wells of a non-denaturing 8% polyacrylamide/0.5X TBE buffer gel and run and stained as above. The DNA band around 110 bp was excised, being careful to avoid primer dimers below. At this point the DNA may easily denature if excessive drying occurs, so all steps were performed as quickly as possible. The crushing of each gel slice was immediately followed by addition of elution buffer and eluted as described above. Precipitation was performed with 20 µg of glycogen and 1 volume of isopropanol at -20˚C for at least 30 minutes. After pelleting and removing the isopropanol, cold 70% ethanol was immediately added. After centrifugation, isopropanol was completely removed and 15 µl of 10 mM Tris-Cl pH 7.5 buffer was added immediately without having allowed the pellet to dry. DNA concentrations were measured using a NanoDrop spectrophotometer. The typical concentration range obtained was 50-70 ng/µl.

For quality control of the libraries before sequencing, 2 µl of each one was treated with Taq polymerase (Roche) to create adenine 3' overhangs and allow cloning into a TOPO-TA vector (pCR-2.1 TOPO vector; Invitrogen). The 4.5 µl reaction included 1X Taq buffer, 0.25 mM dNTPs, and 0.5 U of Taq polymerase, and was incubated at 72˚C for 15 minutes. Then 1 µl of salt solution from the TOPO kit and 0.5 µl of TOPO-vector enzyme mix were added, incubated at room temperature for 30 minutes and transformed by heat shock into chemically competent *E. coli* (E. cloni 5-alpha from Lucigen). Cells were

plated on ampicillin/X-gal plates, grown overnight at 37˚C, and screened for blue/white colonies. White colonies were picked and subjected to PCR in a 20 µl mixture of 1X Roche Taq buffer, 0.2 µM M13 reverse and M13 forward primers, 0.25 mM dNTPs and 0.05 U of Taq polymerase (Roche). Cycle conditions were 94˚C/120 seconds, 30 cycles at 94˚C/20 seconds, 50˚C/20 seconds, 72˚C/30 seconds, and stopped at 4˚C. Size and purity of the products was checked on a 1% agarose/1X TAE gel and each reaction was cleaned using the E.Z.N.A. extraction kit (Omega). After quantification, about 5-10 PCR products per sample were sent for Sanger sequencing using a T7 primer. Sequences were checked for the appropriate barcode, as well as for high frequency of *C. elegans* sequences matching protein coding genes (in the case of TAP cloning) or miRNAs and 21U-RNAs (for direct cloning) using BLAST. Libraries were mixed together for a 1 ng/µl final concentration and submitted to the UMass Deep-Sequencing Core Facility where the sample quality was verified in a Bioanalyzer and sequenced on either an Illumina Genome Analyzer II or Hi-Seq 2000 instrument.

For cloning from single or just a few worms (pg/femtomole range), the following adjustments were made: TAP was used at 0.1 U/µl; 3' and 5' ligations were performed with half the linker amount (25 pmole instead of 50 pmole) with half the amount of T4 RNA ligase (1 U/µl); after 3' ligation, the xylene cyanol was run until it was 2 cm from the bottom of the gel for complete separation from 3' adaptor dimers (crucial because the starting amount of RNA was extremely low); for cDNA synthesis the RT primer amount was reduced from 20 to 10 pmole and Superscript III contentration was reduced to 2 U/µl; finally, the first round of PCR consisted of 20 cycles.

The sequences for all oligos used above are listed in Appendix B (O115-O124).

### mRNA Cloning and Deep-Sequencing

mRNA was isolated from total RNA using the Promega PolyATract mRNA Isolation System IV according to the instructions of the manufacturer. ~100 ng of mRNA was partially degraded by alkaline hydrolysis using 0.2 M of $Na_2CO_3$/$NaHCO_3$ (1:1 ratio), and RNA fragments of 100-200 nt were excised from a denaturing 10% polyacrylamide/ 7 M urea/0.5X TBE gel and purified as described above. The purified RNA was treated with 10 U of CIP (NEB) for 3 hours at 37˚C in a 40 µl reaction, phenol:chloroform extracted, and precipitated at -20˚C with $1/10^{th}$ volume of 3 M sodium acetate, pH 5.2 and 3 volumes of ethanol, followed by pelleting and washing as described previously. Subsequently, the RNA was phosphorylated at the 5' ends with polynucleotide kinase (T4

PNK; NEB) in a 40 µl reaction consisting of 1X PNK buffer, 0.5 U/µl Superasin, 0.4 mM ATP, and 0.5 U/µl T4 PNK, incubated for 1 hour at 37˚C. The RNA was again phenol:chloroform extracted, and precipitated. As described for sRNA cloning, the RNA was ligated with the 3' linker, reverse-transcribed, and ligated with the barcoded 5' linker. Illumina adaptors of full size were added during the PCR step. The final amplicons were sequenced using an Illumina Genome Analyzer II set to sequence 100-nt single reads.

### Analysis of Small RNA and mRNA Deep-Sequencing Libraries

Illumina sequencing reads of 36-100 nt were sorted according to the barcode at the 5' end of each read to de-group the mixed sequences into individual samples, and then the barcodes were removed. Because the 3' end of each read could contain the 3' linker sequence used to build Illumina libraries, our script searched for CTGTAG – the beginning of the 3' linker – and removed this and any other upstream sequence. This trimming process was achieved using a custom PERL (5.10.1) script. Mapping of reads of at least 17 nt to the annotated *C. elegans* genome sequence (WormBase WS215) was performed with Bowtie 0.12.7 (Langmead *et al.*, 2009) with the parameters -v 3 -a --best --strata -m 400. Reads smaller than 17 nt often match multiple genomic loci, and some of these are derived from contaminant *E. coli* RNA that can still match the *C. elegans* genome. These smaller sequences were therefore excluded from the analysis. The parameter '-v 3' allows maximal 3 mismatches in the seed region; '-a --best --strata' returns the best matches within the seed region, using 26 as default; '-m 400' pertains to the filter used to remove reads that match more than 400 loci (repetitive loci). Additionally, our script also maps reads derived from exon-exon splicing junctions.

A custom PERL pipeline was used for post-matching analyses. A maximum number of mismatching nucleotides per read size was allowed as follows: 0 mismatches for size 17–18 nt, 1 for 19–23 nt, 2 for 24–30 nt, and 3 for longer than 30 nt. Again the parameters were used to achieve specificity and high confidence of matches. For RNA reads that mapped to multiple loci, the read number was split evenly among them. The read number was then normalized to 1 million miRNA reads, or 5 million nonstructural sRNAs including miRNAs, 21U-RNAs, and 22G-RNAs, depending on the experiments. The script also summarized the sense and antisense reads derived from each protein-coding gene and non-coding gene. Custom scripts were used to analyze the first nucleotide frequency and the size distribution of reads. The Generic Genome Browser GBrowse 1.70 (Stein *et al.*, 2002) was used to visualize the alignments as first nucleotide histograms, which were

displayed using a log or non-log scale, and whole reads, which were displayed as arrows with height representing the read number using a log scale.

For mRNA sequencing analysis, the parameters used in Bowtie were -n 3 -e 150 -a --best --strata -m 400, and the number of mutations allowed was calculated using the formula (length of read-14)/√(length of read). The parameter '-n 3' allows a maximum of 3 mismatches in the seed region; '-e 150' allows up to 5 mismatches; '-a --best --strata' returns the best matches within the seed region, using 26 as default; '-m 400' is the filter to remove reads that match more than 400 loci (repetitive loci). For mRNA analysis, we used the total sense mRNA reads for normalization. Otherwise, the analysis was very similar to that of sRNAs. Lists of sRNA reads or mRNA fragments matching to each of the four Argonaute pathways were generated by custom PERL scripts. These lists were then parsed using Microsoft Excel to analyze the presented data. Graphs were drawn using GraphPad Prism software.

### Definition of Argonaute-Specific Small RNA Target Loci

Genes targeted by endogenous siRNAs were grouped according to the Argonaute proteins they engage. Such Argonaute-based categories of sRNAs were experimentally defined as follows: (1) WAGO-class sRNAs, targeting 2,911 loci, were based on three mutant hermaphrodite sRNA libraries: *rde-3*, *mut-7* and MAGO12 (a combination of deletion mutations of all 12 WAGO Argonautes), all of which compromise 22G-RNA accumulation. Only targets that were depleted of sRNAs (≤5 normalized reads) in at least two of the mutant libraries were considered (Gu *et al.*, 2009); (2) CSR-1-class sRNAs, targeting 3,659 loci, were based on sRNAs obtained from two independent CSR-1 IP/sRNA cloning libraries and considered sRNAs that overlapped between the libraries with at least a two-fold enrichment in the CSR-1 IP relative to the input libraries (Claycomb *et al.*, 2009; Gu *et al.*, 2009); (3) ERGO-1-class sRNAs (49 loci) were based on 26G-RNAs enriched in an ERGO-1 IP/sRNA cloning library from wild-type embryos, where ERGO-1 26G-RNAs are most abundantly expressed, and corresponding to a two-fold or higher depletion of 22G-RNAs from the target genes in a sRNA library obtained from *ergo-1* deletion mutant embryos (Vasale *et al.*, 2010); (4) The ALG-3/4-class set of 1,684 loci was based on the overlap of two independent sRNA libraries from a *fog-2; alg-3; alg-4* mutant background (male enriched populations), with a requirement for a two-fold or higher depletion of 22G-RNAs in the mutant relative to wild type (Conine *et al.*, 2010). These gene sets are a conservative, yet high-confidence list of sequences targeted by these

pathways. Sequences targeted by CSR-1 and WAGO pathways practically do not overlap (only 33 genes), while many ALG-3/4 targets are also targeted by WAGO (126) or by CSR-1 (420). Ten ERGO-1 targets overlap with WAGO targets.

### Relative Distribution of Small RNAs Along mRNAs (Bin Analysis)

To analyze the relative distribution of 22G-RNAs and 26G-RNAs along all protein coding genes (using WormBase release WS215), each gene was divided into 20 equal bins, with each bin representing 5% of the coding sequence. For genes with multiple splicing forms, all coding exons were collapsed into one spliced genomic locus, before division into 20 bins. 22G-RNAs and 26G-RNA reads for each gene were normalized to the total nonstructural reads, and were mapped to each bin. For sRNAs spanning two bins, 50% of reads for each sRNA were assigned to the 5' bin and 50% were assigned to the 3' bin. The total number of reads in each bin was calculated by adding all reads for all genes in that bin. For this analysis, all annotated protein-coding genes were considered, regardless of whether they contained annotated 3' UTRs or not (annotated 3' UTRs were included in the bins).

### Nuclear Small RNA Isolation and Sequencing

Nuclei were isolated from predominately L4 avr3x animals grown on ivermectin and from seven-day counter-selected *pir-1* homozygotes, as described in Chapter IV Materials and Methods. The yield was lower for *pir-1* animals, possibly because of the relatively smaller germlines. We also noticed that following purification, *pir-1* nuclei tended to clump more after purification than wild-type nuclei. This indicated a tendency to lyse, which could be intrinsic to the *pir-1* phenotype (*i.e.*, nuclei from PIR-1-deficient cells may be structurally weaker). After the second purification over a sucrose solution, nuclear pellets were lysed by resuspension in TRI Reagent (MRC) and processed for RNA extraction as previously described. The integrity of the RNA was verified by running 0.5 µg of denatured RNA in a denaturing gel (15% polyacrylamide/7 M urea/0.5X TBE buffer), followed by staining with ethidium bromide and visualization with UV light. Intact control worms were processed directly with TRI Reagent (as described in Chapter II Materials and Methods). A fraction of the whole animal and purified nuclei samples were set aside for immediate protein extraction in order to test the purity of the nuclear extracts (as described in Chapter III Materials and Methods).

Cloning of sRNAs and read analysis was performed as described above. sRNAs were normalized using total nonstructural reads.

## Cloning, Expression and Purification of Recombinant PIR-1

Wild-type or mutant PIR-1 cDNA sequences lacking the first ATG (of the 233 aa protein) were inserted between the *Nde*I site and *Bam*HI sites of the pET-28a expression vector (Novagen) in fusion with the N-terminal 6x Histidine tag. The resulting constructs were transformed into BL21 (DE3) RIL *E. coli* cells, which were grown in 1 liter of LB medium at 37˚C to an $OD_{600}$ of 0.4, and induced for 4 hours with 1-2 mM IPTG at room temperature. Cells were pelleted at 5,000x *g* for 10 minutes at 4˚C and lysed by sonication in 25 ml of lysis/binding buffer (50 mM Tris-HCl pH 7.5, 700 mM NaCl, 5 mM β-mercaptoethanol, 5% glycerol, 15 mM imidazole, 0.01% NP-40). S100 fractions were prepared by ultracentrifugation at 100,000x *g* at 4˚C for 1 hour. In a 15 ml conical tube, 2 ml of HisPur beads (Thermo Scientific) were washed 3 times with binding buffer and centrifuged at 3,000x *g* between washes. The beads were mixed with the S100 supernatant and transferred to a 50 ml conical tube for rotation at 4˚C for 1 hour. Beads were transferred to an empty Poly-Prep chromatography column (Bio-Rad) and washed with at least 200 bead volumes of the lysis/binding buffer at 4˚C. Elution was performed at 4˚C with 500 µl of imidazole buffer per fraction (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol, 5% glycerol, 400 mM imidazole, 0.01% NP-40). Peak fractions were analyzed by SDS-PAGE (10% gel) followed by Coomassie Blue staining. Proteins were dialyzed using a solution containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 50% glycerol and 0.01% Triton X-100. Final protein concentration was estimated by A280 absorbance using a NanoDrop instrument.

## PIR-1 Dephosphorylation Assay

The reactions from Figure 4.13A were performed in a 10 µl volume containing 1X Terminator exonuclease buffer A, 1 U/µl Superasin and 1 µM of *in vitro* transcribed triphosphorylated single-stranded RNA substrate, ~0.25 µM of recombinant PIR-1 (wild-type or mutant), and 0.025 U/µl of Terminator exonuclease (Epicentre). The reactions, including the no-enzyme control, were incubated for 1 hour at 30˚C. They were then mixed with formamide gel loading buffer II (Ambion), and run on a denaturing 15% polyacrylamide/7 M urea/0.5X TBE buffer gel. The RNA was visualized with UV light after staining with SYBR Gold dye (Invitrogen) as described previously.

The reactions from Fig. 4.13B were performed differently, with a lower amount of substrate, and the Terminator digestion performed after treatment with recombinant PIR-1 or with CIP (NEB). Each 10 µl reaction contained 1X Terminator exonuclease buffer A, 1 U/µl Superasin and 0.5 µM of *in vitro* transcribed triphosphorylated single-stranded RNA substrate, ~0.25 µM of recombinant PIR-1 (wild-type or mutant), or 0.5 U/µl of CIP. The no-enzyme control and PIR-1 reactions were incubated at room temperature for 1 hour, and the CIP reaction was incubated at 37˚C for 1 hour. Subsequently, 2 µl of each reaction was set aside and placed on ice (these were loaded in lanes 1-4 and served as controls for degradation by potential contaminant nucleases in the PIR-1 or CIP enzyme mixtures). To the remaining 8 µl of each reaction, 0.25 µl (or 0.25 U) of Terminator exonuclease was added, and reactions were incubated at 30˚C for 1 hour. All samples were analyzed on a polyacrylamide gel and visualized as described above.

The triphosphorylated RNA substrate consisted of the 26-nt sequence rGrGrArUrCrCrTrTrGrArArArUrGrGrArArCrArUrCrUrGrArArU and was transcribed with T7 polymerase using the MAXIscript T7 Kit (Ambion) according to the manufacturer. The dsDNA template resulted from the annealing of two DNA oligos carrying the T7 promoter (O125 and O126). The RNA product was purified by PAGE in a denaturing 15% polyacrylamide/7M urea/0.5X TBE gel. The band was detected by UV shadowing, excised from the gel, and eluted as detailed in the sRNA cloning procedure. RNA was precipitated, resuspended in ultrapure water, and the concentration was estimated using a NanoDrop instrument.

# CHAPTER V

PIR-1 and Antiviral Immunity

## INTRODUCTION

Following the discovery of RNAi, one of the earliest proposed biological roles for the pathway was to provide defense against viruses. Until recently, the lack of a *C. elegans*-specific virus limited the use of this model organism for the investigation of antiviral immunity. Early studies were performed using a heterologous, transgene-based replication model of the ssRNA Flock House virus (FHV; Lu *et al.*, 2005; Lu *et al.*, 2009), widely used to study antiviral pathways in *Drosophila*, or to infection of *C. elegans* isolated primary cells by the mammalian Vesicular Stomatitis virus (Schott *et al.*, 2005; Wilkins *et al.*, 2005), among a few other systems (reviewed in Diogo and Bratanich, 2014). Nonetheless, this important initial body of work revealed that *C. elegans* relies on the RNAi pathway to restrict viral replication. Components found to be crucial for the antiviral response included the primary Argonaute RDE-1, the dsRNA-binding protein RDE-4, the RdRP RRF-1, DCR-1 and the DEAD-box protein DRH-1.

The discovery of Orsay virus in a wild *C. elegans* strain opened the door for studies of RNAi-based antiviral immunity without the confounding effects of heterologous viral infection models (Felix *et al.*, 2011). Orsay is distantly related to the *Nodaviridae* family, to which FHV belongs. Similarly to nodaviruses, it has a bipartite RNA genome composed of two segments of sense or (+)-strand polarity, containing merely three ORFs (Fig. 5.1A). Genome segment RNA1 encodes an RdRP (ORF A) for viral genome replication. RNA2 encodes the capsid protein (ORF α) and a protein of unknown function (ORF δ). This novel delta product distinguishes Orsay from nodaviruses, and has been recently shown to fuse with the capsid protein by means of a frameshift, and to incorporate mature viral particles (Jiang *et al.*, 2014; Guo *et al.*, 2014). The Orsay 982-aa RdRP shares ~27% amino-acid identity with the RdRPs of known nodaviruses. Curiously, in contrast to nodaviruses, the Orsay genome does not encode any homologs of RNAi inhibitors (*e.g.*, the dsRNA-binding B2 protein of FHV; Li *et al.*, 2002).

Orsay only propagates horizontally, as embryos isolated from infected mothers are never themselves infected (Felix, *et al*., 2011). Both the wild-isolate JU1580 strain, in which the virus was found, and an *rde-1* loss-of-function mutant exhibit severe intestinal morphology defects, including degeneration of cell nuclei, cell fusion and formation of multimembrane structures (Felix, *et al*., 2011). In spite of these defects, animals are still able to move and feed, have a wild-type lifespan and normal brood sizes, although embryos are produced at a slower rate. N2 wild-type animals are not immune to infection

but exhibit ~100-fold lower viral RNA levels than JU1508 or mutants of the RNAi pathway, and do not manifest visible phenotypes. Through northern analysis, genomic, sense (+)RNA and template (-)RNA strands were detected, although the latter were significantly less abundant (Felix *et al.*, 2011). Orsay virus replication and accumulation was found to be limited to the cytoplasm of one to six contiguous intestinal cells in the anterior part of the animal, out of the 20 large cells composing the adult intestine (Franz *et al.*, 2014). Infected cells were identified by the presence of the RdRP, the capsid protein, and of viral RNA, and could be detected in JU1580 and *rde-1* animals, but not in wild-type N2.

Genome sequencing of the JU1580 wild-isolate strain, revealed a deletion in the *drh-1* gene (Ashe *et al.*, 2013). DRH-1 (*D*icer-*R*elated *H*elicase 1) was previously shown to be required for antiviral RNAi (Lu *et al.*, 2009), despite the fact that it is entirely dispensable for exogenously triggered RNAi (Lu *et al.*, 2009) or for endogenous sRNA pathways (Gu *et al.*, 2009). By virtue of its conserved helicase and RIG-I-like C-terminal domains, DRH-1 is considered an ortholog of the cytoplasmic receptor RIG-I of mammals. Through its ability to recognize and bind 5'-triphosporylated viral dsRNA, RIG-I triggers a signaling cascade that leads to the induction of the interferon (IFN) response. As a consequence, an extensive immunity expression program is activated in infected and surrounding cells (Jiang *et al.*, 2011; Luo *et al.*, 2011; reviewed in Chan and Gack, 2015). Despite the absence of an IFN response or knowledge of analogous pathways in *C. elegans*, DRH-1 is considered to be a functional equivalent of RIG-I. This is based on a study using the *C. elegans* FHV-based replicon model in which it was demonstrated that a chimaera of the N-terminal domain of DRH-1 fused to the helicase and C-terminal domain of RIG-1 (replacing those of DRH-1) was able to fully rescue the inability of *drh-1* mutants to execute antiviral RNAi (Guo *et al.*, 2013).

High-throughput sequencing of sRNAs from infected *drh-1* mutant animals revealed that, unlike infected N2 animals, they did not produce a ~23-nt population of sense and antisense viral RNAs. These sRNAs possessed all the signs of Dicer cleavage: they did not exhibit a 5'-nucleotide identity bias, had 5'-monophosphorylated ends, were present at a similar proportion of antisense and sense strands, and exhibited a bias for 2-nt 3' overhangs when paired *in silico* (Ashe *et al.*, 2013). This was concordant with similar observations regarding the antiviral RNAi response against FHV in flies, in which it was shown that such RNAs were the products of Dicer-2 cleavage from a dsRNA replication intermediate rather than from regions of secondary structure within the viral genome (Flynt

*et al.*, 2009). Sequencing of viral siRNAs from additional mutants made it clear that these 23-mer species constituted the primary siRNAs which presumably would be loaded into the primary Argonaute RDE-1, to trigger the secondary amplification of effector 22G-RNAs (Ashe *et al.*, 2013). In agreement with this, abundant antisense 5'-triphosphorylated 22G-RNAs were detected in infected N2 animals, but were eliminated is mutants defective for 22G-RNA synthesis, such as *rrf-1* or *drh-3*. Furthermore, these sRNAs were proven to be *bona fide* effector 22G-RNAs, because they were absent from the MAGO12 mutant, in which none of the 12 WAGO Argonautes are expressed. In large part the silencing of viral RNA was attributed to the cytoplasmic SAGO-2(WAGO-6) Argonaute, shown previously to play a predominant role in exo-RNAi (Yigit *et al.*, 2006). Importantly, Ashe and colleagues also demonstrated that Orsay genome segments bearing 5'-PPP ends do exist in infected animals, by employing a 5'-RACE method in which the ligation of the adaptor depended on the conversion of a 5'-PPP to a 5'-P by a bacterial polyphosphatase. This did not address, however, the question of whether capped genome segments are also produced. Overall, the aforementioned work placed DRH-1 at the top of the viral RNAi pathway in *C. elegans*, leading to the hypothesis that this protein may be responsible for the initial recognition of replicating viral RNA molecules. Lastly, because DRH-1 was found to co-purify with RDE-1, RDE-4, and DCR-1 after immunoprecipitation of RDE-4 or DCR-1 (Tabara *et al.*, 2002; Duchaine *et al.*, 2006), it was also proposed that recognition of viral RNA is accompanied by Dicer-mediated cleavage. A schematic overview of the pathway summarizing these findings is depicted in Figure 5.1B.

Since PIR-1 recognizes RNA 5' triphosphates and stably interacts with Dicer, we hypothesized that PIR-1 could function within or in parallel to the antiviral RNAi response. We began by asking whether *pir-1* arrested animals could suppress Orsay virus replication in comparison to *rde-1* mutant animals. Surprisingly, we found that the *pir-1* mutant was unable to limit viral replication, leading us to conduct further experiments aimed at understanding whether PIR-1 acts in the same pathway initiated by DRH-1. The results we gathered suggest that PIR-1 acts between DRH-1 and RDE-1, although other scenarios are currently not ruled out.

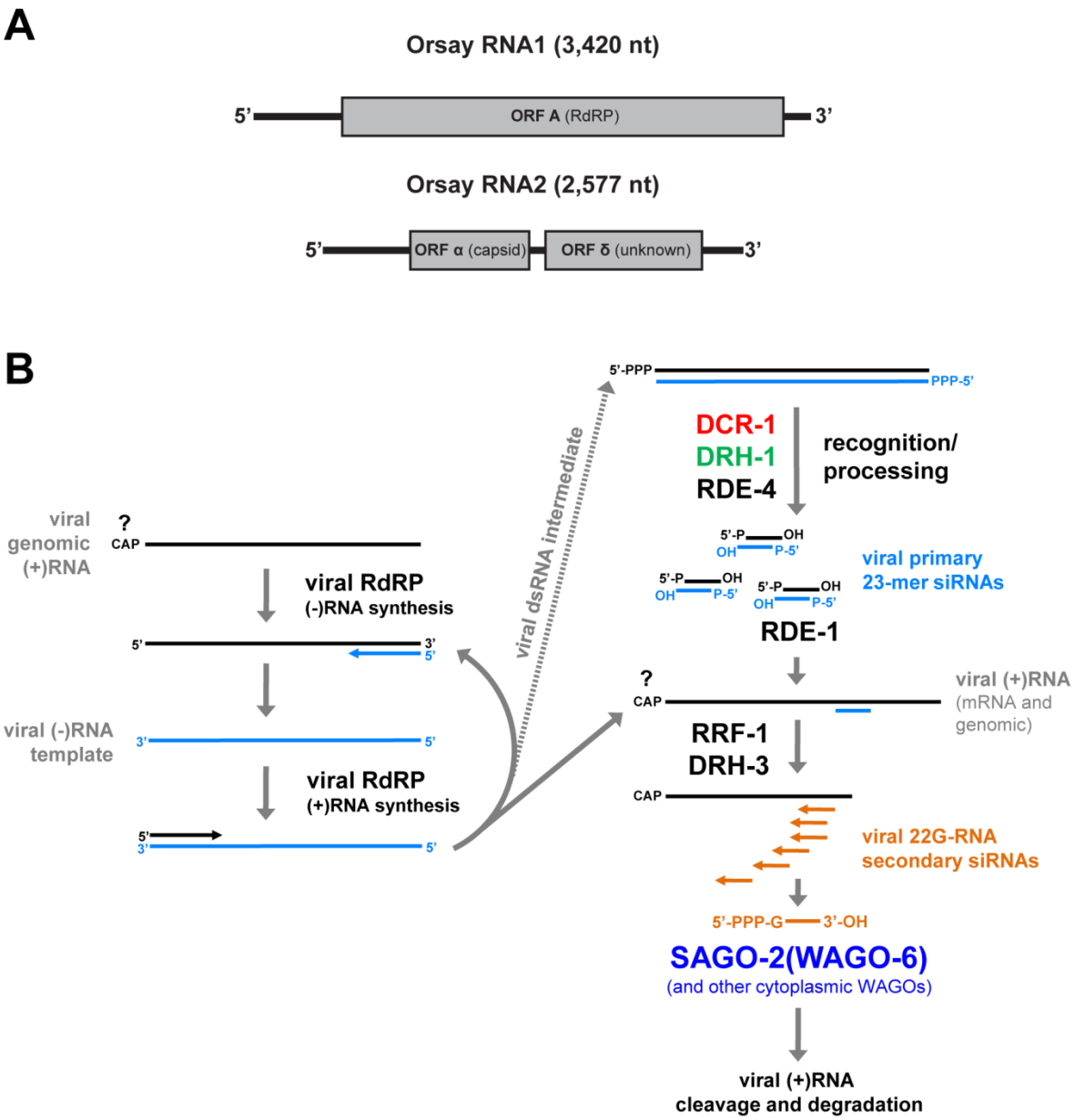



**Figure 5.1. Genomic organization and current model for small RNA-based suppression of Orsay virus. (A)** Schematic of the genomic arrangement of Orsay virus. The sizes of the whole genome segments were defined using reads from our sRNA deep-sequencing libraries of infected *C. elegans*. The gray boxes map the predicted open reading frames. **(B)** Diagram highlighting the major steps in suppressing Orsay virus infection by an RNAi-like mechanism, according to current knowledge. Upon release of the viral RNA into the cell cytoplasm the expression of the viral RdRP allows the replication of the viral genome to begin. Despite no available evidence for Orsay, based on current knowledge of nodaviruses, both genomic and viral mRNA (+) strands are presumed to carry a 5' cap, although the step at which this may occur is unknown. In the course of RNA replication, a low amount of dsRNA intermediate molecules (denoted by the dashed arrow) is presumed to accumulate. Similarly to FHV, and based on the finding that 5'-PPP Orsay viral RNAs exist *in vivo*, the 5'-ends of the dsRNA intermediates are thought to be triphosphorylated, thus providing a substrate for DRH-1 recognition and DCR-1 cleavage. The resulting 23-mer duplexes are incorporated into the primary Argonaute RDE-1, which, upon passenger-strand removal, recognizes viral ssRNAs and triggers the synthesis of abundant secondary anti-viral 22G-RNAs. In association with secondary WAGOs, viral 22G-RNAs lead to the degradation of viral RNA.

## RESULTS

### *pir-1* Mutant Animals Fail to Suppress Orsay Virus Replication

In order to test the susceptibility of *pir-1* mutant animals to infection by Orsay virus, synchronous populations of animals were infected by feeding OP50 *E. coli* mixed with viral filtrates derived from infected JU1580 animal populations. The level of viral replication was estimated from qRT-PCR measurements of Orsay RNA1 from total RNA normalized to the actin mRNA level of the host, and using the RNAi-deficient *rde-1* mutant as a positive control of infection. We found that regardless of exposure to Orsay virus during growth on counter-selective medium or after arresting terminally for a few days, the increase in viral load in *pir-1* animals relative to wild-type always closely followed that of the positive controls (Fig. 5.2A-C). The viral load in *pir-1* was identical to that of the *rde-1* mutant, and did not change in a *pir-1; rde-1* double mutant, suggesting that the two factors may act in the same pathway. We also tested a loss-of-function mutant for the RdRP RRF-1, which compromises the RNAi pathway downstream of RDE-1 by preventing effector 22G-RNAs from being made (Fig. 5.2B and C). Again, we detected as much viral RNA in *pir-1* as in the *rrf-1* mutant, with the *pir-1; rrf-1* double mutant not exhibiting an additive effect on viral replication. A similar viral load was detected in the *dcr-1(ok137)* mutant under the same counter-selection conditions as *pir-1* (Fig. 5.2C). As with other phenotypes, the wild-type *pir-1::gfp* transgene restored viral suppression to the wild-type level, while the catalytically compromised *pir-1::gfp* transgene (C150S) was unable to completely suppress viral replication. Although in the latter case the increase in viral load is not as pronounced as in *pir-1* alone, we conclude that the phosphatase activity is necessary, at least partially, for PIR-1 to exert its antiviral role (Fig. 5.2C).

PIR-1 is required for Dicer-dependent 26G-RNA synthesis. We therefore tested whether Eri 26G-RNA pathway mutants disrupted viral silencing. These included *eri-1*, *rrf-3*, *ergo-1*, and the *dcr-1(mg375)* helicase mutant, which is defective in 26G-RNA biogenesis but is still able to generate miRNAs and primary siRNAs in response to exogenous dsRNA triggers (Pavelec *et al.*, 2009; Gent *et al.*, 2010; Welker *et al.*, 2010). None of these Eri mutants exhibited a significant increase in Orsay viral load compared to wild-type, using *rde-1* as a positive infection control (Fig. 5.2D). This result indicates that unlike its role in 26G-RNA biogenesis, PIR-1 acts in the antiviral pathway likely through association with DCR-1 and probably RDE-4. The dsRNA-binding Dicer co-factor RDE-4 is required both in the canonical exo-RNAi pathway and in 26G-RNA endogenous

pathways (Tabara *et al.*, 1999; Tabara *et al.*, 2002; Duchaine *et al.*, 2006; Lee *et al.*, 2006; Vasale *et al.*, 2010; Welker *et al.*, 2010). We tested the susceptibility of an *rde-4* mutant to infection, in which we observed a higher increase in viral load than in the *rde-1* mutant. This implied that RDE-4 may function in concert with DCR-1 and RDE-1 to promote an adequate initiation of the antiviral sRNA response. The higher viral load measured in the *rde-4* mutant, however, may be indicative of an additional function for RDE-4 in the antiviral process, perhaps related to its reported ability to cooperatively bind long, but not short, dsRNA molecules both *in vitro* (Parker *et al.*, 2006; Parker *et al.*, 2008) and *in vivo* (Tabara *et al.*, 2002). One could imagine that this property could interfere with the replicative cycle of the virus, by not allowing intermediate dsRNA molecules to disassociate into productive ssRNA molecules.

Taken together, these findings indicate that PIR-1 is required to suppress the replication of Orsay virus in somatic cells. Its activity likely occurs in association with DCR-1 and in concert with all other enzymes thus far implicated in the antiviral response. Importantly, since we have shown that this function does not rely on 26G-RNA biogenesis, we have uncovered an additional independent biological role for this RNA phosphatase.

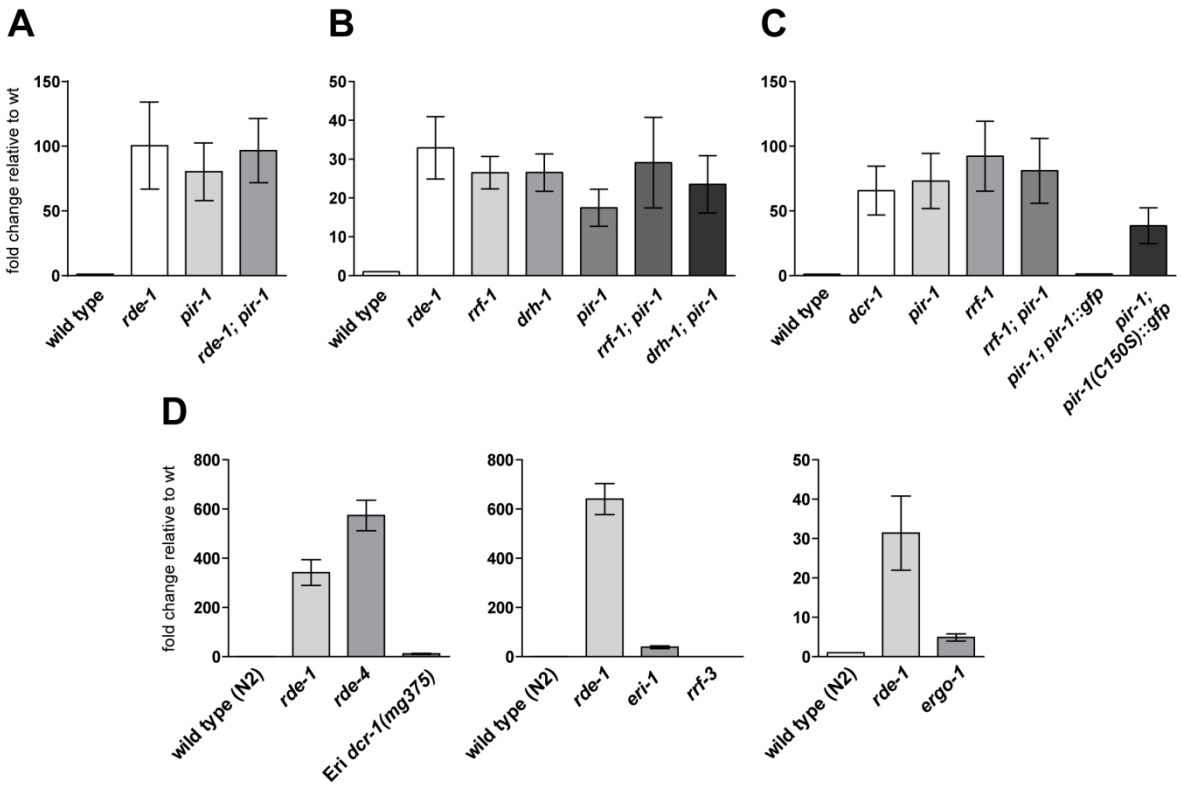



**Figure 5.2. PIR-1 is required to suppress Orsay virus replication.** Orsay RNA1 was measured in different mutants by qRT-PCR. Replicates correspond to three independent infections of the

same strain. Error bars represent the SEM. Levels of Orsay RNA1 were normalized to host actin mRNA and represent the fold-change relative to wild-type normalized RNA1 levels (equal to 1). Each panel represents a different experiment where all strains were of the same age and infected with the same stock of virus for identical periods of time. Strains in A, B, and C are all in the avr3x genetic background, and the wild-type strain is avr3x. **(A)** Animals were counter-selected on ivermectin for seven days before being exposed to virus for two additional days (see Chapter II Materials and Methods, page 102 for a description of selection by ivermectin). Wild-type worms were aged on ivermectin to seven days and purged of all progeny before being infected. **(B)** Animals were aged or counter-selected on ivermectin for four days before being exposed to virus for three days. **(C)** Animals were grown on ivermectin since the L1 larval stage to adulthood/arrested stage for four days with simultaneous exposure to virus. **(D)** Animals were grown since the L1 larval stage to adulthood for 2.5 days with simultaneous exposure to virus.

## PIR-1 Likely Functions Downstream of DRH-1

In order to understand how PIR-1 contributes to suppression of Orsay replication at the molecular level, we cloned and deep-sequenced sRNAs from infected samples of *pir-1* and key RNAi mutants. In order to study the Orsay 23-mer and 22G-RNA populations as extensively as possible, we employed the 5'-dependent/direct cloning method to follow monophosphorylated 23-mers, and the 5'-independent/TAP cloning method to measure triphosphorylated 22G-RNAs. We report most Orsay reads normalized to miRNAs, as these were cloned efficiently by both methods. Consistent with results from previous studies, the majority of sequences cloned from the wild-type N2 background are 22G-RNAs (Fig. 5.3A, normalized-read histograms and percentage of guanosine as the first nucleotide in the antisense sRNA population). Through direct cloning, only a very small amount of antisense 22-mers were cloned (note differences in scale), and even these were primarily 5'G, having likely resulted from 5'-PPP 22G-RNAs that were dephosphorylated intracellularly or during the cloning process of this particular library. Also, as previously reported, anti-viral 22G-RNAs targeted the entire length of the viral genome (Fig. 5.3B). Reflecting the low amount of viral replication, the cloning of duplex 23-mers in wild-type N2 by the direct method was reduced but still detectable (Fig. 5.3A, direct cloning). Conversely, in the *rde-1* mutant background, the 22G-RNA peak was lost in favor of a drastic accumulation of 23-mers exhibiting no nucleotide bias (Fig. 5.3A, histograms and almost equal distribution of the four nucleotides of the antisense sRNA population). The inexistence of a nucleotide bias is consistent with DCR-1 not exhibiting a nucleotide preference, and matches the nucleotide proportions found in the entire Orsay genome (RNA1 + RNA2, rightmost panel of Fig. 5.3A). As expected for products of Dicer cleavage from viral dsRNA precursors (*i.e.*, not loaded into Argonaute proteins), 23-mers

mapped across the entire viral genome and, compatible with being free-standing sRNA duplexes, they were cloned in roughly equal proportions from the sense and antisense strands (Fig. 5.3 A and B).

When we conducted the same analysis in infected *pir-1* animals we observed a strong 23-mer duplex peak (Fig. 5.4B compared to A). These 23-mers still distributed along the entire length of the Orsay genome, with similar levels of sense and antisense reads (Fig. 5.4F). Unlike *rde-1*, however, *pir-1* animals still produced abundant antisense 22G-RNAs. Since the animals used had been infected over the course of four days since the L1 larval stage, we concluded that the 22G-RNAs accumulated while a substantial amount of maternal PIR-1 was still present in the developing animals. Confirming that this was the case, when we profiled the sRNAs of *pir-1* animals that were infected after having been counter-selected for seven days, the 22G-RNA peak decreased substantially (Fig. 5.4C). In both situations, however, the accumulation of large amounts of 23-mers was dramatic and reminiscent of the response observed in the *rde-1* mutant. This pattern was reverted to the wild-type pattern in the rescued *pir-1* background, in which the 23-mer levels were much lower but the 22G-RNA peak remained high (Fig. 5.4D; compare scales between wild-type, *pir-1* and rescued *pir-1*).

In *dcr-1* loss-of-function animals, and consistent with the 23-mer duplexes being the direct product of Dicer cleavage, the 23-mer peak was lost, but again, a strong 22G-RNA peak was still produced (Fig. 5.4E). Unlike *pir-1*, antisense 22G-RNAs accumulated whether *dcr-1* mutant animals were counter-selected as they were being infected, or whether they were infected after a prolonged period of counter-selection. In these libraries, sense read abundance varied substantially (blue bars) due to a high degree of viral (+)RNA degradation during RNA extraction and/or cloning. The heightened levels of sense relative to antisense reads, accompanied by their distribution over a wide range of sizes and an absence of clear peaks, reliably imply nonspecific RNA degradation.

Analysis of *drh-1* mutant animals revealed a loss of the 23-mer peak, consistent with the results of Ashe *et al.* (2013), which placed it as the component that initiates the viral silencing process by recognizing the viral 5'-PPP (Fig. 5.5A). However, we have trouble reconciling this conclusion with the fact that we could still detect a strong 22G-RNA peak of about 15,000 rpm through the TAP cloning method (Fig. 5.5A, lower graph, red bars). Based on the overall size profile of all the sense reads cloned by both the direct and TAP methods (blue bars, Fig. 5.5A), we assumed that the antisense 23-mer reads were mostly degradation products. Even so, we cannot exclude the possibility that a residual

amount of *bona fide* 23-mers are made by Dicer to trigger the production of antisense 22G-RNAs downstream. Given that the viral load in the *drh-1* mutant is high, the 22G-RNAs were probably generated late in the viral replicative cycle, and consequently not able to silence the vast amount of viral RNA that accumulated in the meantime. We next compared the viral sRNA profiles of *pir-1* and *drh-1; pir-1* animals of the same age that had been counter-selected for four days before being infected (Fig. 5.5B and C). We observed that in the absence of DRH-1, the viral sRNA profile of *pir-1* arrested animals resembled that of the *drh-1* single mutant: the prominent 23-mer peak was lost, while abundant antisense 22G-RNAs were still produced. This result strongly suggests that DRH-1 acts upstream of PIR-1 in the antiviral pathway. Finally, supporting the idea that DRH-1 and PIR-1 function in separate complexes, we never detected a physical association between the two proteins by MudPIT, or through IP of PIR-1 and western blotting of DRH-1 with and without Orsay infection (not shown). We cannot, however, rule out that the two proteins interact very transiently or below our experimental detection thresholds, particularly during the most acute stages of viral infection.

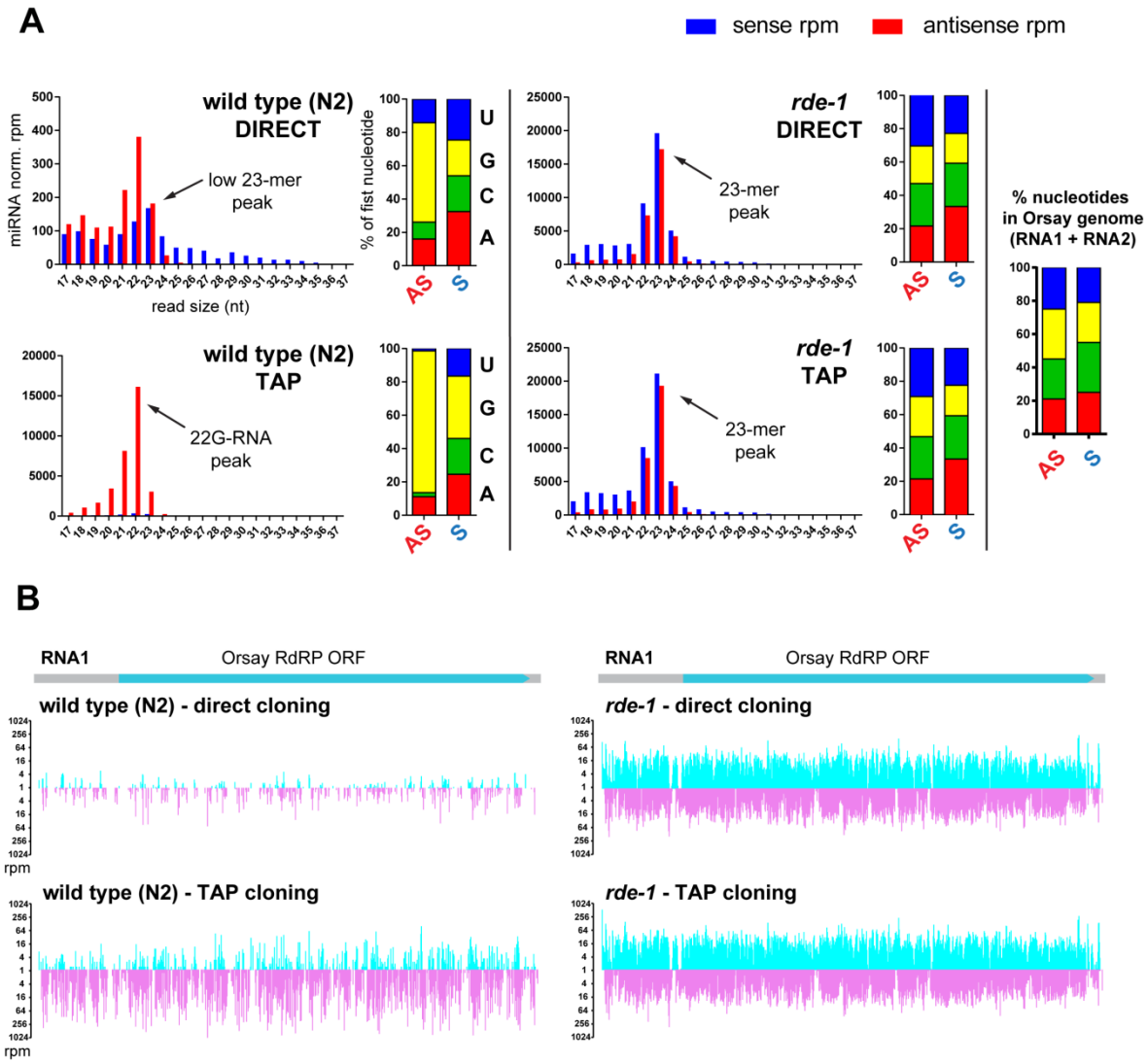



**Figure 5.3. *rde-1* mutant animals infected with Orsay virus accumulate viral 23-mer siRNAs but not downstream 22G-RNAs.** Small RNA deep-sequencing library profiles and distribution of reads across the viral genome in wild-type N2 and *rde-1* mutant animals. **(A)** Comparison of miRNA-normalized reads from direct (5'-dependent) and TAP (5'-independent) cloning libraries with regard to size, polarity and first (5') nucleotide identity for all Orsay reads (RNA1 and RNA2). The *y* axis indicates rpm and the *x* axis the read size in nucleotides. The first nucleotide distribution is expressed as a percentage of all reads combined. The data relative to the N2 and *rde-1* samples were obtained using the same populations used for RNA1 qRT-PCR in Figure 5.2D (left panel, page 226). The rightmost panel shows the proportion of each nucleotide in the entire Orsay genome for comparison with sRNA read first nucleotide fractions. 'AS', antisense. 'S', sense. **(B)** Genome browser representation of miRNA-normalized reads mapping to RNA1 only. Each peak corresponds to the first nucleotide of each read. Blue reads are sense (same polarity as the viral genome and mRNA) and pink reads are antisense. The vertical axis represents rpm in a log2 scale.

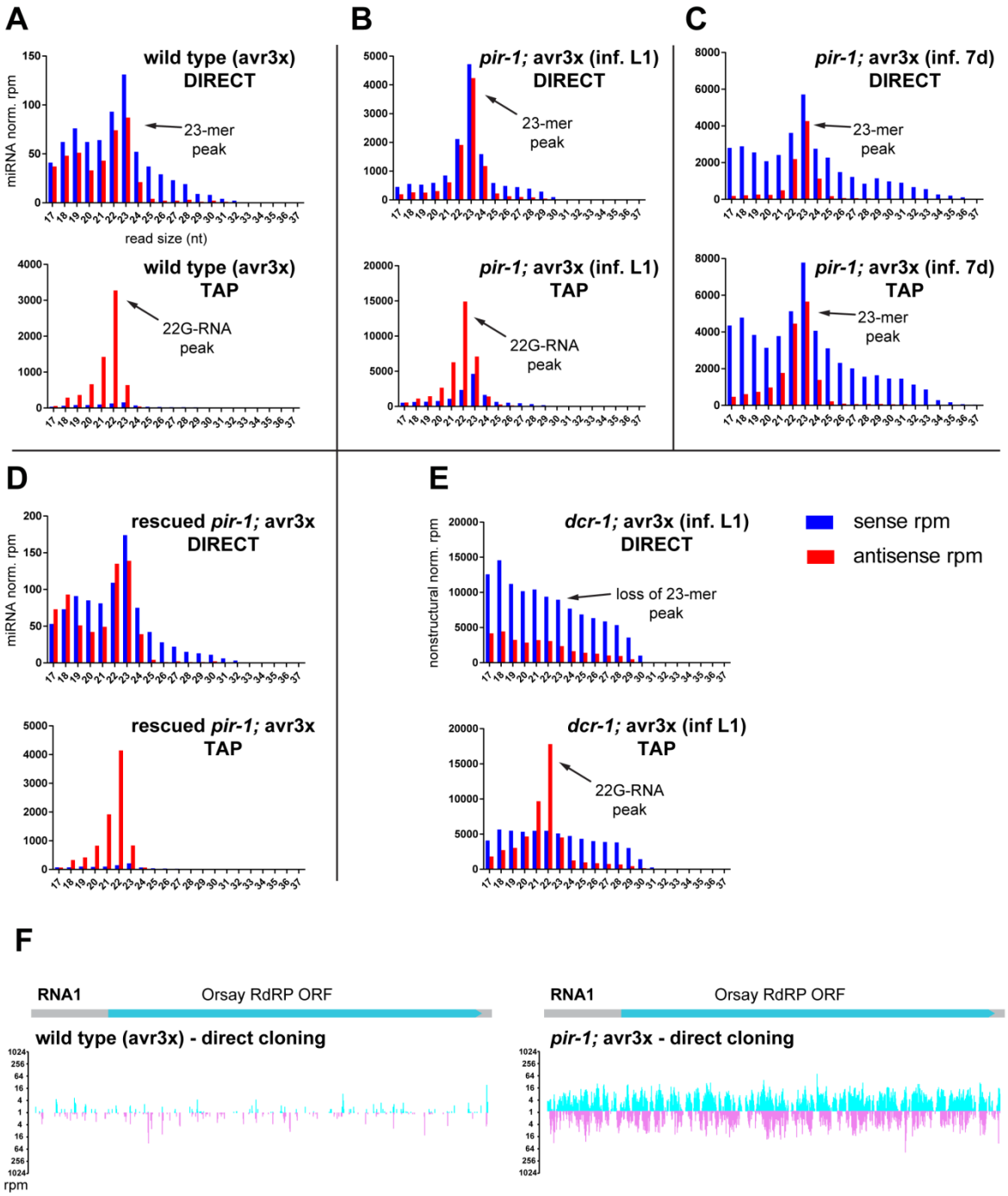



**Figure 5.4. *pir-1* mutant animals infected with Orsay virus accumulate viral 23-mers but still produce 22G-RNAs due to maternally loaded PIR-1.** Small RNA deep-sequencing library size profiles in wild type, *pir-1*, rescued *pir-1* and *dcr-1* mutants, and distribution of reads across the viral genome in *pir-1* mutants. In panels A-E, sRNA reads map to Orsay RNA1 and RNA2. All samples are normalized to miRNAs except for *dcr-1*, in which reads were normalized to total nonstructural reads. A, B, D, and E are samples corresponding to those represented in Fig. 5.2D, page 226. 'inf. L1' refers to infection starting at the L1 larval stage, while 'inf. 7d' means that animals were only infected after seven days of ivermectin counter-selection. The presence of abundant sense reads of all sizes derived from the predominant viral (+) RNA indicates that extensive degradation occurred in those samples (note panels C and E, blue bars). **(A)** In wild type, very few viral 23-mers are produced relative to the abundant effector 22G-RNAs that keep the viral

load at a low level. **(B)** *pir-1* mutant animals accumulate abundant 23-mers, but produce 22G-RNAs even more abundantly, despite having a much higher viral load than wild-type animals. **(C)** This *pir-1* mutant sample, in which animals were infected when the level of PIR-1 protein is presumably very low or null, is meant to show a decrease in the 22G-RNA peak measured by TAP cloning compared to the profile shown in B. This suggests that viral 22G-RNA accumulation depends on PIR-1. **(D)** *pir-1; pir-1::gfp;* avr3x rescued animals exhibit the same viral sRNA profile as avr3x, in agreement with their ability to suppress Orsay virus replication. **(E)** In *dcr-1* deletion mutant animals the 23-mer peak is lost, in contrast to a very abundant 22G-RNA population, in spite of a high viral load. The 22G-RNA accumulation in both *pir-1* (panel B) and *dcr-1* likely reflects early expression of maternally inherited PIR-1 and DCR-1. **(F)** Genome browser representation of miRNA-normalized reads mapping to RNA1 in wild type and *pir-1* mutant animals by direct cloning, to reflect 23-mer reads. Each peak corresponds to the first nucleotide of each read. Blue reads are sense and pink reads are antisense. The vertical axis represents rpm in a log2 scale.

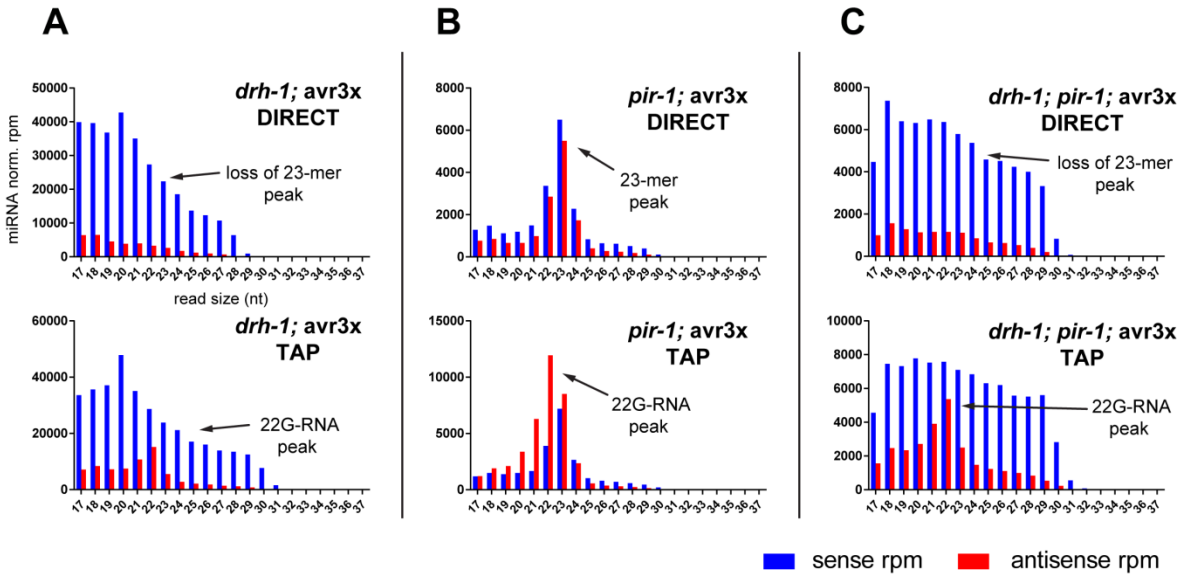



**Figure 5.5. The *drh-1* mutation suppresses the accumulation of 23-mers in *pir-1* mutant animals.** Small RNA reads map to Orsay RNA1 and RNA2. A, B, and C correspond to samples from Fig. 5.2B, page 226. In *drh-1* mutant animals, 23-mer peaks are lost. As in Figure 5.4, the tall blue bars across all sizes indicate extensive degradation of the viral (+)RNA strand. **(A)** *drh-1* mutant animals show a loss of the 23-mer peak and a relatively small 22G-RNA peak (compared to other antisense read sizes). **(B)** *pir-1* mutant animals again show prominent 23-mer and 22G-RNA peaks. **(C)** *drh-1; pir-1* double mutant animals exibit a sRNA profile similar to the *drh-1* single mutant, suggesting that DRH-1 is required for the accumulation of 23-mers upstream of PIR-1.

## PIR-1 Functions Upstream of Secondary 22G-RNA Production along with Factors Required for Primary 23-mer Synthesis and Function

We next analyzed the relationship between viral load, 23-mers and antisense 22G-RNAs in greater detail for some of the libraries (Fig. 5.6). For simplification, and because viral load was measured through qRT-PCR of Orsay RNA1 (Fig. 5.6A), we only present

sequences derived from this genome segment. It should be noted that the qRT-PCR analysis did not distinguish between strands. However, because studies of FHV replication showed that only 1-5% of viral RNA is of the negative polarity (Ball, 1994), we assume that our measurements reflect mostly the accumulation of genomic and mRNA of the positive/sense polarity. We also have preliminary deep-sequencing data suggesting that this is the case in Orsay virus replication (not shown).

Regarding 23-mers, in libraries where degradation was not significant, the ratio of sense to antisense 23-mer reads tended to be close to one (*e.g.*, Fig. 5.4B, page 231). In contrast, in libraries where more degradation products were cloned (*e.g.*, Fig. 5.4E, page 231), the sense 23-mer peak (blue) was higher than the antisense 23-mer peak (red; Fig. 5.6A). Considering this, the actual number of 23-mer duplexes should be judged by the height of the red, antisense bars.

In order to increase the stringency of our analysis, we restricted ourselves to 22G-RNAs with a flanking YR motif, which is defined as a purine (R= A or G) being the first nt (+1) in the sRNA, and a pyrimidine (Y= C or T) being the adjacent 5′ nt. This motif and extended consensus is similar to the initiator element that Pol II uses to initiate transcription in plants, mammals and flies (Smale and Baltimore, 1989; de Hoon and Hayashizaki, 2008; Juven-Gershon *et al.*, 2008). The YR motif was found in our laboratory to be associated with capped sRNA precursors of Pol II-dependent Type II 21U-RNAs (Gu *et al.*, 2012), but also with RdRP-dependent 22G-RNAs (Weifeng Gu, unpublished results). Since the cloning of degradation products varied substantially from library to library, confining our analysis of the TAP-cloned libraries to YG 22-mer reads, eliminated most sense and antisense 22G-RNA reads that derived from the degradation of viral RNA (*i.e.*, not *bona fide* RdRP-dependent 22G-RNAs). Furthermore, we subtracted the normalized antisense 22G-RNAs reads cloned by the direct method from the normalized 22G-RNA reads cloned by the TAP method (Fig. 5.6B). Since Dicer cutting is not precise, not all primary viral siRNAs are exactly 23-nt long (some are 22-nt long, and will start with a 5' guanosine, based on chance alone). Therefore, this subtraction accounted for those Dicer products which were also cloned through the 5'-independent TAP method, but could not be distinguished from real RdRP-generated antisense 22G-RNAs.

When we considered the mutant *rde-1*, the dramatic accumulation of 23-mers (Fig. 5.6A, first and second panels, independent samples) was accompanied by a sharp reduction in 22G-RNAs (Fig. 5.6B). This explains why the viral load is so high in these animals (~400-fold increase relative to wild-type viral RNA1 level, Fig. 5.6A) and implies that

RDE-1 is the main Argonaute using primary 23-mers to trigger 22G-RNA synthesis. *rde-4* mutant animals, on the other hand, despite a very high viral RNA load and a high 23-mer peak, were still able to generate an appreciable amount of 22G-RNAs compared to wild type. As we have discussed before, the fact that so many antisense 22G-RNAs are generated is probably a consequence of the high availability of viral template molecules. However, since 23-mer generation is probably inefficient and thus delayed relative to wild-type, and because the availability of downstream RNAi machinery (such as secondary effector Argonautes) is limited, the replication of the virus outpaces the rate at which the 22G-RNA-loaded Argonautes can destroy the viral (+)RNAs. Therefore, if we factor in the viral load into the number of antisense 22G-RNAs (by dividing it by the qRT-PCR fold-change), the number of 22G-RNAs per viral (+)RNA in the *rde-4* mutant is extremely low compared to wild type.

Perhaps a more accurate way of estimating the ability of each mutant to produce antiviral 22G-RNAs, is to consider the ratio of antisense 22G-RNAs to antisense 23-mers (Fig. 5.6D). Since RDE-1-loaded 23-mers that are antisense to the viral (+)RNA strand trigger the synthesis of antisense 22G-RNAs, this ratio is a measure of the efficiency with which 23-mers can lead to 22G-RNA accumulation. When we calculated this ratio for N2, it was of about 30, meaning that for every antisense 23-mer, approximately 30 secondary 22G-RNAs were produced. In contrast, the ratio was nearly zero for *rde-1*, meaning that the abundant 23-mers are unable to generate 22G-RNAs. In *rde-4*, this ratio also approaches zero, despite the fact that a substantial amount of *bona fide* 22G-RNAs accumulate. In contrast to RDE-1, which is absolutely required for 22G-RNA generation, RDE-4 likely just adds efficiency to the anti-viral silencing response. Without it the response is still elicited but it is not sufficient to keep up with viral replication.

Extending this analysis to other mutants, only *rrf-1* behaved like *rde-1*: it accumulated a very large amount of 23-mers, but failed to generate 22G-RNAs, in agreement with RRF-1 being the main RdRP in the process. *drh-1* mutant animals had a reduced amount of 23-mer duplexes (if we consider the height of the accumulating duplexes to match that of the antisense (red) 23-mers), but still generated an almost wild-type amount of 22G-RNAs. *pir-1* mutant animals (two independent replicates, central panels) accumulated 23-mers, albeit to a lower extent than in the *rde-1* or *rde-4* backgrounds, and still generated a large amount of 22G-RNAs relative to wild-type. The *drh-1; pir-1* double mutant exhibited a 23-mer decrease similar to the *drh-1* single mutant, in contrast to the accumulation observed in the *pir-1* background, confirming that DRH-1

likely acts upstream of PIR-1. It should be noted that in this particular panel of mutants (all grown under the same conditions), the viral loads were relatively low compared to other experiments, reflected in the higher 22G-RNA/23-mer ratios relative to the wild-type ratio. In *dcr-1* mutant animals, the level of 23-mers was similar to those of wild-type animals, regardless of whether the normalization was performed with miRNAs or total nonstructural sRNAs (asterisks). But again, as we have mentioned earlier, most of the antisense 23-mers cloned in the *dcr-1* and *drh-1* backgrounds, based on the size profiles of reads (Figures 5.5 and 5.6) are likely the result of nonspecific degradation and not of Dicer cleavage. Despite the fact that 22G-RNAs in *dcr-1* still accumulated to an appreciable level, the 22G-RNA/23-mer ratio still revealed a defect in the production of 22G-RNAs from 23-mers. These 23-mers were likely produced by residual maternal DCR-1 protein. Finally, in rescued *pir-1* animals the 23-mer/22G-RNA ratio was restored to a wild-type value (Fig. 5.6, rightmost panels), showing that, similarly to all other mutants analyzed, *pir-1* is required for a timely and robust generation of effector antiviral 22G-RNAs.

Lastly, when we took into account the high viral load in *pir-1* mutant animals, the amount of 23-mer reads was consistently lower than in wild-type (Fig. 5.6C). This argument also held true for *rde-1*, *rde-4*, *dcr-1*, and *drh-1*, all of which act at the 23-mer production/function levels. In contrast, the *rrf-1* mutant accumulated 23-mers to a level almost ten-fold higher than in *rde-1* without exhibiting a proportional increase in viral load. As a result, when we normalize the 23-mer amount in *rrf-1* animals to the respective viral load, the levels approximate the 23-mer levels in wild-type. As expected from the absence of RdRP activity, this indicates that 23-mer RNA synthesis occurs normally and that the only block occurs at the 22G-RNA level. Collectively, these results reveal that PIR-1 plays an important role in limiting Orsay virus replication. We hypothesize that Dicer-associated PIR-1 promotes early and robust accumulation of 23-mers, downstream of DRH-1, to ensure an effective 22G-RNA response. The fact that 23-mers are still detected in *pir-1* animals can be tentatively rationalized by a lowered, but still present, activity of Dicer on viral dsRNA intermediates.

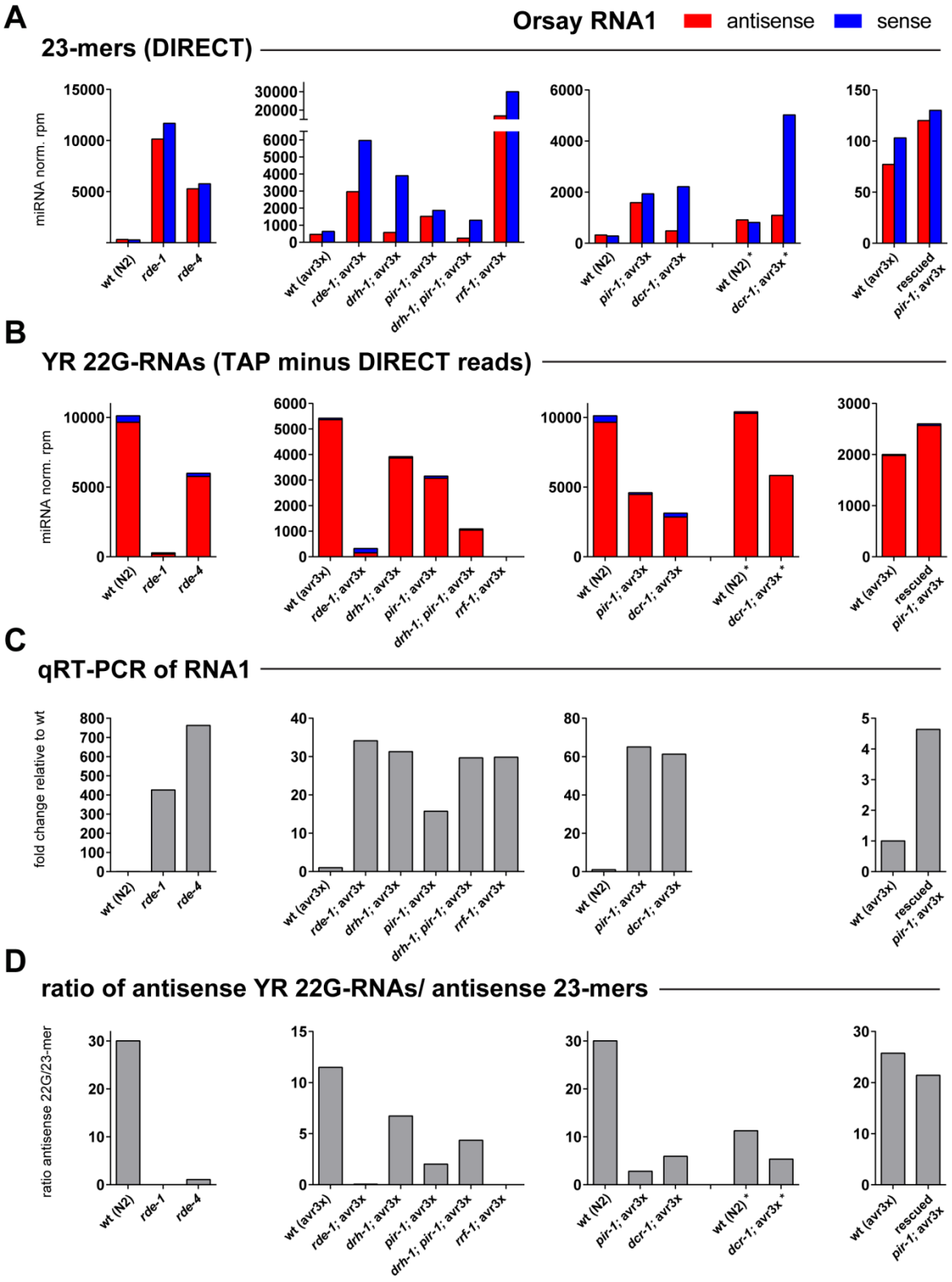



**Figure 5.6. PIR-1 is necessary for the robust accumulation of antiviral 22G-RNAs.** Comparison between Orsay RNA1 23-mer reads, antisense 22G-RNA reads and viral RNA levels in different mutants. **(A)** miRNA-normalized RNA1 23-mer reads obtained by direct cloning. Higher sense read columns are indicative of degradation in the samples. **(B)** miRNA-normalized YR-motif 22G-RNAs (see main text for explanation) obtained by TAP cloning, from which YR 22G-RNA reads obtained by direct cloning were subtracted. This subtraction accounted for 5'-P 22-nt products of Dicer cleavage that were indistinguishably cloned together with bona fide 5'-PPP

22G-RNAs through the TAP-based method. **(C)** qRT-PCR measurements of RNA1 from Figure 5.2 for individual samples subjected to sRNA deep-sequencing. **(D)** Ratio of antisense YR 22G-RNA reads to antisense 23-mer reads, as a measure of the ability of 23-mers to trigger secondary 22G-RNA synthesis. A lower ratio indicates a decreased ability to produce 22G-RNAs from existing 23-mers. Asterisks in the N2 and *dcr-1* samples indicate normalization with total nonstructural reads.

## The Interaction of Viral Primary 23-mers with RDE-1 Is Partially Compromised in *pir-1* Mutant Animals

Having found that *pir-1* mutant animals accumulate viral 23-mers to high levels without, however, generating enough 22G-RNAs to control viral replication, we asked whether such 23-mers were properly loaded into the RDE-1 Argonaute. To answer this question we generated transgenic strains carrying multi-copy extrachromosomal arrays of a *gfp::rde-1* construct under the control of native regulatory sequences. One strain was generated in an *rde-1* null mutant background and another in a genetically balanced *pir-1; rde-1* double mutant background. The lines we obtained were tested for rescue of RDE-1 activity through recovery of the muscle twitching phenotype elicited upon RNAi of *unc-22* (not shown). We assessed transgene expression by performing GFP IP/western blotting from rescued lines exhibiting the highest transgene transmission rates (Fig. 5.7A). Additionally, we demonstrated that the *gfp::rde-1* transgene could also rescue the deficiency in Orsay virus suppression of the *rde-1* mutant (Fig. 5.7B). We then counter-selected *rde-1*-rescued *pir-1* mutant animals (referred to as '*pir-1*') for seven days followed by a three-day infection with Orsay virus, and grew the rescued *rde-1* line on virus for three days until they reached the L4-YA stage (referred to as 'wt'). We then TAP-cloned and deep-sequenced sRNAs from both input (extracted from intact animals) and GFP::RDE-1 immunoprecipitates, and analyzed Orsay-derived sRNAs.

As a control for specificity of binding to viral 23-mers, we also included the miRNA Argonaute ALG-1, which, similarly to RDE-1, also predominately binds 5'-P 22- and 23-mer RNAs and is ubiquitously expressed (Vasquez-Rifo *et al.*, 2012). Since loss-of-function mutants for ALG-1 or its redundant sister Argonaute ALG-2 were still able to suppress Orsay virus infection as well as wild-type animals, we expected that few viral 23-mers would be loaded into these Argonautes (Fig. 5.7C). We therefore analyzed sRNAs associated with ALG-1 not only in the wild-type background, but also in the *rde-1* mutant background in order to increase the availability of viral 23-mers in infected cells.

Analysis of the sRNA composition of the different libraries (from absolute, non-normalized reads) allowed us to make important observations (Fig. 5.7D). First, as a percentage of all reads mapping to expressed genomic loci, reads corresponding to Orsay sequences never exceeded 15% of the total. Second, in wt *rde-1* rescued animals, Orsay sRNAs were enriched by 40% compared to input, whereas in the *pir-1* mutant background they were enriched by just ~20%. Third, ALG-1 IPs specifically enriched the libraries with miRNAs by 75-80% relative to inputs, irrespective of genetic background. Additionally, the interaction of ALG-1 with Orsay sRNAs was almost null and did not increase with higher viral loads, despite the fact that the fraction of viral sRNAs in inputs increased from 0.7% in N2 animals to 5% in *rde-1* animals. These results show that indeed RDE-1 specifically and abundantly interacts with Orsay sRNAs, and, further, that this interaction may be perturbed by the absence of PIR-1.

We next analyzed these numbers in greater detail, by decomposing the reads into 23-mers and YR-motif 22G-RNAs (Fig. 5.8). In our initial analyses, we chose miRNAs as the normalization standard because they are mostly expressed in the soma, where Orsay infection also takes place. We found, however, that using total miRNAs was not ideal for our comparisons primarily for two reasons: (1) RDE-1 engages a large proportion of sRNAs mapping to protein coding RNAs (Fig. 5.7D, gray lines), but a small proportion of miRNAs compared to input; and (2) the proportion of miRNAs associating with RDE-1 in the *pir-1* mutant was almost 20% higher than in the wild-type rescued animals (Fig. 5.7D, orange lines). Instead, we resorted to normalization using only miR-243, which has been shown to specifically interact with RDE-1 but not ALG-1, to direct the synthesis of 22G-RNAs from the somatic *Y47H10A.5* transcript (Correa *et al.*, 2010). We confirmed this result in our libraries, where miR-243 was enriched in RDE-1 IP libraries relative to inputs, but not in ALG-1 IP libraries (Fig. 5.8B). Focusing on miR-243-normalized Orsay RNA1-matching reads, we observed a predominance of 22G-RNAs in the wild-type *rde-1* rescued input library (Fig. 5.8A; compare to the overall nucleotide composition of the RNA1 genome segment on the far right panels). After RDE-1 IP, a clear peak at 23 nt with roughly equal read numbers of the sense and antisense polarities was obtained, with the proportion of first nucleotides losing their G bias (due to depletion of antisense 22G-RNA reads). This shows that RDE-1 preferentially interacts with viral 23-mers without a strand bias. In the *pir-1* input library, the 22G-RNA peak was still detected, although the large numbers of sense reads of all sizes denoted cloning of degraded viral RNA. After RDE-1 IP, a 23-mer duplex peak was recovered but, curiously, both polarities exhibited a bias

toward sRNAs beginning with a uridine with a concomitant reduction in the proportion of sequences starting with A and G compared to wild-type 23-mers (in wild-type, U-23-mers occurred at frequency of ~20%, whereas in *pir-1* they constituted 40-45% of total Orsay 23-mers). Mapping of these U-23-mers revealed that they were still distributed along the length of the entire Orsay RNA1 sequence, as were A-, G-, and C-23-mers (not shown).

When we compared the RDE-1 IPs (Fig. 5.8C), 23-mers were recovered in the *pir-1* mutant background at a ~4.5-fold lower abundance than in the wild-type (*rde-1* rescued) animals. As the cloning was performed through the 5'-independent TAP method, it must be considered that in the input antisense 23-mer reads there is a significant contribution of 23-nt RdRP sRNAs (*i.e.*, 22G-RNAs that are actually 23-nt long, represented in the yellow portion of each bar). In the wild-type input, these constitute ~75% of the antisense reads, while in the *pir-1* input these make up ~60% of all antisense 23-mer reads. If we take into account the roughly equitative proportions of the first nucleotide found in the Orsay genome, then an approximate estimate for the real number of Dicer-produced 23-mers of each polarity should be of ~9,000 reads in the wild-type input and ~5,000 reads in the *pir-1* input. This means that while in the wild-type RDE-1 IP sample there is an enrichment of about five-fold in 23-mers relative to the input, in *pir-1* the enrichment is only about two-fold. Not surprisingly, antiviral 22G-RNAs were heavily depleted in RDE-1 IP samples from both wt and *pir-1* animals (Fig. 5.8D). This result, together with the peculiar bias towards 5'U-23-mers, hints that in the *pir-1* background 23-mers are not loaded into RDE-1 as effectively as in a wild-type context.

Finally, in both wild-type N2 animals and *rde-1* mutant animals, 23-mers were depleted in ALG-1 IP samples compared to reads measured in inputs (Fig. 5.8C bottom panel). In the case of the N2 input, considering that 84% of reads begin with a guanosine, the depletion was of almost four-fold. In the *rde-1* sample, there was no 5'-nucleotide bias in the antisense reads, and the depletion of reads in the ALG-1 IP was of approximately three-fold. Predictably, 22G-RNAs were also not enriched in the ALG-1 IPs (Fig. 5.8D bottom panel). Since the miR-243 read numbers in the ALG-1 IP samples are very low (Fig. 5.8B, bottom panel), the numbers of viral 23-mers in these samples are overestimated, such that the real depletion of 23-mers is much higher than three- to four-fold (we have confirmed that this is the case when the normalization was performed with total miRNAs). These results demonstrate that the miRNA Argonaute ALG-1 is not used by cells to induce the antiviral RNAi response.

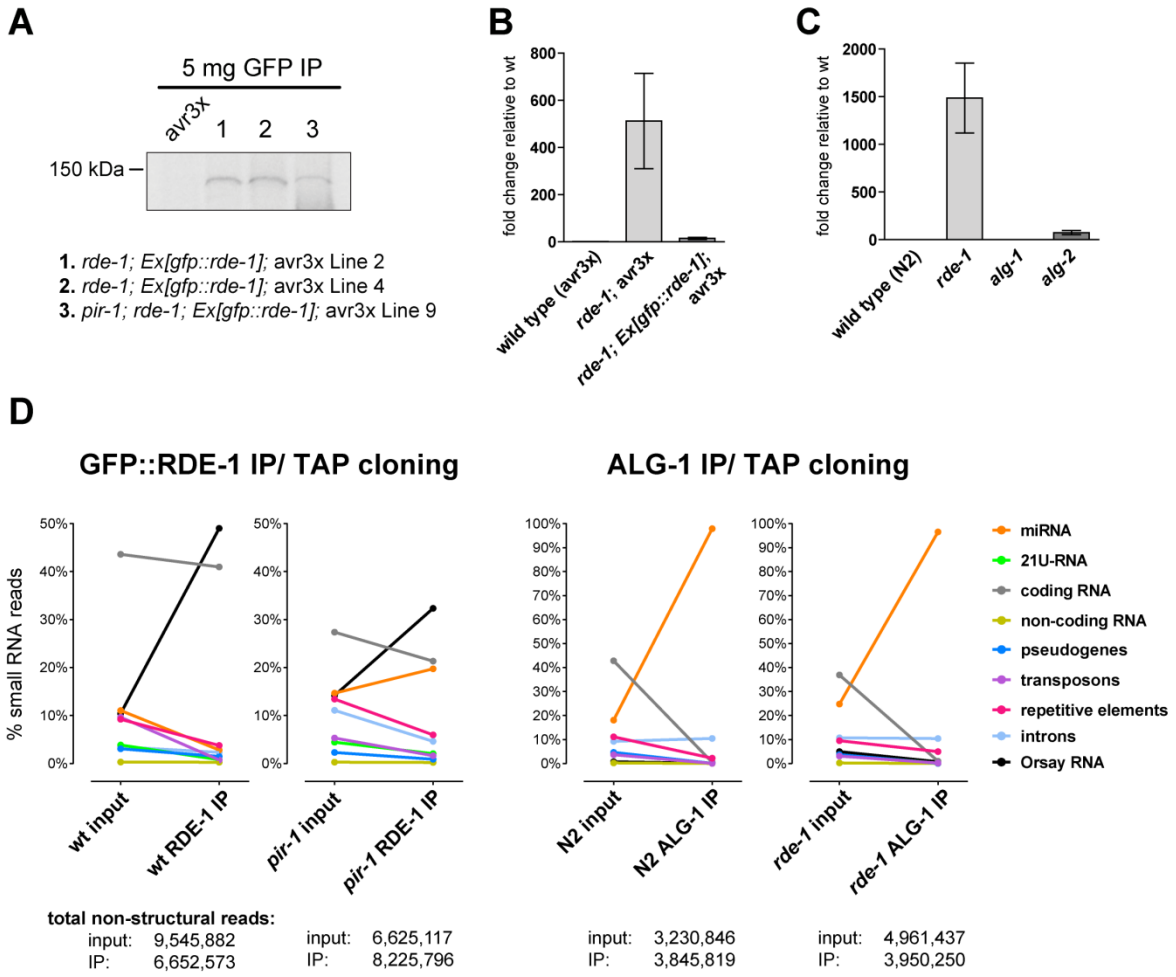



**Figure 5.7. GFP::RDE-1 transgenic lines and fraction of small RNA classes cloned from RDE-1 and ALG-1 immunoprecipitates. (A)** Western blot of immunoprecipitated GFP::RDE-1 expressed in non-integrated lines where the *rde-1* mutation was rescued (assessed by RNAi of *unc-22*). Lines from lanes 2 and 3 were used for IP/sRNA cloning. **(B)** qRT-PCR of Orsay RNA1 normalized to actin from infected *rde-1* and *rde-1* transgenic line 4 animals showing that viral suppression is restored by the *gfp::rde-1* transgene. **(C)** qRT-PCR of Orsay RNAi in infected *alg-1* and *alg-2* loss-of-function mutants. **(D)** Graphic comparison between the fraction of major sRNA groups cloned from total (inputs) or immunoprecipitated sRNA as a percentage (*y* axis) of total nonstructural reads (indicated below graphs). 'wt' refers to the *rde-1* rescued transgenic line and '*pir-1*' refers to the *pir-1; rde-1* double mutant, in which only the *rde-1* phenotype is rescued.

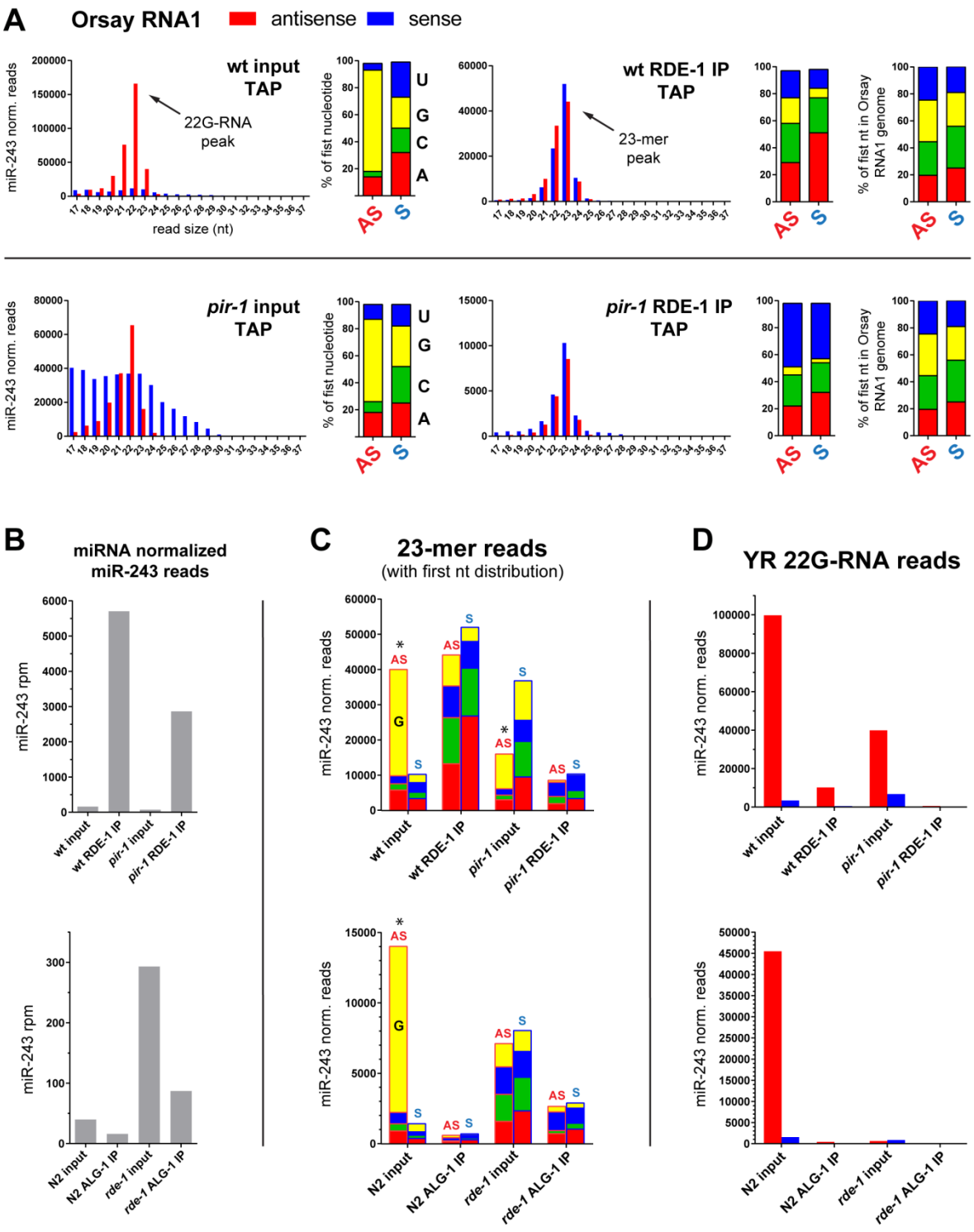



**Figure 5.8. The specific association of Orsay 23-mers with RDE-1 is weakened in the *pir-1* mutant background.** Cloning of sRNAs from total RNA or RDE-1 and ALG-1 immunoprecipitates was performed using the 5'-independent TAP method. Reads were normalized to miR-243, which interacts preferentially with RDE-1. For simplicity, only Orsay RNA1 sequences were considered. **(A)** Size distribution of reads and fraction of the identity of the first nucleotide according to read polarity in RDE-1 IP samples and respective inputs. The far right panels represent the proportion of each nucleotide present in the entire RNA1 genome segment for comparison with the sRNA read nucleotide proportions. **(B)** Total miRNA-normalized miR-243 read numbers in the eight different libraries. **(C)** 23-mer read numbers in each sample, separated

into sense and antisense reads and further into first nucleotide identity. The asterisks indicate that the antisense 23-mer reads have a large proportion of reads with a 5'G, indicating that they correspond to longer products of RdRP synthesis, or "23-nt long 22G-RNAs". Note that 23-mers binding RDE-1 in the *pir-1* mutant background exhibit a 5'U bias. **(D)** YR-motif 22G-RNA reads in each sample. 'AS', antisense reads; 'S', sense reads.

## Contributions

Daniel Chaves constructed and grew all *C. elegans* strains, performed viral infections, RNA extractions and qRT–PCRs. Some sRNA cloning libraries were prepared by Daniel Chaves, but most were prepared in our laboratory by Weifeng Gu. Mapping of reads and preliminary analyses of deep-sequencing data were performed by Weifeng Gu, and subsequent analyses were performed by Daniel Chaves and Weifeng Gu. The GFP::RDE-1 IP experiment was performed by Daniel Chaves, including sRNA library preparation, while the ALG-1 IP experiment was entirely performed by Weifeng Gu. Additional libraries were prepared by Ruidong Li, a student at the University of California Riverside in the laboratory of Weifeng Gu. Sequencing of all libraries was performed at UMass Medical School.

## DISCUSSION

The experiments presented in this chapter have demonstrated that *pir-1* mutant animals are unable to suppress replication of Orsay virus, a (+)-strand ssRNA virus related to nodaviruses (Felix *et al.*, 2011). By showing that the suppression of Orsay virus replication does not depend on 26G-RNAs, we have discovered an additional biological function for PIR-1 in *C. elegans*, whereby it promotes an effective antiviral 22G-RNA response. When the *pir-1* deletion was combined with loss-of-function mutations for factors previously implicated in sRNA-based immunity against Orsay including DRH-1, RDE-1, and RRF-1 (Felix *et al.*, 2011; Ashe *et al.*, 2013), the relative viral loads measured in the double mutants never exceeded the viral loads of the individual mutants. This suggested that PIR-1 acts in concert with these factors to mediate viral defense rather than being part of a parallel immune pathway. Analysis of viral sRNAs in these mutant backgrounds revealed that PIR-1 functions downstream of DRH-1, which has been postulated to sense 5'-triphosphorylated dsRNA viral molecules (Guo *et al.*, 2013; Ashe *et al.*, 2013). Moreover, we showed for the first time that RDE-1 specifically interacts with Dicer-produced viral 23-mers and that in the absence of PIR-1, loading of RDE-1 with these sRNAs is partially compromised.

The addition of PIR-1 to the repertoire of proteins required for the RNAi antiviral response is somewhat puzzling, considering that at least two Dicer protein complexes with enzymes capable of recognizing RNA 5' triphosphates act in the same pathway: PIR-1/DCR-1 and DRH-1/DCR-1. Since much remains to be understood about the dynamics of viral RNA recognition and the molecular details of how Dicer complexes engage and process viral dsRNA, characterizing PIR-1 activity in this context is likely to provide key insights. Given that PIR-1 is conserved in all metazoans, and that RNAi has been reported as an important antiviral barrier evolutionarily extending to vertebrates (Parameswaran *et al.*, 2010; Li *et al.*, 2013; Maillard *et al.*, 2013), further elucidation of this process will get us closer to understanding the interplay and balance that must occur between distinct, co-existing antiviral immunity pathways in metazoans. Below, we propose a model for Orsay small-RNA based suppression taking into consideration the RNA phosphatase activity of PIR-1, and discuss the potential for functional conservation of this protein in antiviral pathways of more complex animals.

## A Model for PIR-1 Function in Promoting Small RNA-Based Antiviral Immunity

Based on the results reported here and elsewhere, we propose an integrated working model for Dicer-associated PIR-1 activity in suppressing the replication of the nematode-infecting Orsay ssRNA virus (Fig. 5.9). Since our data suggest that DRH-1 acts upstream of PIR-1, the first step in the control of viral infection is likely to be the DRH-1-mediated recognition of triphosphorylated viral dsRNA molecules. 5'-PPP dsRNA is generated during the synthesis of (-)RNA template strands and during the amplification of (+)RNA by the viral RdRP. The latter provide genomic RNA to be incorporated into new viral particles and mRNA to be translated into structural and functional components of the virus. We speculate that DRH-1, perhaps by virtue of a very high affinity to its substrate, is able to recognize viral RNAs at very low cellular concentrations, during the early stage of the viral replicative cycle. As DRH-1 has long been found to form a stable complex with DCR-1, RDE-4 and RDE-1 (Tabara *et al.*, 2002), recognition of viral dsRNA may trigger Dicer cleavage and loading of 23-mers into the RDE-1 primary Argonaute. Since this step occurs very early in infection and dsRNA intermediates are probably unstable, loaded RDE-1 complexes do not accumulate to a sufficiently high level to trigger a robust 22G-RNA response. Rather, these early events could activate signaling pathways or act directly to recruit additional factors that lead to a potent response capable of restricting the increasing concentration of viral RNA. One attractive and unexplored possibility is that initial recognition by the DRH-1 complex may nucleate the binding of RDE-4 molecules along the regions of dsRNA. RDE-4 has been shown to promote the accumulation of dsRNA *in vivo* (Tabara *et al.*, 2002) and to bind dsRNA in a cooperative manner *in vitro* (Parker *et al.*, 2006; Parker *et al.*, 2008). With these properties in mind, one can imagine that multiple RDE-4 molecules could cover regions of duplexed viral RNA, triggered by the initial binding of DRH-1 complexes at dsRNA termini. This could not only result in the sequestration of dsRNA away from viral machinery, but perhaps also establish a molecular platform for the recruitment of additional processing factors, including the DCR-1/PIR-1 complex. Such a mechanism could be especially relevant in the battle against Orsay virus considering that its replication, similarly to that of its cousin Flock House Virus in *Drosophila* cells, could be restricted to specialized cellular compartments. Specifically, the RdRP of FHV has the capacity to insert itself into the outer membrane of mitochondria of the host cell, leading to invaginations inside of which replication complexes and dsRNA

intermediates are thought to be sequestered and "hidden" from host defense systems (Miller *et al.*, 2001; Miller and Ahlquist, 2002; Kopek *et al.*, 2007).

Given that the phosphatase activity of PIR-1 is required to suppress Orsay replication, we speculate that the dephosphorylation of the ends of dsRNA molecules may interfere with the ability of the virus to separate the two strands. In this scenario, PIR-1 would be in direct competition with the potential unwinding and capping activities of the viral RdRP (several viral RdRPs incorporate RNA capping activity; reviewed in Aartjan and te Velthuis, 2014). This would lead to the accumulation of stable, full-length viral dsRNA that would be cleaved by DCR-1 to provide a steady source of primary 23-mer viral siRNAs to be loaded into RDE-1. Alternatively, the stabilization of dsRNA molecules could occur not as a consequence of dephosphorylation, but rather due to the elimination of terminal sequences in the viral RNA required for viral RdRP initiation. In FHV, negative strand synthesis is dictated by elements within the 3'-proximal 108 and 50 nucleotides of (+)RNA1 and (+)RNA2, respectively, while positive strand synthesis relies on 3-14 nucleotides at the 3' ends of (-)RNA strands (Venter and Schneemann, 2008). Assuming that Orsay replication requires similar elements, PIR-1 could act as a sensor for terminal 5'-PPP followed by a Dicer cleavage event that would destroy such critical sites and halt initiation of further rounds of viral RdRP synthesis. Regardless of the dsRNA stabilization mechanism, this would guarantee an abundant cytoplasmic synthesis of antisense 22G-RNAs by the host RdRP RRF-1 using the existing viral (+)RNA strands as templates. Since the stabilized viral dsRNA substrates would encompass the entire viral genome sequence, primary 23-mers, and consequently secondary 22G-RNAs, would also cover the entire sequence, in agreement with our results.

Regarding RDE-1 loading, while the DRH-1 complex incorporates RDE-1 (Tabara *et al.*, 2002), the PIR-1 complex does not (Chapter III), such that it may require other factors to couple the cleavage of 23-mers to RDE-1 loading. It is plausible that upon the first cleavage by the PIR-1/DCR-1 complex at either end of the dsRNA intermediates, PIR-1 is no longer required for additional rounds of cleavage, since the remaining dsRNA would now have Dicer-generated 5'-P receded ends. In this case, a DCR-1/RDE-4/RDE-1 complex would be sufficient to cleave the rest of the dsRNA molecule, with the added advantage of ensuring coupling of Dicer cleavage to RDE-1 loading. In this scenario, it is not clear to us how to interpret our observation that in the *pir-1* mutant the interaction of RDE-1 with 23-mers is less pronounced (especially considering that 23-mers still accumulate and are distributed along the entire length of the viral genome). We cannot rule

out, however, that in the context of viral infection, RDE-1 could be recruited to PIR-1/DCR-1/RDE-4 complexes already localized to foci of viral replication, and facilitate cleavage and/or loading by an unknown mechanism.

Due to gaps in our knowledge regarding the molecular details of the Orsay virus replicative cycle and the biochemical properties (*e.g.*, RNA binding affinities, rate and processiveness of cleavage) of several of the *C. elegans* factors involved in this process, this is a highly speculative model. As such, it is far from being the only possible model, but it is one that tries to integrate the little we have learned from recent studies and our own work. As a working model it generates interesting and testable hypotheses that are worthy of future pursuit. Perhaps the most interesting hypothesis concerns stabilization of dsRNA precursors for Dicer cleavage. Specifically, the accumulation of dsRNAs could be tested in a *dcr-1* mutant, where they would potentially be preserved by the absence of cleavage (assuming that other RNA degradation pathways would not take over). The proof that PIR-1 has a stabilizing effect on viral dsRNA intermediates would come from the inability to detect these molecules in a *dcr-1; pir-1* double mutant. Experiments such as this one could not only help us define the precise molecular function of PIR-1, but also expand the repertoire of mechanisms that cells employ to combat viral infections at the RNA level.

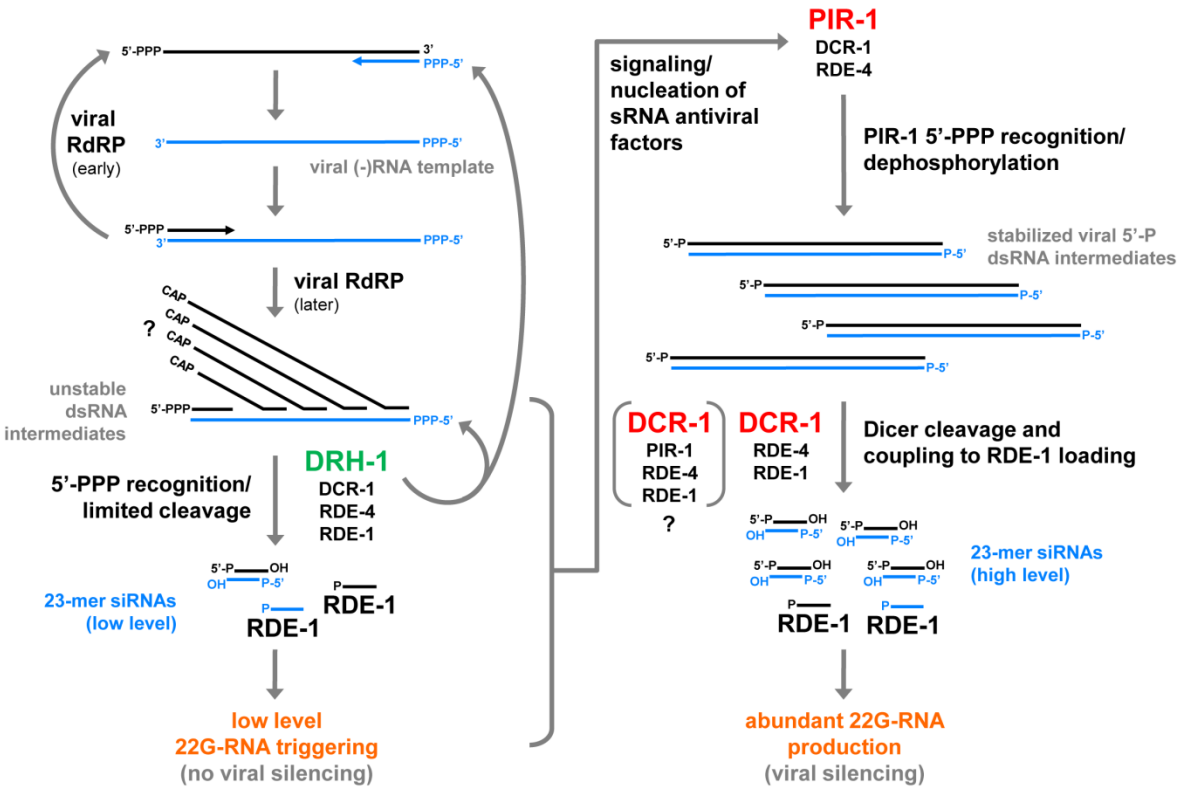



**Figure 5.9. Model for PIR-1 function in the control of Orsay virus replication.** After entry of viruses into the cell and translation of the first viral RdRP molecules from the genomic segments, the generation of negative polarity template strands is initiated. As enough template strands accumulate, the balance shifts toward a more abundant production of positive strands to be incorporated into new virions. Many of these molecules are likely immediately capped by the viral RdRP or the host capping enzyme. In competition with this activity, DRH-1, with very high affinity for RNA 5' triphosphates, is able to bind the ends of some of these temporary dsRNA intermediates. Although it may contain DCR-1, the DRH-1 complex is not able to generate enough 23-mers to produce sufficient effector 22G-RNAs, but may rather provide a signal for the recruitment of additional dsRNA processing factors. Among these could be a PIR-1/DCR-1 complex capable of dephosphorylating the ends of the labile dsRNA intermediates and perhaps permit their stabilization by keeping viral factors, or cellular factors recruited by the virus, from separating the strands. This would allow DCR-1 to initiate cleavage of the whole viral sequence providing a steady source of 23-mers to be loaded into RDE-1. In turn, RDE-1/sRNA complexes would bind viral (+)RNA molecules to trigger the abundant primer-independent synthesis of effector 22G-RNAs by RRF-1, thereby allowing the host cell to destroy viral RNAs at a fast rate. Since PIR-1-mediated dephosphorylation would only be required at the ends of the full-length viral dsRNAs, the majority of cleavage could be performed by DCR-1/RDE-1 complexes, facilitating the loading of primary viral 23-mers. Alternatively or in parallel, RDE-1 may transiently associate with a complex of DCR-1/PIR-1/RDE-4 (in parentheses) to execute the cleavage, coupled to RDE-1 loading.

## Why Do 22G-RNAs Accumulate in Mutants That Cannot Suppress Viral Replication?

The abundant production of 22G-RNAs in mutants where Orsay virus replicates uncontrollably seems paradoxical. When taking into account the high viral load present in these mutants, however, the relative amount of antiviral 22G-RNAs is lower than in wild-type animals, and clearly unable to keep up with viral replication. Nonetheless, the fact that they are produced in mutants where maternal effects are not at play is suggestive that the system is relatively permissive to the lack of accessory components, such as DRH-1 or RDE-4. In a context where these components are absent, Dicer may be able to elicit cleavage of dsRNA intermediates after viral RNAs reach a high enough cytoplasmic concentration, bypassing, for instance, the DRH-1-mediated 5'-PPP recognition step. In the case of the *drh-1* mutant, the virus takes advantage of a prolonged period without cellular surveillance during which replication is unobstructed. Viral RNA therefore accumulates to a level that can no longer be downregulated by 22G-RNAs due to a delayed accumulation of trigger 23-mers.

In *dcr-1* mutants, it is likely that by the time RNA is extracted, enough 22G-RNAs have been triggered by 23-mers made earlier with maternal DCR-1. 23-mer primary species may have a much more rapid turnover than 22G-RNA/Argonaute complexes, explaining the disappearance of the 23-mer peak. Again, the lack of 23-mers does not allow cells to keep up enough 22G-RNA synthesis to destroy viral RNAs. In the *rde-1* background, the accumulation of 23-mers is made possible by unperturbed production by Dicer. Assuming that RDE-1 is the only Argonaute that can engage viral 23-mers, this implies that their production rate amply exceeds the rate at which cells can degrade them. Additionally, 23-mer duplexes may be intrinsically more stable than other cellular RNAs, also contributing to their accumulation. In the case of *rde-1*, however, as with *rrf-1* mutant animals, 22G-RNAs are never synthesized, and the unrestricted synthesis of viral RNA is easily explained.

## In the Absence of PIR-1, RDE-1 Engages Fewer and 5'U-Biased Antiviral 23-mers

While it is undeniable that PIR-1 is required for suppression of Orsay virus replication, the reason for accumulation of 23-mers is unclear and confounded by the maternal load effect. In this regard, *pir-1* behaves like the *rde-1* mutant. As shown in Figure 5.4 (page 233), when maternally loaded PIR-1 is reduced by counter-selecting

animals for a longer period before infection, the 22G-RNA peak is decreased relative to the 23-mer peak. This suggests that in the theoretical complete absence of maternal load, only 23-mers would accumulate, and that the reduction in 22G-RNAs would be even more pronounced. Since proper loading of RDE-1 with 23-mers is absolutely necessary to generate 22G-RNAs, we believe that this outcome holds a connection to the observation that immunoprecipitated RDE-1 in the *pir-1* background is able to engage a lower and qualitatively different (5'U bias) fraction of 23-mers than in a wild-type context. We do not understand, however, how PIR-1 could increase the efficiency of loading of RNA into RDE-1. In one possibility, the PIR-1 complex may engage accessory factors that mediate this process during infection. Based on this reasoning, we tentatively place PIR-1 upstream of RDE-1 in the viral silencing pathway.

Some important caveats to the RDE-1 IP/cloning experiments should be mentioned. First, no perfect normalization standards applied to this experiment. miR-243 is adequate for normalization between RDE-1 IPs, but may skew results for inputs, due to low read numbers in those samples. Despite this, when we normalized the data with total miRNAs, although the numbers produced were different, the overall conclusions were maintained (*i.e.*, RDE-1 IP enrichment of 23-mers in wild type, much lower enrichment in *pir-1*, and depletion in ALG-1 IPs.). Second, in addition to TAP cloning, the preparation of direct cloning libraries would have helped to provide better estimates of Dicer-dependent 23-mers. Third, since we did not sequence more 23-mer reads in *pir-1* than in wild type, we suspect that the expression levels of the rescuing *gfp::rde-1* transgene in the *pir-1* background could have been lower than in the control strain, at least in the intestinal cells where Orsay virus replicates. While this does not invalidate our conclusion (23-mers still heavily interact with RDE-1 in the *pir-1* background), future studies should eliminate this source of variability by using integrated transgenic lines with uniform and ubiquitous expression of the transgene. Nonetheless, we believe that these results suggesting that PIR-1 acts upstream of RDE-1 loading are important enough to warrant further investigation.

Finally, we exclude the possibility that *pir-1* mutants are not able to produce enough 22G-RNAs due to a decreased level of RDE-1 because (1) arrested mutants can sustain a normal RDE-1-dependent exo-RNAi response (Chapter IV), (2) they express higher levels of *rde-1* mRNA than wild-type (mRNA-seq, Appendix D), (3) and *pir-1(tm1496)* animals express wild-type levels of RDE-1 protein (not shown).

## The Pattern of PIR-1 Expression Is Compatible with a Constitutive Role in Antiviral Immunity

The ubiquitous spatio-temporal expression of PIR-1 in *C. elegans* is in agreement with it being a key component of antiviral immunity. The Orsay replicative cycle, for instance, is restricted to the cytoplasm of a few somatic cells (Felix *et al.*, 2011; Franz *et al.*, 2014), but the fact that PIR-1 is expressed in practically every cell of the organism suggests that cells are permanently primed for defense against (yet-to-be-discovered) viruses that exhibit wider cell and tissue tropisms than Orsay. Additionally, although we think that the expression of PIR-1 in the nucleus may be related to its role in 26G-RNA biogenesis, and perhaps in broader functions encompassing gene expression regulation and cell division (discussed in Chapter II), it may also be required for suppression of viruses that replicate in cell nuclei. These include viruses belonging to the *Orthomyxoviridae* family (including Influenza viruses), which also contains a vertebrate/invertebrate-infecting genus named *Thogotovirus*. Orthomyxoviruses carry multipartite (-)RNA genomes and replicate in the nuclei of infected cells, with the likely formation of dsRNA intermediates. A role in the defense against such viruses would explain why the nuclear localization of PIR-1 is maintained throughout the entire life of *C. elegans*. The hypothesis that PIR-1 may restrict viral replication in the nucleus will be easily testable once such viruses are identified in *C. elegans*.

## Does PIR-1 Contribute to Antiviral Immunity in Other Animals?

Our findings concerning the involvement of PIR-1 in suppressing the replication of an RNA virus led us to question whether this highly conserved phosphatase is required for antiviral defense in other metazoans. In Chapter III we showed that in human cells PIR1 can interact with Dicer, at least when PIR1 is overexpressed. Those results were extended by a former member of the laboratory, Darryl Conte, who asked whether human orthologs of DRH-3 interacted with Dicer in human cells. He considered the three cytoplasmic RIG-I-like receptors (RLRs) that sense viral RNA: RIG-I, MDA5 (*M*elanoma *D*ifferentiation-*A*ssociated protein 5) and LGP2 (*L*aboratory of *G*enetics and *P*hysiology 2) (reviewed in Bruns and Horvath, 2014; Errett and Gale, 2015). All three proteins are highly homologous to *C. elegans* DRH-1, DRH-2 and DRH-3, namely by the shared presence of a DExD/H box and a RIG-I-like C-terminal domain. Notably, the worm RLR orthologs lack the N-terminal tandem caspase activation and recruitment (CARD) domains found in RIG-I and MDA5 (for an interspecies comparison of viral RNA sensing by duplex RNA-activated

ATPases, consult Paro *et al.*, 2015). Upon dsRNA binding through the helicase domain of these proteins, the CARD domains provide a platform onto which signaling cascade-initiating complexes can assemble. Such signaling leads to expression of IFN genes, which in turn heighten immune response pathways in the infected and neighboring cells. The experiment consisted of overexpressing tagged versions of the human RLRs in HEK293T cells, followed by immunoprecipitation and western analysis. The IP of LGP2, but not RIG-I or MDA5, robustly co-immunoprecipitated not only endogenous Dicer, but also endogenous PIR1 (Darryl Conte, unpublished results). The interaction of Dicer with LGP2 was later independently confirmed in a proteomics-based study of HEK293 cells with an activated IFN response (Li *et al.*, 2011). This reinforced the idea that the Dicer-PIR1 interaction is maintained in vertebrates, but also led to the hypothesis that human PIR1 is involved in antiviral immunity.

LGP2 has the highest RNA binding affinity among the RLRs, and is able to recognize dsRNAs regardless of length or 5'-phosphorylation state. It cannot, however, trigger the IFN response on its own due to the lack of CARD domains (Bamming and Horvath, 2009; Takahasi *et al.*, 2009). Despite a variety of reports showing that LGP2 can positively modulate RIG-I and MDA5 activity (Satoh *et al.*, 2009; Moresco and Beutler, 2010 and references therein; Bruns and Hovarth, 2015), the robust association of LGP2 with Dicer and PIR1 in the absence of viral infection, tempts us to speculate that this may represent a constitutive complex functioning outside of the canonical IFN response. In this context, LGP2 may act as a highly sensitive cytosolic dsRNA sensor that may rapidly place Dicer and PIR1 on invading dsRNA to promote the generation of viral siRNAs. This idea is especially plausible in light of a study demonstrating that an effective Dicer-dependent RNAi-based antiviral immunity response was mounted in mice infected with Nodamura virus, a bipartite (+)RNA virus related to FHV (Li *et al.*, 2013). In this study, the IFN response was not manipulated and therefore assumed to occur concurrently with the siRNA-based response. One can imagine that because the IFN pathway often leads to apoptosis of the infected cells (reviewed in Barber, 2001), it would be beneficial for the host to avoid this response as much as possible, in favor of a non-destructive RNAi-based response. The balance between the two pathways may depend on both the type of virus and the number of initial viral particles with which cells are challenged. It is unclear how the dephosphorylation activity of PIR1 could contribute to the efficacy of antiviral RNAi in conjunction with Dicer and LGP2. One interesting possibility could involve the dephosphorylation of 5' tri- and diphosphate viral dsRNAs in order to prevent recognition

by RIG-I, which has been demonstrated to specifically bind to 5'-PP and 5'-PPP ends of viral dsRNA (Hornung *et al.*, 2006; Lu *et al.*, 2010; Goubau *et al.*, 2014). This would favor an initial antiviral response based solely on RNA restriction, relegating the IFN response to a second barrier of defense that would only be activated if the initial response failed to contain viral replication. Based on our laboratory's discovery that Dicer, LGP2 and PIR1 may function together in a complex, a *PIR1* knockout mouse was created in the laboratory of Robert Finberg at UMass Medical School. Thus far, results suggest that mouse PIR1 modulates the response to viral infection, although it is still unclear how it does so, and whether it acts in synergy with the IFN pathway or as part of an independent pathway (Christopher MacKay, personal communication). Lastly, as described in Chapter III, the LGP2 helicase ortholog DRH-3, which is also required for suppression of Orsay virus replication in *C. elegans* (Ashe *et al.*, 2013), was found to stabilize the smaller PIR-1a isoform, which appears to be confined to the cytoplasm (Fig. 3.7, page 133 and Fig. 3.8, page 136). It is therefore tempting to propose that a DRH-3/PIR-1a complex could constitute the *C. elegans* equivalent of an LGP2/PIR1 complex. This possibility should not be disregarded in future work on the involvement of *C. elegans* PIR-1 in antiviral immunity.

In *Drosophila*, as in other insects, there are no direct orthologs of RIG-I, MDA5 or LGP2. Flies express two Dicer proteins: Dicer-1, which produces miRNAs, and Dicer-2, which generates siRNAs from dsRNA. Viral dsRNA is sensed by Dicer-2, which in association with the dsRNA-binding protein R2D2 (an ortholog of *C. elegans* RDE-4) generates 21-nt siRNAs that are loaded into the effector Argonaute-2 (Ago-2) to destroy viral RNA (Galiana-Arnoux *et al.*, 2006; van Rij *et al.*, 2006; Wang *et al.*, 2006; Zambon *et al.*, 2006). In addition to this activity, upon viral dsRNA binding, Dicer-2 is capable of activating the Jak-STAT (Janus Kinase and Signal Transducer and Activator of Transcription) signaling cascade, which induces a more global innate immune response, and is reminiscent of the signaling events that lead to the vertebrate IFN response (Deddouche *et al.*, 2008; Paradkar *et al.*, 2012; reviewed in Kingsolver *et al.*, 2013).

Despite the fact that in flies the association of the PIR-1 ortholog CG13197 has not yet been shown to occur with either of the organism's two Dicer proteins, we have gathered preliminary results showing that *Drosophila* PIR1 is required for suppression of FHV replication in whole animals. The experiment consisted of injecting FHV into one-day old male flies and following the accumulation of FHV RNA1 by qRT-PCR and survival of the animals over a 20-day period (unlike Orsay in *C. elegans*, FHV is lethal to

flies). In *Ago-2* and *Dcr-2* loss-of-function mutants, which served as positive controls, the viral load increased approximately 1,000-fold relative to wild-type flies. In *Drosophila CG13197* mutants, the accumulation of FHV RNA1 was three orders of magnitude higher than in the positive controls. This dramatic difference presumably results from the presence of maternally loaded PIR1 in infected animals (as with *C. elegans*, mutants are sterile and are maintained as genetically-balanced heterozygotes). This maternal load results in a later onset of death associated with infection, giving the virus more time to replicate than in the *Dcr-2* and *Ago2* mutants, which die earlier (experiments performed by Keith Boundy, in the laboratory of Phillip Zamore). While this result awaits full confirmation through rescue of the mutant by a wild-type *CG13197* transgene, it is a strong indication that *Dm* PIR1 functions in concert with Dicer-2 to provide antiviral immunity in flies. Since it is currently not known how Dicer-2 is able to distinguish non-viral dsRNA (which only triggers siRNA production) from viral dsRNA (which triggers both siRNA production and the Jak-STAT response), it is plausible that PIR-1 functions in a mechanism that allows this distinction based on the 5'-phosphorylation state of the dsRNA molecules engaged by Dicer-2.

In summary, there is strong evidence supporting the involvement of the mammalian and insect orthologs of PIR-1 in suppressing the replication of RNA viruses. The fact that PIR-1 exhibits functional conservation makes it a very attractive target for deeper studies regarding its molecular mechanisms, namely through the use of simpler and more technically amenable immortalized cell models, currently inexistent for *C. elegans*.

# MATERIALS AND METHODS

## Preparation of Orsay Virus Filtrates for Infection

Preparation of Orsay virus filtrates was performed essentially as described in Felix, *et al.*, 2011 from the original wild isolate *C. elegans* strain JU1580 (a kind gift from Dr. David Wang from Washington University in Saint Louis). Briefly, per batch of filtrate prepared, JU1580 animals were grown on six large (15 cm) NGM plates with 3 ml of concentrated OP50 per plate at 20˚C until they reached starvation. Worms were collected from the plates by washing them 2x with 15 ml of sterile 20 mM Tris-HCl pH 7.5 buffer, followed by transfer to 15 ml conical tubes and centrifugation to eliminate the animals in the suspension. This suspension was then passed through successively finer filters (5 µm, 1.2 µ and 0.22 µm) attached to a 30 ml syringe for elimination of debris and contaminating fungi and bacteria. The filtrates were aliquoted into sterile 1.5 ml microtubes and frozen at -80˚C.

## Growth and Infection of *C. elegans* Strains

In general, strains used for infection were expanded in 15 cm NGM plates and bleached for recovery of clean embryos and establishment of synchronized populations. Infections were established by mixing 20-40,000 synchronized animals (depending on stage at time of infection or stage to be collected post-infection) with 1 ml of concentrated OP50 and 150 µl of freshly thawed viral filtrate. Infection periods varied from 2-4 days at 20˚C. Strains requiring ivermectin counter-selection were selected under the same conditions as described in Chapter II Materials and Methods. For aging of non-sterile strains during or prior to infection, progeny were eliminated as L1/L2 larvae every day until final collection. This was achieved by collecting all animals from their plates in M9 buffer into 50 ml conical tubes, and letting most adult animals settle by gravity for about 2-3 minutes. Larvae remaining in suspension were aspirated and fresh M9 buffer was added. This process was repeated at least five times to eliminate all larvae. This step was required to ensure that viral mRNA would not be measured from newly infected larvae.

## Extraction of RNA

20,000-40,000 worms growing on 10 cm plates were washed 5x with M9 buffer followed by a 30-minute to 1-hour incubation in M9 buffer to eliminate as much virus as possible from the intestinal lumen and outer surface of the animals, in order to restrict our

analysis to intracellular viral RNA. The resulting pellets (100-200 µl) were resuspended in 1 ml of TRI Reagent, transferred to 1.5 ml Safe-Lock tubes (Eppendorf) with 200 µl of fine glass beads (Sigma), and vortexed vigorously for 20 minutes at 4˚C using a foam adaptor to accommodate multiple tubes. After addition of 100 µl of BCP, samples were vortexed for 2-3 minutes at room temperature and further processed according to the TRI Reagent instructions. Pellets were dissolved in ultrapure water and quantified with a NanoDrop instrument.

### qRT-PCR of Orsay Virus

cDNA synthesis and qRT-PCR were performed as described in Chapter II and SEM was calculated as described in Chapter IV. Orsay RNA1 was amplified using the primers GW194 and GW195 from Felix *et al.,* 2011 (O127/O128, Appendix B) and normalized to actin mRNA levels. The magnitude of the increase in viral RNA fold-change varied considerably from experiment to experiment. Among other potential causes for this variation, are the fact that the extent of viral RNA accumulation varies according to the stage at which animals are infected as well as the duration of infection (Sterken *et al.*, 2014), and that we could not titrate the concentration of infectious Orsay virions used in each experiment as no such method is currently available. Such variability in qRT-PCR measurements was also observed by Ashe and colleagues (2013). We therefore always compared samples that were infected in parallel under identical conditions of viral stock, worm age, and infection period.

### Small RNA Cloning

Small RNA libraries for deep sequencing were prepared as described in Chapter IV Materials and Methods.

### IP/Cloning of RDE-1- and ALG-1-Associated Orsay Viral Small RNAs

Due to the lack of enough anti-RDE-1 antibody, transgenic animals expressing GFP::RDE-1 were generated to permit IP with an anti-GFP antibody. For this, both *rde-1(ne300); unc-119(ed3);* avr3x and *pir-1(tm3198)/mnC1*; rde-1(ne300); unc-119(ed3);* avr3x strains were microinjected with a construct carrying the wild-type *C. briggsae unc-119* gene and a *gfp::rde-1* fusion under the control of its own regulatory sequences (provided by Masaki Shirayama in the laboratory). Progeny whose *unc-119* paralysis phenotype was rescued were picked and assessed for genetic transmission of the rescuing

non-integrated transgenic arrays formed by the injected constructs. Only lines with >90% transmission rates were selected to test for rescue of RNAi resistance due to the *rde-1* mutation. For this, animals were placed on *unc-22* (expressed in somatic muscle cells) or *pos-1* (germline) dsRNA-expressing bacteria as early larvae. In the case of *pir-1* animals, balanced heterozygotes were used. The majority of lines were still resistant to *pos-1* RNAi (an expected outcome due to germline silencing of high-copy transgenic arrays), but sensitive to *unc-22* RNAi with manifestation of the characteristic twitching/paralysis phenotype. GFP signal could be observed in many tissues of the selected lines and GFP::RDE-1 could be detected from IPs of *rde-1; gfp::rde-1* and ivermectin counterselected *pir-1; rde-1; gfp::rde-1* lines. Exposure of the selected *rde-1; gfp::rde-1* line to Orsay virus also led to rescue of viral suppression to wild-type levels compared *to rde-1(ne300)* alone.

For the GFP::RDE-1 IP experiment, animals were grown on ivermectin plates. The *rde-1* rescued animals were grown from L1s on OP50 with Orsay virus for 3 days and harvested as L4-young adults. *pir-1* animals were counter-selected for 7 days prior to exposure to Orsay for 3 days. For the ALG-1 IP experiment, wild-type N2 and *rde-1(ne300)* strains were used. Animals were grown for 2.5 days in the presence of Orsay and harvested as young adults. All worms were cultured at 20˚C.

The IP protocol was based on an ALG-1 IP procedure optimized in the laboratory of Victor Ambros in our department. Frozen worm pellets were thawed and lysed in 2-3 volumes of a buffer composed of HEPES-KOH, pH 7.4, 50 mM KCl, 0.1% Triton X-100, 2 mM $MgCl_2$, and 10% glycerol supplemented with 2 mM DTT, Superasin RNase inhibitor (50 µl/10 ml; Ambion), EDTA-free protease inhibitor cocktail (1 tablet/10 ml; Roche), Phostop phosphatase inhibitor cocktail (1 tablet/10 ml; Roche). 15 mg of each lysate at a 5 mg/ml concentration were split into 3 tubes and each was pre-cleared with washed beads from 200 µl of magnetic Dynabeads Protein G suspension (Life Technologies) for 1 hour at 4˚C. 80 µl of pre-cleared lysate was set aside to collect total RNA to be used as input. For GFP IPs, 5 µl (~6 µg) of monoclonal anti-GFP antibody (Wako) was used per 5 mg of protein. For ALG-1 IPs, 15 µg of polyclonal ALG-1 antibody (kindly provided by Victor Ambros) was used per 5 mg of protein. The lysates were incubated with the antibodies for 1 hour at 4˚C and then transferred to tubes with washed Dynabeads for an additional 1-hour incubation at 4˚C, using 50 µl and 125 µl of bead suspension for GFP and ALG-1 IPs, respectively. Beads were then washed 3x 10 minutes with cold wash buffer (same composition as lysis buffer except for an increased

salt concentration to 100 mM KCl) and transferred to fresh tubes (to avoid recovering nonspecific RNA stuck to the tube walls). For elution, the beads were incubated with 50 µl of TRI Reagent twice. 30 µl of ultrapure water and 20 µl of BCP phase separation reagent were added to the 100 µl of TRI Reagent and vortexed. Phase separation was carried out in 250 µl PCR tubes adapted as mini Phaselock tubes to which ~30 µl of phase-lock gel was added. This minimized loss of material from the aqueous phase. All subsequent phase separation and precipitation steps were carried out in small tubes. Centrifugations for phase separation were performed at 10,000x *g* for 4 minutes at room temperature. Precipitation took place at -20˚C overnight after the addition of 20 µg of glycogen (Roche) and 1 volume of isopropanol. Samples were centrifuged at 15,000x *g* for 15 minutes at 4˚C and then washed with cold 75% ethanol. Pellets were resuspended and pooled in 10 µl of TE buffer to which 10 µl of gel loading buffer II with formamide (Ambion) was added. For inputs, 80 µl of pre-cleared lysate was subjected to two consecutive phenol:chloroform extractions using the adapted Phase-lock PCR tubes. Precipitation was carried out by addition of 1/10 volumes 3 M sodium acetate pH 5.2, 20 µg of glycogen and 1 volume of isopropanol. The rest of the procedure was as described above. Input material was resuspended in 30 µl of TE buffer and quantified in a NanoDrop instrument. 10 µg were set aside to mix with gel loading buffer for a final volume of 20 µl. All samples were denatured at 65˚C for 5 minutes and run on a 15% polyacrylamide/7 M urea/0.5X TBE gel for size separation, followed by staining with ethidium bromide in 0.5X TBE buffer, and destaining for 10-20 minutes. Gel slices corresponding to ~18-40 nt were excised and processed as described in Chapter IV Materials and Methods.

Small RNAs were cloned using the 5'-independent TAP method described in Chapter IV with the exception of the scheme for the post-ligation PCRs which was adjusted to the very low amount of starting material. For this, different numbers of cycles were tested in the first round of PCR (12, 16, 20, 24 and 28 cycles) with 0.1 µM of the short oligos, using 2 µl of cDNA in a 50 µl reaction volume. After running 3 µl of each condition on a non-denaturing 10% polyacrylamide/0.5X TBE gel, the optimal cycle number was chosen for each sample and a 50 µl production PCR was performed and run on another non-denaturing gel. The ~70 bp products were excised from the gel (between 60 and 80 bp), eluted, precipitated and resuspended in 20 µl of Tris-HCl pH 7.5 buffer. This purification was intended to get rid of very abundant primer dimers in the IP samples which would prevent amplification of the desired sequences in the second round. For the second PCR with the long oligos containing the Illumina sequences, a 50 µl test reaction

was again performed, using oligos at 0.5 µM and 2 µl of the purified first-round PCR products. Optimal production was assessed at 4, 7, 10 and 13 PCR cycles and a 3x 50 µl production PCR was prepared for each sample, followed by gel purification between 90 and 110 bp. Purified products were diluted to 10 ng/µl and some samples were TOPO cloned and sequenced for quality control. Samples were then mixed for a 1 ng/µl final concentration of each individually barcoded library and submitted for deep-sequencing in an Illumina HiSeq instrument.

### Assembly of the Orsay Virus Genome and Small RNA Analysis

We assembled the Orsay virus genome using sRNAs derived from Orsay virus using a custom PERL script before a complete genome sequence became publicly available (Felix *et al.*, 2011). Our sequences are almost identical to the published ones with two exceptions: (1) for the viral segment encoding the alpha-delta fusion protein (capsid protein and protein delta; GenBank: HM030970.2), our sequence is 1 nt shorter due to a single deletion in the CCCCCCCCCCC region. This is likely a homopolymer mutation caused by sequencing errors in either the published sequences or our own. Otherwise the two sequences are 99.7% identical, with the remaining differences consisting of 9 different single-nucleotide polymorphisms (SNPs); (2) for the viral segment encoding the RdRP (HM030971.2), our sequence has an additional AG at the 5' end. Otherwise the two sequences are 99.7% identical with the remaining differences being 8 SNPs. The assembly custom PERL script and fully assembled sequences are available upon request.

Mapping of reads to the Orsay genome and analysis was performed as described in Chapter IV Materials and Methods, with some modifications. The direct cloning mostly cloned primary siRNAs (including viral 23-mers), 21U-RNAs and miRNAs. However, the TAP procedure cloned not only secondary triphosphate 22G-RNAs, but also monophosphate viral primary siRNAs that were 22-nt long and began with a G. To remove contamination of primary 22G-RNAs from secondary 22G-RNAs, only 22G-RNAs preceded with C or T (the ‘YR’ motif described in the Results section) were considered as real secondary 22G-RNAs, because we previously demonstrated that secondary 22G-RNAs prefer a non-coded 5'C or T. For the RDE-1 IP, miR-243 was used to normalize sRNA reads, because miR-243 specifically interacts with RDE-1 and can therefore be used as a good indicator for the extent of RDE-1 enrichment in each sample. For non-IP experiments, all miRNAs were used for normalization unless otherwise noted.

# CONCLUSION

This work constitutes the first characterization of the highly conserved metazoan RNA 5' phosphatase PIR1/DUSP11 in a whole animal. Our studies uncovered two distinct biological functions for *C. elegans* PIR-1. As expected from its association with the ribonuclease Dicer (Duchaine *et al.*, 2006), both of these functions are related to the production of small RNAs. In the context of a stable association with the sRNA-producing ERI complex *in vivo*, we found PIR-1 to be required in the germline for the biogenesis of nematode-specific endogenous primary 26G sRNAs, which promote proper sperm cell differentiation (Han *et al.*, 2009; Pavelec *et al.*, 2009; Conine *et al.*, 2010). In somatic cells, and upon infection by the naturally-occurring *C. elegans* Orsay RNA virus (Felix *et al.*, 2011), we found PIR-1 to be required for the suppression of viral replication by participating in an RNAi-like antiviral pathway. These functions correlate well with the nearly ubiquitous expression of PIR-1 along development in both the soma and germline. Additionally, we showed that endogenous germline sRNAs targeting protein-coding transcripts are significantly downregulated in *pir-1* mutant animals, suggesting that PIR-1 may promote their synthesis and consequently play a potentially crucial role in the coordination of germline sRNA-dependent processes. Experimental constraints, however, precluded further investigation of this interesting result.

Importantly, we have established that *C. elegans* PIR-1 indeed functions as an RNA triphosphatase *in vitro*, converting 5'-triphosphorylated RNA to 5'-monophosphorylated molecules, and that, similarly to its human ortholog, this activity is abolished when the active site cysteine residue is substituted (Deshpande *et al.*, 1999). *In vivo*, both functions that PIR-1 promotes depend on an intact catalytic site. Moreover, the developmental arrest and sterility manifested in *pir-1* null mutant animals cannot be overcome by catalytically inactive PIR-1. As neither of the processes in which we have implicated PIR-1 is required for general development, we conclude that PIR-1 possesses additional functions that we have not yet been able to unveil.

Since RNA-dependent RNA polymerases in *C. elegans* generate 5'-triphosphorylated RNA products starting with a guanosine (Aoki *et al.*, 2007; Pak and Fire, 2007; Sijen *et al.*, 2007), we presume that the precursors of 26G-RNAs that are synthesized by the RdRP RRF-3 are originally triphosphorylated. The discovery that PIR-1 is necessary for the accumulation of these 5'-monophosphorylated RNAs solves the mystery of how 26G-RNAs lose their two terminal phosphates. This modification could be necessary for loading of 26G-RNAs into the correct Argonaute proteins, or, alternatively, to facilitate Dicer-mediated cleavage of RNA duplexes formed by RRF-3 from template

mRNAs. In the context of viral infection, *pir-1* mutant animals are unable to generate enough antiviral 22G-RNAs to suppress replication of the virus. We propose that by converting the ends of triphosphorylated (+) and (-) viral RNA strands to monophosphates, viral replication/transcription intermediates may be stabilized in a dsRNA form. This would provide Dicer with enough substrate molecules to generate abundant primary viral siRNAs, which would in turn lead to a potent secondary amplification of silencing 22G-RNAs. The common denominator between these two PIR-1-dependent processes is the need for a dsRNA substrate that Dicer can cleave. It is therefore possible that the molecular activity of PIR-1 serves the same purpose in these two distinct pathways. While we have not gathered evidence that this is the case, experiments designed to measure the accumulation of dsRNA intermediates in different mutants should be considered in the future. Recent studies have successfully characterized dsRNA populations at a genome-wide level in *C. elegans* through the use of dsRNA-enrichment strategies such as ssRNA-specific RNases to digest total RNA, or dsRNA immunoprecipitation with anti-dsRNA antibodies, followed by cloning and deep sequencing (Li *et al.*, 2012; Saldi *et al.*, 2014; Whipple *et al.*, 2015). The potential for PIR-1 to regulate dsRNA levels should be studied using animals undergoing spermatogenesis or viral infection, in both wild-type and mutant genetic backgrounds where dsRNA may be more stable (*e.g.*, in a *dcr-1* null background or in the *dcr-1(mg375)* helicase mutant, in which 26G-RNA production is specifically abrogated (Pavelec *et al.*, 2009; Welker *et al.*, 2010)).

We have demonstrated here that PIR-1 strongly concentrates in the nucleus of most cells, where it also associates with chromatin. We found that PIR-1 maintains its interactions with Dicer and other components of the ERI complex in nuclear protein extracts. Most interestingly, when we profiled the sRNA content of wild-type nuclei, we found that PIR-1-dependent 26G-RNAs were the only species to show an appreciable enrichment, strongly suggesting that synthesis of Dicer-dependent sRNAs occurs in the nuclei of *C. elegans*. In further support of this possibility, one of the two main protein isoforms of PIR-1 that we were able to identify, is highly enriched in nuclei and strictly requires Dicer for stability. These findings add to the increasing body of work in diverse organisms that demonstrates the pervasiveness of sRNA biogenesis and effector pathways in cell nuclei (reviewed in Cecere and Grishok, 2014; Huang and Li, 2014; Schraivogel and Meister, 2014). Additionally, we observed that cells undergoing high DNA replication and transcription, such as the endoreduplicating cells of the intestine or the mitotic cells of the hermaphrodite distal germline, contain the highest levels of nuclear PIR-1. This

correlates well with the impaired growth and germline underproliferation phenotypes of *pir-1* mutant animals. As human PIR1 has also been shown to localize to nuclei, to interact with proteins implicated in RNA processing, and to participate in rRNA processing (Yuan *et al.*, 1998; Andersen *et al.*, 2005; Caprara *et al.*, 2009; Tafforeau *et al.*, 2013), we would ultimately like to know whether PIR-1 can function in processes beyond the realm of sRNAs to modulate growth and development in all animals. Therefore, potential alterations in splicing patterns and transcription levels of mRNAs, as well as in levels and processing states of non-coding RNAs (such as long non-coding RNAs, rRNAs, snRNAs, snoRNAs and tRNAs) should be addressed in greater detail, as such changes could plausibly lead to the phenotypes observed in *pir-1* loss-of-function mutant animals.

One surprising result of this study was the specific upregulation of innate immunity and stress response gene expression in *pir-1* mutant animals. This matched the results of a prior study where a similar gene expression change was identified in loss-of-function mutants for DCR-1, RDE-4 and RDE-1, but for which the mechanistic basis and biological purpose were not found (Welker *et al.*, 2007). The effect is unlikely to derive from the downregulation of miRNAs, as Dicer is the only enzyme of the four required for their accumulation. Given that (1) PIR-1 may have an important role in dsRNA accumulation destined for Dicer cleavage of primary sRNAs, that (2) RDE-4 binds dsRNA and may assist Dicer in the cleavage/loading process (Parrish and Fire, 2001; Tabara *et al.*, 2002; Parker *et al.*, 2006), and that (3) RDE-1 is the Argonaute into which primary siRNAs are loaded (Yigit *et al.*, 2006; Steiner *et al.*, 2007), the four proteins may cooperate to silence the innate immune response by an RNAi-like mechanism, analogous to the suppression of viral replication. This raises the hypothesis that during a viral infection these factors are co-opted to combat the virus, loosening the repression of innate immunity response genes. The result would be the amplification of the antiviral response by the concomitant activity of the sRNA-based pathway and of a broader, sRNA-independent immune response. As *C. elegans* lacks the potent IFN response that is central to antiviral defense in vertebrates (reviewed in Stifter and Feng, 2015), studying the expression of these genes during acute viral infection in diverse genetic backgrounds could potentially lead to the discovery of complementary antiviral immune pathways in nematodes and other invertebrates. But perhaps more interestingly, this would demonstrate the existence of an incredibly well-orchestrated mechanism of gene regulation whereby the very same proteins required for the suppression of viral infection have the ability to constitutively repress gene expression programs that should only be activated during an infection. Such a regulatory mechanism

would fit well with the fact that PIR-1 is constantly and ubiquitously expressed, as if in a permanent state of alert to incoming viruses.

Notably, we have gathered preliminary evidence strongly suggesting that the antiviral immunity role of PIR-1 is conserved. In a *Drosophila* mutant of the PIR-1 counterpart, Flock House Virus replicates uncontrollably (Keith Boundy, unpublished results), as it does in *Dicer-2* and *Argonaute-2* loss-of-function mutants (reviewed in Ding, 2010; Kingsolver *et al.*, 2013). In human cells, we have shown that PIR1 physically interacts with Dicer. Furthermore, both Dicer and PIR1 are specifically pulled down by LGP2 (Darryl Conte, unpublished results), an important dsRNA binding cytoplasmic receptor implicated in the response against RNA viruses in vertebrates (reviewed in Rodriguez *et al.*, 2014). The link to LGP2 supports the hypothesis that mammalian PIR1 also plays a role in the defense against RNA viruses, perhaps as part of the antiviral RNAi response already demonstrated to occur in mice (Li *et al.*, 2013; Maillard *et al.*, 2013). Regarding other pathways, functional conservation is less certain: insects and vertebrates seem to lack RdRPs, which are central to PIR-1-dependent sRNA biogenesis. However, a strong indication that functional conservation extends beyond antiviral immunity stems from the observation that adult PIR-1-deficient flies are unable to reproduce, implying that *Drosophila* PIR-1 may participate in central aspects of germline development and/or gamete differentiation. As more examples of Dicer-dependent endogenous sRNA regulatory pathways emerge in flies and vertebrates (reviewed in Piatek and Werner, 2014), the potential involvement of PIR-1 in these processes should not be disregarded.

Is PIR-1 activity circumscribed to sRNA pathways or does it participate in broader gene regulatory processes in the nucleus? The answer to this fundamental question will depend on the identification of the RNAs and chromatin loci with which PIR-1 associates at different stages along development. Future experiments must extend our efforts to identify PIR-1 targets by employing iPAR-CLIP (i*n vivo P*hoto*a*ctivatable *R*ibonucleoside-Enhanced UV *C*ross-*L*inking and *IP*; Rybak-Wolf *et al.*, 2014) and optimized versions of ChIP-seq (*Ch*romatin *I*mmuno*p*recipitation followed by deep-sequencing of associated DNA). Such experiments should focus on PIR-1 obtained from isolated nuclei, since we believe that the sensitivity of these methods is decreased by the large pool of PIR-1 residing in the cytoplasm. As both *C. elegans* and cultured cells are currently equally amenable to these techniques, the parallel study and comparison of results obtained from both systems would be especially powerful. The success of these approaches would compensate for limitations such as the inheritance of a variable maternal load of PIR-1 in

*pir-1* mutant animals, the inability to silence *pir-1* by RNAi, and the inexistence of conditional mutants, all of which prevented us from studying the effects arising from lack of PIR-1 at distinct developmental stages.

Lastly, in addition to the questions that this study raised, there are some evolutionary peculiarities regarding this phosphatase that are equally worthy of future inquiry. For instance, what is the role of the PIR-1 paralog F54C8.4? Orthologs of this protein can be found throughout the nematode lineage, but are also present in insects, suggesting that it evolved specifically in the common ancestor of this group of animals (both are protostomes grouped in *ecdysozoa*). Since the catalytic site of F54C8.4 contains the conserved asparagine that distinguishes RNA 5' triphosphatases from protein tyrosine phosphatases, it is presumed to have *in vivo* RNA 5' phosphatase activity (see Fig. 1.6, page 62; Changela *et al.*, 2005). Intrigued by this, we superficially characterized F54C8.4, showing that (1) its mRNA expression levels across development are similar to those of *pir-1*, that (2) homozygote animals for a null deletion allele die as early-stage larvae after about seven days, and that (3) RNAi results in embryonic lethality. Unlike PIR-1, F54C8.4 was not identified via proteomic analysis of DCR-1 complexes. Studying F54C8.4 could therefore reveal important parallels and differences to help us understand PIR-1 function across the evolutionary spectrum of metazoans. Another intriguing PIR-1-related evolutionary mystery concerns the presence of homologous RNA 5' phosphatases in animal-infecting dsDNA viruses from the *Baculoviridae* and *Poxviridae* families (the latter comprising not only invertebrate-, but also vertebrate-infecting viruses, such as Vaccinia and Smallpox). Why did these viruses acquire RNA triphosphatases when they already encode their own mRNA capping machinery? Are they used to promote viral expression, or are they used to regulate or disable host systems? The fact these viruses incorporated and still maintain PIR-1-like enzymes in their genomes lends strength to the idea that they must perform important roles in all metazoans.

We believe that our demonstration of specific biological functions for *C. elegans* PIR-1 will propel further studies to understand how it promotes the synthesis of Dicer-dependent sRNAs and to address its essential requirement for growth and development. Ultimately, we hope that our work will spark a renewed interest in the biology of this multi-functional RNA phosphatase in more complex animals.

# REFERENCES

Abbott, A.L. (2011). Uncovering new functions for microRNAs in Caenorhabditis elegans. Current biology *21*, R668-671.

Alder, M.N. (2003). Gene silencing in Caenorhabditis elegans by transitive RNA interference. Rna *9*, 25-32.

Albers, H.M., Kuijl, C., Bakker, J., Hendrickx, L., Wekker, S., Farhou, N., Liu, N., Blasco-Moreno, B., Scanu, T., den Hertog, J., Celie, P., Ovaa, H., Neefjes, J. (2014) Integrating chemical and genetic silencing strategies to identify host kinase-phosphatase inhibitor networks that control bacterial infection. ACS Chemical biology *9*,:414-422.

Alonso, A., Sasin, J., Bottini, N., Friedberg, I., Friedberg, I., Osterman, A., Godzik, A., Hunter, T., Dixon, J., and Mustelin, T. (2004). Protein tyrosine phosphatases in the human genome. Cell *117*, 699-711.

Alvarez-Saavedra, E., and Horvitz, H.R. (2010). Many families of C. elegans microRNAs are not essential for development or viability. Current biology *20*, 367-373.

Altun, Z.F., Herndon, L.A., Crocker, C., Lints, R. and Hall, D.H. (ed.s) 2002-2015. WormAtlas, http://www.wormatlas.org

Ambros, V. (2003) MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. Cell 113, 673-676.

Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science *328*, 1534-1539.

Andersen, J.S., Lam, Y.W., Leung, A.K., Ong, S.E., Lyon, C.E., Lamond, A.I., and Mann, M. (2005). Nucleolar proteome dynamics. Nature *433*, 77-83.

Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K., and Tabara, H. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. EMBO journal *26*, 5007-5019.

Ashe, A., Belicard, T., Le Pen, J., Sarkies, P., Frezal, L., Lehrbach, N.J., Felix, M.A., and Miska, E.A. (2013). A deletion polymorphism in the Caenorhabditis elegans RIG-I homolog disables viral RNA dicing and antiviral immunity. eLife *2*, e00994.

Ashe, A., Sapetschnig, A., Weick, E.M., Mitchell, J., Bagijn, M.P., Cording, A.C., Doebley, A.L., Goldstein, L.D., Lehrbach, N.J., Le Pen, J., *et al.* (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. Cell *150*, 88-99.

Asikainen, S., Storvik, M., Lakso, M., and Wong, G. (2007). Whole genome microarray analysis of C. elegans rrf-3 and eri-1 mutants. FEBS letters *581*, 5050-5054.

Audhya, A., Hyndman, F., McLeod, I.X., Maddox, A.S., Yates, J.R., 3rd, Desai, A., and Oegema, K. (2005). A complex containing the Sm protein CAR-1 and the RNA helicase CGH-1 is required for embryonic cytokinesis in Caenorhabditis elegans. The Journal of cell biology *171*, 267-279.

Austin, J., and Kimble, J. (1987). glp-1 is required in the germ line for regulation of the decision between mitosis and meiosis in C. elegans. Cell *51*, 589-599.

Avgousti, D.C., Palani, S., Sherman, Y., and Grishok, A. (2012). CSR-1 RNAi pathway positively regulates histone expression in C. elegans. The EMBO journal *31*, 3821-3832.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes & development *22*, 2773-2785.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., Pasquinelli, A.E. (2005) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell *122*, 553-563.

Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.M., Bouasker, S., Lehrbach, N.J., Simard, M.J., and Miska, E.A. (2012). Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science *337*, 574-578.

Ball, L.A. (1994). Replication of the genomic RNA of a positive-strand RNA animal virus from negative-sense transcripts. Proceedings of the National Academy of Sciences of the United States of America *91*, 12443-12447.

Bamming, D., and Horvath, C.M. (2009). Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. The Journal of biological chemistry *284*, 9700-9712.

Barber, G.N. (2001) Host defense, viruses and apoptosis.Cell death and differentiation. *8*, 113-126.

Barraud, P., Emmerth, S., Shimada, Y., Hotz, H.R., Allain, F.H., and Buhler, M. (2011). An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. The EMBO journal *30*, 4223-4235.

Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. Cell *136*215-233.

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., *et al.* (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Molecular cell *31*, 67-78.

Bender, L.B., Cao, R., Zhang, Y., and Strome, S. (2004). The MES-2/MES-3/MES-6 complex and regulation of histone H3 methylation in C. elegans. Current biology *14*, 1639-1643.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Berry, L.W., Westlund, B., and Schedl, T. (1997). Germ-line tumor formation caused by activation of glp-1, a Caenorhabditis elegans member of the Notch family of receptors. Development *124*, 925-936.

Beshore, E.L., McEwen, T.J., Jud, M.C., Marshall, J.K., Schisa, J.A., and Bennett, K.L. (2011). C. elegans Dicer interacts with the P-granule component GLH-1 and both regulate germline RNPs. Developmental biology *350*, 370-381.

Billi, A.C., Alessi, A.F., Khivansara, V., Han, T., Freeberg, M., Mitani, S., and Kim, J.K. (2012). The Caenorhabditis elegans HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. PLoS genetics *8*, e1002617.

Billi, A.C., Freeberg, M.A., Day, A.M., Chun, S.Y., Khivansara, V., and Kim, J.K. (2013). A conserved upstream motif orchestrates autonomous, germline-enriched expression of Caenorhabditis elegans piRNAs. PLoS genetics *9*, e1003392.

Blanchard, D., Parameswaran, P., Lopez-Molina, J., Gent, J., Saynuk, J.F., and Fire, A. (2014). On the nature of in vivo requirements forrde-4in RNAi and developmental pathways in C. elegans. RNA biology *8*, 458-467.

Blumenthal, T. (2012). Trans-splicing and operons in C. elegans. WormBook : the online review of C elegans biology, 1-11.

Bouasker, S., Simard, M.J. (2012) The slicing activity of miRNA-specific Argonautes is essential for the miRNA pathway in C. elegans. Nucleic acids research. *40*, 10452-10462.

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.

Brodigan, T.M., Liu, J., Park, M., Kipreos, E.T., and Krause, M. (2003). Cyclin E expression during development in Caenorhabditis elegans. Developmental biology *254*, 102-115.

Bruns, A.M., Horvath, C.M. (2014) Antiviral RNA recognition and assembly by RLR family innate immune sensors. Cytokine growth factor reviews. *25*, 507-512.

Bruns, A.M., and Horvath, C.M. (2015). LGP2 synergy with MDA5 in RLR-mediated RNA recognition and antiviral signaling. Cytokine *74*, 198-206.

Buck, A.H., and Blaxter, M. (2013). Functional diversification of Argonautes in nematodes: an expanding universe. Biochemical Society transactions *41*, 881-886.

Buckley, B.A., Burkhart, K.B., Gu, S.G., Spracklin, G., Kershner, A., Fritz, H., Kimble, J., Fire, A., and Kennedy, S. (2012). A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. Nature *489*, 447-451.

Burke, S.L., Hammell, M., and Ambros, V. (2015). Robust Distal Tip Cell Pathfinding in the Face of Temperature Stress Is Ensured by Two Conserved microRNAs in Caenorhabditis elegans. Genetics *115*,179184.

Burkhart, K.B., Guang, S., Buckley, B.A., Wong, L., Bochner, A.F., and Kennedy, S. (2011). A pre-mRNA-associating factor links endogenous siRNAs to chromatin regulation. PLoS genetics *7*, e1002249.

Burton, N.O., Burkhart, K.B., and Kennedy, S. (2011). Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America *108*, 19683-19688.

Campbell, A.C., and Updike, D.L. (2015). CSR-1 and P granules suppress sperm-specific transcription in the C. elegans germline. Development *142*, 1745-1755.

Caprara, G., Zamponi, R., Melixetian, M., and Helin, K. (2009). Isolation and characterization of DUSP11, a novel p53 target gene. Journal of cellular and molecular medicine *13*, 2158-2170.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nature reviews Genetics *14*, 100-112.

Castel, S.E., Ren, J., Bhattacharjee, S., Chang, A.Y., Sanchez, M., Valbuena, A., Antequera, F., and Martienssen, R.A. (2014). Dicer promotes transcription termination at sites of replication stress to maintain genome stability. Cell *159*, 572-583.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., *et al.* (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393-1406.

Caudy, A.A., Ketting, R.F., Hammond, S.M., Denli, A.M., Bathoorn, A.M., Tops, B.B., Silva, J.M., Myers, M.M., Hannon, G.J., and Plasterk, R.H. (2003). A micrococcal nuclease homologue in RNAi effector complexes. Nature *425*, 411-414.

Cecere, G., and Grishok, A. (2014). A nuclear perspective on RNAi pathways in metazoans. Biochimica et biophysica acta *1839*, 223-233.

Cecere, G., Hoersch, S., O'Keeffe, S., Sachidanandam, R., and Grishok, A. (2014). Global effects of the CSR-1 RNA interference pathway on the transcriptional landscape. Nature structural & molecular biology *21*, 358-365.

Cecere, G., Zheng, G.X., Mansisidor, A.R., Klymko, K.E., and Grishok, A. (2012). Promoters recognized by forkhead proteins exist for individual 21U-RNAs. Molecular cell *47*, 734-745.
Chan, Y.K., and Gack, M.U. (2015). RIG-I-like receptor regulation in virus infection and immunity. Current opinion in virology *12*, 7-14.

Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M., Micklem, G., Piano, F., Snyder, M., Stein, L., White, K.P., Waterston, R.H.; modENCODE Consortium. (2009) Unlocking the secrets of the genome. Nature. *459*, 927-930.

Chan, Y.K., Gack, M.U. (2015) RIG-I-like receptor regulation in virus infection and immunity. Current opinion in virology *12*, 7-14.

Changela, A., Martins, A., Shuman, S., and Mondragon, A. (2005). Crystal structure of baculovirus RNA triphosphatase complexed with phosphate. The Journal of biological chemistry *280*, 17848-17856.

Chen, C.C., Simard, M.J., Tabara, H., Brownell, D.R., McCollough, J.A., and Mello, C.C. (2005). A member of the polymerase beta nucleotidyltransferase superfamily is required for RNA interference in C. elegans. Current biology *15*, 378-383.

Chesney, M.A., Lam, N., Morgan, D.E., Phillips, B.T., and Kimble, J. (2009). C. elegans HLH-2/E/Daughterless controls key regulatory cells during gonadogenesis. Developmental biology *331*, 14-25.

Chintapalli, V.R., Wang, J., Dow, J.A.T. (2007) Using FlyAtlas to identify better Drosophila models of human disease. Nature genetics *39*, 715-720.

Chu, C.Y., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. PLoS biology *4*, e210.

Chu, D.S., Liu, H., Nix, P., Wu, T.F., Ralston, E.J., Yates, J.R., 3rd, and Meyer, B.J. (2006). Sperm chromatin proteomics identifies evolutionarily conserved fertility factors. Nature *443*, 101-105.

Claycomb, J.M., Batista, P.J., Pang, K.M., Gu, W., Vasale, J.J., van Wolfswinkel, J.C., Chaves, D.A., Shirayama, M., Mitani, S., Ketting, R.F., *et al.* (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell *139*, 123-134.

Colmenares, S.U., Buker, S.M., Buhler, M., Dlakic, M., and Moazed, D. (2007). Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. Molecular cell *27*, 449-461.

Conine, C.C., Batista, P.J., Gu, W., Claycomb, J.M., Chaves, D.A., Shirayama, M., and Mello, C.C. (2010). Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America *107*, 3588-3593.

Conine, C.C., Moresco, J.J., Gu, W., Shirayama, M., Conte, D., Jr., Yates, J.R., 3rd, and Mello, C.C. (2013). Argonautes promote male fertility and provide a paternal memory of germline gene expression in C. elegans. Cell *155*, 1532-1544.

Correa, R.L., Steiner, F.A., Berezikov, E., and Ketting, R.F. (2010). MicroRNA-directed siRNA biogenesis in Caenorhabditis elegans. PLoS genetics *6*, e1000903.

Crittenden, S.L., Troemel, E.R., Evans, T.C., and Kimble, J. (1994). GLP-1 is localized to the mitotic region of the C. elegans germ line. Development *120*, 2901-2911.

Dang, Y., Yang, Q., Xue, Z., Liu, Y. (2011) RNA interference in fungi: pathways, functions, and applications. Eukaryotic cell *10*, 1148-1155.

Dardousis, K., Voolstra, C., Roengvoraphoj, M., Sekandarzad, A., Mesghenna, S., Winkler, J., Ko, Y., Hescheler, J., Sachinidis, A. (2007) Identification of differentially expressed genes involved in the formation of multicellular tumor spheroids by HT-29 colon carcinoma cells. Molecular therapy *15*, 94-102.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., *et al.* (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Molecular cell *31*, 79-90.

de Albuquerque, B.F., Luteijn, M.J., Cordeiro Rodrigues, R.J., van Bergeijk, P., Waaijers, S., Kaaij, L.J., Klein, H., Boxem, M., and Ketting, R.F. (2014). PID-1 is a novel factor that operates during 21U-RNA biogenesis in Caenorhabditis elegans. Genes & development *28*, 683-688.

De Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic acids research *34*(Web Server issue), W362-5.

de Hoon, M., and Hayashizaki, Y. (2008). Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. BioTechniques *44*, 627-628, 630, 632.

Deddouche, S., Matt, N., Budd, A., Mueller, S., Kemp, C., Galiana-Arnoux, D., Dostert, C., Antoniewski, C., Hoffmann, J.A., and Imler, J.L. (2008). The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in drosophila. Nature immunology *9*, 1425-1432.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

Dent, J.A., Smith, M.M., Vassilatis, D.K., and Avery, L. (2000). The genetics of ivermectin resistance in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America *97*, 2674-2679.

Deshpande, T., Takagi, T., Hao, L., Buratowski, S., and Charbonneau, H. (1999). Human PIR1 of the protein-tyrosine phosphatase superfamily has RNA 5'-triphosphatase and diphosphatase activities. The Journal of biological chemistry *274*, 16590-16594.

Ding, L., Spencer, A., Morita, K., and Han, M. (2005). The developmental timing regulator AIN-1 interacts with miRISCs and may target the argonaute protein ALG-1 to cytoplasmic P bodies in C. elegans. Molecular cell *19*, 437-447.

Ding, S.W. (2010). RNA-based antiviral immunity. Nature reviews Immunology *10*, 632-644.
Diogo, J., and Bratanich, A. (2014). The nematode Caenorhabditis elegans as a model to study viruses. Archives of virology *159*, 2843-2851.

Doyle, M., Badertscher, L., Jaskiewicz, L., Guttinger, S., Jurado, S., Hugenschmidt, T., Kutay, U., and Filipowicz, W. (2013). The double-stranded RNA binding domain of human Dicer functions as a nuclear localization signal. Rna *19*, 1238-1252.

Drake, M., Furuta, T., Suen, K.M., Gonzalez, G., Liu, B., Kalia, A., Ladbury, J.E., Fire, A.Z., Skeath, J.B., and Arur, S. (2014). A Requirement for ERK-Dependent Dicer Phosphorylation in Coordinating Oocyte-to-Embryo Transition in C. elegans. Developmental cell *31*, 614-628.

Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte, D., Jr., Pang, K., Brownell, D.R., Harding, S., Mitani, S., Ruvkun, G., *et al.* (2006). Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. Cell *124*, 343-354.

Ellis, R.E., Stanfield, G.M. (2014) The regulation of spermatogenesis and sperm function in nematodes. Seminars in cell and developmental biology *29*, 17-30.

Emmerth, S., Schober, H., Gaidatzis, D., Roloff, T., Jacobeit, K., and Buhler, M. (2010). Nuclear retention of fission yeast dicer is a prerequisite for RNAi-mediated heterochromatin assembly. Developmental cell *18*, 102-113.

Ermolaeva, M.A., and Schumacher, B. (2014). Insights from the worm: the C. elegans model for innate immunity. Seminars in immunology *26*, 303-309.

Errett, J.S., and Gale, M. (2015). Emerging complexity and new roles for the RIG-I-like receptors in innate antiviral immunity. Virologica Sinica *30*, 163-173.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.F., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. Genes & development *21*, 2558-2570.

Ezkurdia, I., Juan, D., Rodriguez, J.M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., Tress, M.L. (2014) Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Human molecular genetics *23*, 5866-5878.

Fares, H., and Greenwald, I. (2001). Genetic analysis of endocytosis in Caenorhabditis elegans: coelomocyte uptake defective mutants. Genetics *159*, 133-145.

Fay, D.S., Han, M. (2000) Mutations in cye-1, a Caenorhabditis elegans cyclin E homolog, reveal coordination between cell-cycle control and vulval development. Development *127*, 4049-4060.

Felix, M.A., Ashe, A., Piffaretti, J., Wu, G., Nuez, I., Belicard, T., Jiang, Y., Zhao, G., Franz, C.J., Goldstein, L.D., *et al.* (2011). Natural and experimental infection of Caenorhabditis nematodes by novel viruses related to nodaviruses. PLoS biology *9*, e1000586.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-11.

Fischer, S.E., Butler, M.D., Pan, Q., and Ruvkun, G. (2008). Trans-splicing in C. elegans generates the negative RNAi regulator ERI-6/7. Nature *455*, 491-496.

Fischer, S.E., Montgomery, T.A., Zhang, C., Fahlgren, N., Breen, P.C., Hwang, A., Sullivan, C.M., Carrington, J.C., and Ruvkun, G. (2011). The ERI-6/7 helicase acts at the first stage of an siRNA amplification pathway that targets recent gene duplications. PLoS genetics *7*, e1002369.

Flynt, A., Liu, N., Martin, R., and Lai, E.C. (2009). Dicing of viral replication intermediates during silencing of latent Drosophila viruses. Proceedings of the National Academy of Sciences of the United States of America *106*, 5270-5275.

Francis, R., Barton, M.K., Kimble, J., and Schedl, T. (1995). gld-1, a tumor suppressor gene required for oocyte development in Caenorhabditis elegans. Genetics *139*, 579-606.

Frank, F., Hauver, J., Sonenberg, N., and Nagar, B. (2012). Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. The EMBO journal *31*, 3588-3595.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818-822.

Franz, C.J., Renshaw, H., Frezal, L., Jiang, Y., Felix, M.A., and Wang, D. (2014). Orsay, Santeuil and Le Blanc viruses primarily infect intestinal cells in Caenorhabditis nematodes. Virology *448*, 255-264.

Frokjaer-Jensen, C., Davis, M.W., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen, S.P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. Nature genetics *40*, 1375-1383.

Fuller-Pace, F.V. (2006) DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic acids research *34*, 4206-4215.

Gabel, H.W., and Ruvkun, G. (2008). The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. Nature structural & molecular biology *15*, 531-533.

Gagnon, K.T., Li, L., Chu, Y., Janowski, B.A., and Corey, D.R. (2014). RNAi factors are present and active in human cell nuclei. Cell reports *6*, 211-221.

Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J.A., and Imler, J.L. (2006). Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. Nature immunology *7*, 590-597.

Gallo, C.M., Munro, E., Rasoloson, D., Merritt, C., and Seydoux, G. (2008). Processing bodies and germ granules are distinct RNA granules that interact in C. elegans embryos. Developmental biology *323*, 76-87.

Gassmann, R., Rechtsteiner, A., Yuen, K.W., Muroyama, A., Egelhofer, T., Gaydos, L., Barron, F., Maddox, P., Essex, A., Monen, J., *et al.* (2012). An inverse relationship to germline transcription defines centromeric chromatin in C. elegans. Nature *484*, 534-537.

Gent, J.I., Lamm, A.T., Pavelec, D.M., Maniar, J.M., Parameswaran, P., Tao, L., Kennedy, S., and Fire, A.Z. (2010). Distinct phases of siRNA synthesis in an endogenous RNAi pathway in C. elegans soma. Molecular cell *37*, 679-689.

Gent, J.I., Schvarzstein, M., Villeneuve, A.M., Gu, S.G., Jantsch, V., Fire, A.Z., and Baudrimont, A. (2009). A Caenorhabditis elegans RNA-directed RNA polymerase in sperm development and endogenous RNA interference. Genetics *183*, 1297-1314.

Ghildiyal, M., Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. Nature reviews genetics *10*, 94-108.

Giannone, R.J., McDonald, H.W., Hurst, G.B., Shen, R.F., Wang, Y., and Liu, Y. (2010). The protein network surrounding the human telomere repeat binding factors TRF1, TRF2, and POT1. PloS one *5*, e12407.

Goh, W.S., Seah, J.W., Harrison, E.J., Chen, C., Hammell, C.M., and Hannon, G.J. (2014). A genome-wide RNAi screen identifies factors required for distinct stages of C. elegans piRNA biogenesis. Genes & development *28*, 797-807.

Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A.J., Zillinger, T., Goldeck, M., Schuberth, C., Van der Veen, A.G., Fujimura, T., Rehwinkel, J., *et al.* (2014). Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5'-diphosphates. Nature *514*, 372-375.

Grishok, A. (2005). RNAi mechanisms in Caenorhabditis elegans. FEBS letters *579*, 5932-5939.

Grishok, A., Tabara, H., Mello, C.C. (2000) Genetic requirements for inheritance of RNAi in C. elegans. Science *287*, 2494-2497.

Gu, S.G., Pak, J., Guang, S., Maniar, J.M., Kennedy, S., and Fire, A. (2012). Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. Nature genetics *44*, 157-164.

Gu, W., Lee, H.C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., Jr., and Mello, C.C. (2012b). CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. Cell *151*, 1488-1500.

Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., *et al.* (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Molecular cell *36*, 231-244.

Guang, S., Bochner, A.F., Burkhart, K.B., Burton, N., Pavelec, D.M., and Kennedy, S. (2010). Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. Nature *465*, 1097-1101.

Guang, S., Bochner, A.F., Pavelec, D.M., Burkhart, K.B., Harding, S., Lachowiec, J., and Kennedy, S. (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. Science *321*, 537-541.

Guo, X., Zhang, R., Wang, J., Ding, S.W., and Lu, R. (2013). Homologous RIG-I-like helicase proteins direct RNAi-mediated antiviral immunity in C. elegans by distinct mechanisms. Proceedings of the National Academy of Sciences of the United States of America *110*, 16085-16090.

Guo, Y.R., Hryc, C.F., Jakana, J., Jiang, H., Wang, D., Chiu, W., Zhong, W., and Tao, Y.J. (2014). Crystal structure of a nematode-infecting virus. Proceedings of the National Academy of Sciences of the United States of America *111*, 12781-12786.

Hammell, C.M., Lubin, I., Boag, P.R., Blackwell, T.K., and Ambros, V. (2009). nhl-2 Modulates microRNA activity in Caenorhabditis elegans. Cell *136*, 926-938.

Han, T., Manoharan, A.P., Harkins, T.T., Bouffard, P., Fitzpatrick, C., Chu, D.S., Thierry-Mieg, D., Thierry-Mieg, J., and Kim, J.K. (2009). 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America *106*, 18674-18679.

Harris, T.W., Baran, J., Bieri, T., Cabunoc, A., Chan, J., Chen, W.J., Davis, P., Done, J., Grove, C., Howe, K., Kishore, R., Lee, R., Li, Y., Muller, H.M., Nakamura, C., Ozersky, P., Paulini, M., Raciti, D., Schindelman, G., Tuli, M.A., Van Auken, K., Wang, D., Wang, X., Williams, G., Wong, J.D., Yook, K., Schedl, T., Hodgkin, J., Berriman, M., Kersey, P., Spieth, J., Stein, L., Sternberg, P. W. (2014) WormBase 2014: new views of curated biology. Nucleic acids research, *42*, D789-793.

Häsler, R., Kerick, M., Mah, N., Hultschig, C., Richter, G., Bretz, F., Sina, C., Lehrach, H., Nietfeld, W., Schreiber, S., Rosenstiel, P. (2011) Alterations of pre-mRNA splicing in human inflammatory bowel disease. European journal of cell biology *90*, 603-611.

Hasnain, S.E., Jain, A., Habib, S., Ghosh, S., Chatterji, U., Ramachandran, A., Das, P., Venkaiah, B., Pandey, S., Liang, B., *et al.* (1997). Involvement of host factors in transcription from baculovirus very late promoters -- a review. Gene *190*, 113-118.

He, Y., and Smith, R. (2009). Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. Cellular and molecular life sciences *66*, 1239-1256.

Hoefig, K.P., Rath, N., Heinz, G.A., Wolf, C., Dameris, J., Schepers, A., Kremmer, E., Ansel, K.M., and Heissmeyer, V. (2013). Eri1 degrades the stem-loop of oligouridylated histone mRNAs to induce replication-dependent decay. Nature structural & molecular biology *20*, 73-81.

Hornung, V., Ellegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K.K., Schlee, M., *et al.* (2006). 5'-Triphosphate RNA is the ligand for RIG-I. Science *314*, 994-997.

Hu, P.J. (2007) Dauer. WormBook *8*,1-19.

Huang, V., and Li, L.C. (2014). Demystifying the nuclear function of Argonaute proteins. RNA biology *11*, 18-24.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nature reviews Molecular cell biology *9*, 22-32.

Inukai, S., and Slack, F. (2013). MicroRNAs and the genetic network in aging. Journal of molecular biology *425*, 3601-3608.

Jiang, F., Ramanathan, A., Miller, M.T., Tang, G.Q., Gale, M., Jr., Patel, S.S., and Marcotrigiano, J. (2011). Structural basis of RNA recognition and activation by innate immune receptor RIG-I. Nature *479*, 423-427.

Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., and Liu, Q. (2005). Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. Genes & development *19*, 1674-1679.

Jiang, H., Franz, C.J., Wu, G., Renshaw, H., Zhao, G., Firth, A.E., and Wang, D. (2014). Orsay virus utilizes ribosomal frameshifting to express a novel protein that is incorporated into virions. Virology *450-451*, 213-221.

Jonas, S., Izaurralde, E. (2015) Towards a molecular understanding of microRNA-mediated gene silencing. Nature reviews genetics *16*, 421-433.

Jones, A.R., Francis, R., and Schedl, T. (1996). GLD-1, a cytoplasmic protein essential for oocyte differentiation, shows stage- and sex-specific expression during Caenorhabditis elegans germline development. Developmental biology *180*, 165-183.

Juang, B.T., Gu, C., Starnes, L., Palladino, F., Goga, A., Kennedy, S., and L'Etoile, N.D. (2013). Endogenous nuclear RNAi mediates behavioral adaptation to odor. Cell *154*, 1010-1022.

Jungkamp, A.C., Stoeckius, M., Mecenas, D., Grun, D., Mastrobuoni, G., Kempa, S., and Rajewsky, N. (2011). In vivo and transcriptome-wide identification of RNA binding protein target sites. Molecular cell *44*, 828-840.

Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., Kadonaga, J.T. (2008) The RNA polymerase II core promoter - the gateway to transcription. Current opinion in cell biology *20*, 253-259.

Kamita, S.G., Nagasaka, K., Chua, J.W., Shimada, T., Mita, K., Kobayashi, M., Maeda, S., and Hammock, B.D. (2005). A baculovirus-encoded protein tyrosine phosphatase gene induces enhanced locomotory activity in a lepidopteran host. Proceedings of the National Academy of Sciences of the United States of America *102*, 2584-2589.

Kamminga, L.M., Luteijn, M.J., den Broeder, M.J., Redl, S., Kaaij, L.J., Roovers, E.F., Ladurner, P., Berezikov, E., and Ketting, R.F. (2010). Hen1 is required for oocyte development and piRNA stability in zebrafish. The EMBO journal *29*, 3688-3700.

Kamminga, L.M., van Wolfswinkel, J.C., Luteijn, M.J., Kaaij, L.J., Bagijn, M.P., Sapetschnig, A., Miska, E.A., Berezikov, E., and Ketting, R.F. (2012). Differential impact of the HEN1 homolog HENN-1 on 21U and 26G RNAs in the germline of Caenorhabditis elegans. PLoS genetics *8*, e1002702.

Kennedy, S., Wang, D., and Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. Nature *427*, 645-649.

Kelly, W.G., Xu, S., Montgomery, M.K., and Fire, A. (1997). Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics *146*, 227-238.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes & development *15*, 2654-2659.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133-141.

Kim, D., Weaver, R.F. (1993) Transcription mapping and functional analysis of the protein tyrosine/serine phosphatase (PTPase) gene of the Autographa californica nuclear polyhedrosis virus.Virology *195*, 587-595.

Kim, J.K., Gabel, H.W., Kamath, R.S., Tewari, M., Pasquinelli, A., Rual, J.F., Kennedy, S., Dybbs, M., Bertin, N., Kaplan, J.M. et al. (2005). Functional genomic analysis of RNA interference in C. elegans. Science *308*, 1164-1167.

Kimble, J., and Crittenden, S.L. (2007). Controls of germline stem cells, entry into meiosis, and the sperm/oocyte decision in Caenorhabditis elegans. Annual review of cell and developmental biology *23*, 405-433.

Kingsolver, M.B., Huang. Z., Hardy, R.W. (2013) Insect antiviral innate immunity: pathways, effectors, and connections. Journal of molecular biology *425*, 4921-4936.

Knight, S.W., and Bass, B.L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science *293*, 2269-2271.

Ko, S., Kawasaki, I., and Shim, Y.H. (2013). PAB-1, a Caenorhabditis elegans poly(A)-binding protein, regulates mRNA metabolism in germline by interacting with CGH-1 and CAR-1. PloS one *8*, e84798.

Ko, S., Park, J.H., Lee, A.R., Kim, E., Jiyoung, K., Kawasaki, I., and Shim, Y.H. (2010). Two mutations in pab-1 encoding poly(A)-binding protein show similar defects in germline stem cell proliferation but different longevity in C. elegans. Molecules and cells *30*, 167-172.

Kopek, B.G., Perkins, G., Miller, D.J., Ellisman, M.H., and Ahlquist, P. (2007). Three-dimensional analysis of a viral RNA replication complex reveals a virus-induced mini-organelle. PLoS biology *5*, e220.

Korta, D.Z., and Hubbard, E.J. (2010). Soma-germline interactions that influence germline proliferation in Caenorhabditis elegans. Developmental dynamics : an official publication of the American Association of Anatomists *239*, 1449-1459.

Kudron, M.M., and Reinke, V. (2008). C. elegans nucleostemin is required for larval growth and germline stem cell division. PLoS genetics *4*, e1000181.

Kuznicki, K.A., Smith, P.A., Leung-Chiu, W.M., Estevez, A.O., Scott, H.C., and Bennett, K.L. (2000). Combinatorial RNA interference indicates GLH-4 can compensate for GLH-1; these two P granule components are critical for fertility in C. elegans. Development *127*, 2907-2916.

Kuzuoglu-Ozturk, D., Huntzinger, E., Schmidt, S., and Izaurralde, E. (2012). The Caenorhabditis elegans GW182 protein AIN-1 interacts with PAB-1 and subunits of the PAN2-PAN3 and CCR4-NOT deadenylase complexes. Nucleic acids research *40*, 5651-5665.

Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E., and Corbett, A.H. (2007). Classical nuclear localization signals: definition, function, and interaction with importin alpha. The Journal of biological chemistry *282*, 5101-5105.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G.
(2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Lee, H.C., Gu, W., Shirayama, M., Youngman, E., Conte, D., Jr., and Mello, C.C. (2012). C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. Cell *150*, 78-87.

Lee, M., Kim, B., and Kim, V.N. (2014). Emerging roles of RNA modification: m(6)A and U-tail. Cell *158*, 980-987.

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., Hammell, C.M., and Ambros, V. (2006). Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. Rna *12*, 589-597.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Leuschner, P.J., Ameres, S.L., Kueng, S., and Martinez, J. (2006). Cleavage of the siRNA passenger strand during RISC assembly in human cells. EMBO reports *7*, 314-320.

Li, H., Li, W.X., and Ding, S.W. (2002). Induction and suppression of RNA silencing by an animal virus. Science *296*, 1319-1321.

Li, S., Wang, L., Berman, M., Kong, Y.Y., and Dorf, M.E. (2011). Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. Immunity *35*, 426-440.

Li, Y., and Guarino, L.A. (2008). Roles of LEF-4 and PTP/BVP RNA triphosphatases in processing of baculovirus late mRNAs. Journal of virology *82*, 5573-5583.

Li, Y., Miller, L.K. (1995a).Expression and localization of a baculovirus protein phosphatase. Journal of general virology *76*, 2941-2948.

Li, Y., Lu, J., Han, Y., Fan, X., and Ding, S.W. (2013). RNA interference functions as an antiviral immunity mechanism in mammals. Science *342*, 231-234.

Liao, J.Y., Ma, L.M., Guo, Y.H., Zhang, Y.C., Zhou, H., Shao, P., Chen, Y.Q., and Qu, L.H. (2010). Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. PloS one *5*, e10563.

Lin, T., Meng, L., Li, Y., and Tsai, R.Y. (2010). Tumor-initiating function of nucleostemin-enriched mammary tumor cells. Cancer research *70*, 9444-9452.

Linder, P., Jankowsky, E. (2011) From unwinding to clamping - the DEAD box RNA helicase family. Nature reviews molecular and cell biology *12*, 505-516.

Lozano, E., Saez, A.G., Flemming, A.J., Cunha, A., and Leroi, A.M. (2006). Regulation of growth by ploidy in Caenorhabditis elegans. Current biology *16*, 493-498.

Lu, C., Xu, H., Ranjith-Kumar, C.T., Brooks, M.T., Hou, T.Y., Hu, F., Herr, A.B., Strong, R.K., Kao, C.C., and Li, P. (2010). The structural basis of 5' triphosphate double-stranded RNA recognition by RIG-I C-terminal domain. Structure *18*, 1032-1043.

Lu, R., Maduro, M., Li, F., Li, H.W., Broitman-Maduro, G., Li, W.X., and Ding, S.W. (2005). Animal virus replication and RNAi-mediated antiviral silencing in Caenorhabditis elegans. Nature *436*, 1040-1043.

Lu, R., and Wang, G.G. (2013). Tudor: a versatile family of histone methylation 'readers'. Trends in biochemical sciences *38*, 546-555.

Lu, R., Yigit, E., Li, W.X., and Ding, S.W. (2009). An RIG-I-Like RNA helicase mediates antiviral RNAi downstream of viral siRNA biogenesis in Caenorhabditis elegans. PLoS pathogens *5*, e1000286.

Lui, D.Y., Colaiácovo, M.P. (2013) Meiotic development in Caenorhabditis elegans. Advances in experimental medicine and biology *757*, 133-170.

Luo, D., Ding, S.C., Vela, A., Kohlway, A., Lindenbach, B.D., and Pyle, A.M. (2011). Structural insights into RNA recognition by RIG-I. Cell *147*, 409-422.

Luteijn, M.J., and Ketting, R.F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nature reviews Genetics *14*, 523-534.

Luteijn, M.J., van Bergeijk, P., Kaaij, L.J., Almeida, M.V., Roovers, E.F., Berezikov, E., and Ketting, R.F. (2012). Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. The EMBO journal *31*, 3422-3430.

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature *434*, 666-670.

Ma, L., and Horvitz, H.R. (2009). Mutations in the Caenorhabditis elegans U2AF large subunit UAF-1 alter the choice of a 3' splice site in vivo. PLoS genetics *5*, e1000708.

MacRae, I.J., Zhou, K., and Doudna, J.A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. Nature structural & molecular biology *14*, 934-940.

Maillard, P.V., Ciaudo, C., Marchais, A., Li, Y., Jay, F., Ding, S.W., and Voinnet, O. (2013). Antiviral RNA interference in mammalian cells. Science *342*, 235-238.

Maine, E.M., Hauth, J., Ratliff, T., Vought, V.E., She, X., and Kelly, W.G. (2005). EGO-1, a putative RNA-dependent RNA polymerase, is required for heterochromatin assembly on unpaired dna during C. elegans meiosis. Current biology *15*, 1972-1978.

Martins, A., and Shuman, S. (2000). Mechanism of phosphoanhydride cleavage by baculovirus phosphatase. The Journal of biological chemistry *275*, 35070-35076.

Martins, A., and Shuman, S. (2002a). The Domain Order of Mammalian Capping Enzyme Can Be Inverted and Baculovirus Phosphatase Can Function in Cap Formation in Vivo. Virology *304*, 167-175.

Martins, A., and Shuman, S. (2002b). Mutational analysis of baculovirus phosphatase identifies structural residues important for triphosphatase activity in vitro and in vivo. Biochemistry *41*, 13403-13409.

Matranga, C., and Pyle, A.M. (2010). Double-stranded RNA-dependent ATPase DRH-3: insight into its role in RNAsilencing in Caenorhabditis elegans. The Journal of biological chemistry *285*, 25363-25371.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607-620.

McDonald, W.H., Tabb, D.L., Sadygov, R.G., MacCoss, M.J., Venable, J., Graumann J., Johnson J.R., Cociorva, D., Yates, J.R. 3rd. (2004) MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. Rapid communications in mass spectrometry *18*, 2162-2168.

McJunkin, K., and Ambros, V. (2014). The embryonic mir-35 family of microRNAs promotes multiple aspects of fecundity in Caenorhabditis elegans. G3 (Bethesda, Md) *4*, 1747-1754.

Meister, G., Landthaler, M., Peters, L., Chen, P.Y., Urlaub, H., Luhrmann, R., and Tuschl, T. (2005). Identification of novel argonaute-associated proteins. Current biology *15*, 2149-2155.

Mello, C.C., Kramer, J.M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. The EMBO journal *10*, 3959-3970.

Merritt, C., Rasoloson, D., Ko, D., and Seydoux, G. (2008). 3' UTRs are the primary regulators of gene expression in the C. elegans germline. Current biology *18*, 1476-1482.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., *et al.* (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell *133*, 116-127.

Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. Rna *10*, 1692-1694.

Miller, D.J., and Ahlquist, P. (2002). Flock House Virus RNA Polymerase Is a Transmembrane Protein with Amino-Terminal Sequences Sufficient for Mitochondrial Localization and Membrane Insertion. Journal of virology *76*, 9856-9867.

Miller, D.J., Schwartz, M.D., and Ahlquist, P. (2001). Flock house virus RNA replicates on outer mitochondrial membranes in Drosophila cells. Journal of virology *75*, 11664-11676.

Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D., Ambros, V., and Horvitz, H.R. (2005). Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. PLoS genetics e215.

Moerman, D.G., Benian, G.M., Barstead, R.J., Schriefer, L.A., and Waterston, R.H. (1988). Identification and intracellular localization of the unc-22 gene product of Caenorhabditis elegans. Genes & development *2*, 93-105.

Montgomery, T.A., Rim, Y.S., Zhang, C., Dowen, R.H., Phillips, C.M., Fischer, S.E., and Ruvkun, G. (2012). PIWI associated siRNAs and piRNAs specifically require the Caenorhabditis elegans HEN1 ortholog henn-1. PLoS genetics *8*, e1002616.

Moresco, E.M., and Beutler, B. (2010). LGP2: positive about viral sensing. Proceedings of the National Academy of Sciences of the United States of America *107*, 1261-1262.

Morgan, D.E., Crittenden, S.L., and Kimble, J. (2010). The C. elegans adult male germline: stem cells and sexual dimorphism. Developmental biology *346*, 204-214.

Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789-802.

Myong, S., Cui, S., Cornish, P.V., Kirchhofer, A., Gack, M.U., Jung, J.U., Hopfner, K.P., and Ha, T. (2009). Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. Science *323*, 1070-1074.

Nakamura, M., Ando, R., Nakazawa, T., Yudazono, T., Tsutsumi, N., Hatanaka, N., Ohgake, T., Hanaoka, F., and Eki, T. (2007). Dicer-related drh-3 gene functions in germ-line development by maintenance of chromosomal integrity in Caenorhabditis elegans. Genes to cells : devoted to molecular & cellular mechanisms *12*, 997-1010.

Pak, J., and Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science *315*, 241-244.

Pak, J., Maniar, J.M., Mello, C.C., and Fire, A. (2012). Protection from feed-forward amplification in an amplified RNAi mechanism. Cell *151*, 885-899.

Paradkar, P.N., Trinidad, L., Voysey, R., Duchemin, J.B., and Walker, P.J. (2012). Secreted Vago restricts West Nile virus infection in Culex mosquito cells by activating the Jak-STAT pathway. Proceedings of the National Academy of Sciences of the United States of America *109*, 18915-18920.

Parameswaran, P., Sklan, E., Wilkins, C., Burgon, T., Samuel, M.A., Lu, R., Ansel, K.M., Heissmeyer, V., Einav, S., Jackson, W., *et al.* (2010). Six RNA viruses and forty-one hosts: viral small RNAs and modulation of small RNA repertoires in vertebrate and invertebrate systems. PLoS pathogens *6*, e1000764.

Park, S.K., Venable, J.D., Xu, T., Yates, J.R. 3rd. (2008) A quantitative analysis software tool for mass spectrometry-based proteomics. Nature methods. *5*, 319-322.

Parker, G.S., Eckert, D.M., and Bass, B.L. (2006). RDE-4 preferentially binds long dsRNA and its dimerization is necessary for cleavage of dsRNA to siRNA. Rna *12*, 807-818.

Parker, G.S., Maity, T.S., and Bass, B.L. (2008). dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi. Journal of molecular biology *384*, 967-979.

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. Molecular cell *25*, 635-646.

Paro, S., Imler, J.L., and Meignin, C. (2015). Sensing viral RNAs by Dicer/RIG-I like ATPases across species. Current opinion in immunology *32C*, 106-113.

Parrish, S., and Fire, A. (2001). Distinct roles for RDE-1 and RDE-4 during RNA interference in Caenorhabditis elegans. Rna *7*, 1397-1402.

Parry, D.H., Xu, J., and Ruvkun, G. (2007). A whole-genome RNAi Screen for C. elegans miRNA pathway genes. Current biology *17*, 2013-2022.

Patterson, K.I., Brummer, T., O'Brien, P.M., Daly, R.J. (2009) Dual-specificity phosphatases: critical regulators with diverse cellular targets. Biochemistry journal *418*, 475-489.

Pavelec, D.M., Lachowiec, J., Duchaine, T.F., Smith, H.E., and Kennedy, S. (2009). Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in Caenorhabditis elegans. Genetics *183*, 1283-1295.

Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., Gygi, S.P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. Journal of proteome research *2*, 43-50.

Pepper, A.S., Killian, D.J., and Hubbard, E.J. (2003). Genetic analysis of Caenorhabditis elegans glp-1 mutants suggests receptor interaction or competition. Genetics *163*, 115-132.

Petryszak, R., Burdett, T., Fiorelli, B., Fonseca, N.A., Gonzalez-Porta, M., Hastings, E., Huber, W., Jupp, S., Keays, M., Kryvych, N., McMurry, J., Marioni, J.C., Malone, J., Megy, K., Rustici, G., Tang, A.Y., Taubert, J., Williams, E., Mannion, O., Parkinson, H.E., Brazma, A. (2014) Expression Atlas update - a database of gene and transcript expression from microarray and sequencing-based functional genomics experiments. Nucleic acids research *42*, D926-32.

Phillips, C.M., McDonald, K.L., and Dernburg, A.F. (2009). Cytological analysis of meiosis in Caenorhabditis elegans. Methods in molecular biology *558*, 171-195.

Phillips, C.M., Montgomery, T.A., Breen, P.C., and Ruvkun, G. (2012). MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the C. elegans germline. Genes & development *26*, 1433-1444.

Piano, F., Schetter, A.J., Morton, D.G., Gunsalus, K.C., Reinke, V., Kim, S.K., and Kemphues, K.J. (2002). Gene clustering based on RNAi phenotypes of ovary-enriched genes in C. elegans. Current biology *12*, 1959-1964.

Piatek, M.J., and Werner, A. (2014). Endogenous siRNAs: regulators of internal affairs. Biochemical Society transactions *42*, 1174-1179.

Pitt, J.N., Schisa, J.A., and Priess, J.R. (2000). P granules in the germ cells of Caenorhabditis elegans adults are associated with clusters of nuclear pores and contain RNA. Developmental biology *219*, 315-333.

Praitis, V., Casey, E., Collar, D., and Austin, J. (2001). Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. Genetics *157*, 1217-1226.

Qu, J., and Bishop, J.M. (2012). Nucleostemin maintains self-renewal of embryonic stem cells and promotes reprogramming of somatic cells to pluripotency. The Journal of cell biology *197*, 731-745.

Raizen, D.M., Lee, R.Y., and Avery, L. (1995). Interacting genes required for pharyngeal excitation by motor neuron MC in Caenorhabditis elegans. Genetics *141*, 1365-1382.

Rajyaguru, P., and Parker, R. (2009). CGH-1 and the control of maternal mRNAs. Trends in cell biology *19*, 24-28.

Ren, Z., and Ambros, V.R. (2015). Caenorhabditis elegans microRNAs of the let-7 family act in innate immune response circuits and confer robust developmental timing against pathogen stress. Proceedings of the National Academy of Sciences of the United States of America *112*, E2366-2375.

Rice, P., Longden, I., Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends in genetics. *16*, 276-277.

Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nature structural & molecular biology *12*, 340-349.

Robert, V.J., Sijen, T., van Wolfswinkel, J., and Plasterk, R.H. (2005). Chromatin and RNAi factors protect the C. elegans germline against repetitive sequences. Genes & development *19*, 782-787.

Rocheleau, C.E., Cullison, K., Huang, K., Bernstein, Y., Spilker, A.C., and Sundaram, M.V. (2008). The Caenorhabditis elegans ekl (enhancer of ksr-1 lethality) genes include putative components of a germline small RNA pathway. Genetics *178*, 1431-1443.

Rocheleau, C.E., Yasuda, J., Shin, T.H., Lin, R., Sawa, H., Okano, H., Priess, J.R., Davis, R.J., and Mello, C.C. (1999). WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in C. elegans. Cell *97*, 717-726.

Rodriguez, K.R., Bruns, A.M., and Horvath, C.M. (2014). MDA5 and LGP2: accomplices and antagonists of antiviral signal transduction. Journal of virology *88*, 8194-8200.

Rual, J.F., Ceron, J., Koreth, J., Hao, T., Nicot, A.S., Hirozane-Kishikawa, T., Vandenhaute, J., Orkin, S.H., Hill, D.E., van den Heuvel, S., *et al.* (2004). Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. Genome research *14*, 2162-2168.

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.

Rybak-Wolf, A., Jens, M., Murakawa, Y., Herzog, M., Landthaler, M., and Rajewsky, N. (2014). A Variety of Dicer Substrates in Human and C. elegans. Cell *159*, 1153-1167.

Saijou, E., Fujiwara, T., Suzaki, T., Inoue, K., and Sakamoto, H. (2004). RBD-1, a nucleolar RNA-binding protein, is essential for Caenorhabditis elegans early development through 18S ribosomal RNA processing. Nucleic acids research *32*, 1028-1036.

Saldi, T.K., Ash, P.E., Wilson, G., Gonzales, P., Garrido-Lecca, A., Roberts, C.M., Dostal, V., Gendron, T.F., Stein, L.D., Blumenthal, T., *et al.* (2014). TDP-1, the Caenorhabditis elegans ortholog of TDP-43, limits the accumulation of double-stranded RNA. The EMBO journal *33*, 2947-2966.

Sanchez-Jimenez, F., and Sanchez-Margalet, V. (2013). Role of Sam68 in post-transcriptional gene regulation. International journal of molecular sciences *14*, 23402-23419.

Sankhala, R.S., Lokareddy, R.K., and Cingolani, G. (2014). Structure of human PIR1, an atypical dual-specificity phosphatase. Biochemistry *53*, 862-871.

Sarkies, P., Ashe, A., Le Pen, J., McKie, M.A., and Miska, E.A. (2013). Competition between virus-derived and endogenous small RNAs regulates gene expression in Caenorhabditis elegans. Genome research *23*, 1258-1270.

Satoh, T., Kato, H., Kumagai, Y., Yoneyama, M., Sato, S., Matsushita, K., Tsujimura, T., Fujita, T., Akira, S., and Takeuchi, O. (2010). LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. Proceedings of the National Academy of Sciences of the United States of America *107*, 1512-1517.

Schaner, C.E., and Kelly, W.G. (2006). Germline chromatin. WormBook : the online review of C elegans biology, 1-14.

Schisa, J.A., Pitt, J.N., and Priess, J.R. (2001). Analysis of RNA associated with P granules in germ cells of C. elegans adults. Development *128*, 1287-1298.

Schott, D.H., Cureton, D.K., Whelan, S.P., and Hunter, C.P. (2005). An antiviral role for the RNA interference machinery in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America *102*, 18420-18424.

Schraivogel, D., Meister, G. (2014) Import routes and nuclear functions of Argonaute and other small RNA-silencing proteins. Trends in biochemical science *39*, 420-31.

Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2013). The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. Developmental cell *27*, 656-663.

Seydoux, G., and Dunn, M.A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of Caenorhabditis elegans and Drosophila melanogaster. Development *124*, 2191-2201.

Seydoux, G., Mello, C.C., Pettitt, J., Wood, W.B., Priess, J.R., and Fire, A. (1996). Repression of gene expression in the embryonic germ lineage of C. elegans. Nature *382*, 713-716.

She, X., Xu, X., Fedotov, A., Kelly, W.G., and Maine, E.M. (2009). Regulation of heterochromatin assembly on unpaired chromosomes during Caenorhabditis elegans meiosis by components of a small RNA-mediated pathway. PLoS genetics *5*, e1000624.

Sheng, Z., and Charbonneau, H. (1993). The baculovirus Autographa californica encodes a protein tyrosine phosphatase. The Journal of biological chemistry *268*, 4728-4733.

Sheth, U., Pitt, J., Dennis, S., and Priess, J.R. (2010). Perinuclear P granules are the principal sites of mRNA export in adult C. elegans germ cells. Development *137*, 1305-1314.

Shi, Z., Montgomery, T.A., Qi, Y., and Ruvkun, G. (2013). High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. Genome research *23*, 497-508.

Shih, J.D., and Hunter, C.P. (2011). SID-1 is a dsRNA-selective dsRNA-gated channel. Rna *17*, 1057-1065.

Shirayama, M., Seth, M., Lee, H.C., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell *150*, 65-77.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell *107*, 465-476.

Sijen, T., Steiner, F.A., Thijssen, K.L., and Plasterk, R.H. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science *315*, 244-247.

Sijen, T., and Plasterk, R.H. (2003). Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. Nature *426*, 310-314.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S.P., Nonet, M.L., Fire, A., Ahringer, J., and Plasterk, R.H. (2002). Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi. Current biology *12*, 1317-1319.

Sinha, N.K., Trettin, K.D., Aruscavage, P.J., and Bass, B.L. (2015). Drosophila dicer-2 cleavage is mediated by helicase- and dsRNA termini-dependent states that are modulated by Loquacious-PD. Molecular cell *58*, 406-417.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nature reviews Molecular cell biology *12*, 246-258.

Smale, S.T., Baltimore, D. (1989) The "initiator" as a transcription control element. Cell *57*,103-113.

Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Current biology *10*, 169-178.

Soennichsen, B., Koski, L.B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.M., Artelt, J., Bettencourt, P., Cassin, E. et al. (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature *434*, 462-469.

Song, R., Hennig, G.W., Wu, Q., Jose, C., Zheng, H., and Yan, W. (2011). Male germ cells express abundant endogenous siRNAs. Proceedings of the National Academy of Sciences of the United States of America *108*, 13159-13164.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harbor perspectives in biology *3*.

Stadler, M., Artiles, K., Pak, J., and Fire, A. (2012). Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of C. elegans heterochronic miRNA targets. Genome research *22*, 2418-2426.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A., Lewis, S. (2002) The generic genome browser: a building block for a model organism system database. Genome research *12*, 1599-1610.

Stein, P., Svoboda, P., and Schultz, R.M. (2003). Transgenic RNAi in mouse oocytes: a simple and fast approach to study gene function. Developmental biology *256*, 188-194.

Steiner, F.A., Okihara, K.L., Hoogstrate, S.W., Sijen, T., and Ketting, R.F. (2009). RDE-1 slicer activity is required only for passenger-strand cleavage during RNAi in Caenorhabditis elegans. Nature structural & molecular biology *16*, 207-211.

Sterken, M.G., Snoek, L.B., Bosman, K.J., Daamen, J., Riksen, J.A., Bakker, J., Pijlman, G.P., and Kammenga, J.E. (2014). A heritable antiviral RNAi response limits Orsay virus infection in Caenorhabditis elegans N2. PloS one *9*, e89760.

Stifter, S.A., and Feng, C.G. (2015). Interfering with immunity: detrimental role of type I IFNs during infection. Journal of immunology *194*, 2455-2465.

Stoeckius, M., Grun, D., and Rajewsky, N. (2014). Paternal RNA contributions in the Caenorhabditis elegans zygote. The EMBO journal *33*, 1740-1750.

Stoeckius, M., Maaskola, J., Colombo, T., Rahn, H.P., Friedlander, M.R., Li, N., Chen, W., Piano, F., and Rajewsky, N. (2009). Large-scale sorting of C. elegans embryos reveals the dynamics of small RNA expression. Nature methods *6*, 745-751.

Svoboda, P. (2014). Renaissance of mammalian endogenous RNAi. FEBS letters *588*, 2550-2556.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.

Tabara, H., Yigit, E., Siomi, H., and Mello, C.C. (2002). The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. Cell *109*, 861-871.

Tabb, D.L., McDonald, W.H., Yates, J.R. 3rd (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. Journal of proteome research *1*, 21-26.

Tafforeau, L., Zorbas, C., Langhendries, J.L., Mullineux, S.T., Stamatopoulou, V., Mullier, R., Wacheul, L., Lafontaine, D.L. (2013) The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. Molecular cell *51*, 539-551.

Tahbaz, N., Kolb, F.A., Zhang, H., Jaronczyk, K., Filipowicz, W., and Hobman, T.C. (2004). Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. EMBO reports *5*, 189-194.

Takagi, T., Moore, C.R., Diehn, F., and Buratowski, S. (1997). An RNA 5'-triphosphatase related to the protein tyrosine phosphatases. Cell *89*, 867-873.

Takagi, T., Taylor, G.S., Kusakabe, T., Charbonneau, H., and Buratowski, S. (1998). A protein tyrosine phosphatase-like protein from baculovirus has RNA 5'-triphosphatase and diphosphatase activities. Proceedings of the National Academy of Sciences of the United States of America *95*, 9808-9812.

Takagi, T., Walker, A.K., Sawa, C., Diehn, F., Takase, Y., Blackwell, T.K., and Buratowski, S. (2003). The Caenorhabditis elegans mRNA 5'-capping enzyme. In vitro and in vivo characterization. The Journal of biological chemistry *278*, 14174-14184.

Takahasi, K., Kumeta, H., Tsuduki, N., Narita, R., Shigemoto, T., Hirai, R., Yoneyama, M., Horiuchi, M., Ogura, K., Fujita, T., *et al.* (2009). Solution structures of cytosolic RNA sensor MDA5 and LGP2 C-terminal domains: identification of the RNA recognition loop in RIG-I-like receptors. The Journal of biological chemistry *284*, 17465-17474.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. Plant & cell physiology *49*, 493-500.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., *et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534-538.

te Velthuis, A.J. (2014). Common and unique features of viral RNA-dependent polymerases. Cellular and molecular life sciences : CMLS *71*, 4403-4420.

Thivierge, C., Makil, N., Flamand, M., Vasale, J.J., Mello, C.C., Wohlschlegel, J., Conte, D., Jr., and Duchaine, T.F. (2012). Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. Nature structural & molecular biology *19*, 90-97.

Tian, Y., Simanshu, D.K., Ma, J.B., and Patel, D.J. (2011). Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. Proceedings of the National Academy of Sciences of the United States of America *108*, 903-910.

Tijsterman, M., Okihara, K.L., Thijssen, K., and Plasterk, R.H. (2002b). PPW-1, a PAZ/PIWI protein required for efficient germline RNAi, is defective in a natural isolate of C. elegans. Current biology *12*, 1535-1540.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004). A protein sensor for siRNA asymmetry. Science *306*, 1377-1380.

Tops, B.B., Tabara, H., Sijen, T., Simmer, F., Mello, C.C., Plasterk, R.H., and Ketting, R.F. (2005). RDE-2 interacts with MUT-7 to mediate RNA interference in Caenorhabditis elegans. Nucleic acids research *33*, 347-355.

Tsai, H.Y., Chen, C.C., Conte, D., Jr., Moresco, J.J., Chaves, D.A., Mitani, S., Yates, J.R., 3rd, Tsai, M.D., and Mello, C.C. (2015). A Ribonuclease Coordinates siRNA Amplification and mRNA Cleavage during RNAi. Cell *160*, 407-419.

Tsai, R.Y. (2014). Turning a new page on nucleostemin and self-renewal. Journal of cell science *127*, 3885-3891.

Tsai, R.Y., Meng, L. (2009) Nucleostemin: a latecomer with new tricks. International journal of biochemistry and cell biology *41*, 2122-2124.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., Pontén, F. (2015) Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

Updike, D., and Strome, S. (2010). P granule assembly and function in Caenorhabditis elegans germ cells. Journal of andrology *31*, 53-60.

van Rij, R.P., Saleh, M.C., Berry, B., Foo, C., Houk, A., Antoniewski, C., and Andino, R. (2006). The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes & development *20*, 2985-2995.

van Wolfswinkel, J.C., Claycomb, J.M., Batista, P.J., Mello, C.C., Berezikov, E., and Ketting, R.F. (2009). CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. Cell *139*, 135-148.

Vasale, J.J., Gu, W., Thivierge, C., Batista, P.J., Claycomb, J.M., Youngman, E.M., Duchaine, T.F., Mello, C.C., and Conte, D., Jr. (2010). Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. Proceedings of the National Academy of Sciences of the United States of America *107*, 3582-3587.

Vasquez-Rifo, A., Jannot, G., Armisen, J., Labouesse, M., Bukhari, S.I., Rondeau, E.L., Miska, E.A., and Simard, M.J. (2012). Developmental characterization of the microRNA-specific C. elegans Argonautes alg-1 and alg-2. PloS one *7*, e33750.

Vastenhouw, N.L., Fischer, S.E.J., Robert, V.J.P., Thijssen, K.L., Fraser, A.G., Kamath, R.S., Ahringer, J., and Plasterk, R.H.A. (2003). A Genome-Wide Screen Identifies 27 Genes Involved in Transposon Silencing in C. elegans. Current Biology *13*, 1311-1316.

Venter, P.A., and Schneemann, A. (2008). Recent insights into the biology and biomedical applications of Flock House virus. Cellular and molecular life sciences : CMLS *65*, 2675-2687.

Vought, V.E., Ohmachi, M., Lee, M.H., and Maine, E.M. (2005). EGO-1, a putative RNA-directed RNA polymerase, promotes germline proliferation in parallel with GLP-1/notch signaling and regulates the spatial organization of nuclear pore complexes and germline P granules in Caenorhabditis elegans. Genetics *170*, 1121-1132.

Walker, A.K., Boag, P.R., and Blackwell, T.K. (2007). Transcription reactivation steps stimulated by oocyte maturation in C. elegans. Developmental biology *304*, 382-393.

Wang, G., and Reinke, V. (2008). A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. Current biology *18*, 861-867.

Wang, X.H., Aliyari, R., Li, W.X., Li, H.W., Kim, K., Carthew, R., Atkinson, P., and Ding, S.W. (2006). RNA interference directs innate immunity against viruses in adult Drosophila. Science *312*, 452-454.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., *et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Wedeles, C.J., Wu, M.Z., and Claycomb, J.M. (2013). Protection of germline gene expression by the C. elegans Argonaute CSR-1. Developmental cell *27*, 664-671.

Welker, N.C., Maity, T.S., Ye, X., Aruscavage, P.J., Krauchuk, A.A., Liu, Q., and Bass, B.L. (2011). Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. Molecular cell *41*, 589-599.

Welker, N.C., Pavelec, D.M., Nix, D.A., Duchaine, T.F., Kennedy, S., and Bass, B.L. (2010). Dicer's helicase domain is required for accumulation of some, but not all, C. elegans endogenous siRNAs. Rna *16*, 893-903.

Wightman, B., Ha, I., 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.

Whipple, J.M., Youssef, O.A., Aruscavage, P.J., Nix, D.A., Hong, C., Johnson, W.E., and Bass, B.L. (2015). Genome-wide profiling of the C. elegans dsRNAome. Rna *21*, 786-800.

White, E., Schlackow, M., Kamieniarz-Gdula, K., Proudfoot, N.J., and Gullerova, M. (2014). Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. Nature structural & molecular biology *21*, 552-559.

Wianny, F., and Zernicka-Goetz, M. (2000). Specific interference with gene function by double-stranded RNA in early mouse development. Nature cell biology *2*, 70-75.

Wilkins, C., Dishongh, R., Moore, S.C., Whitt, M.A., Chow, M., and Machaca, K. (2005). RNA interference is an antiviral defence mechanism in Caenorhabditis elegans. Nature *436*, 1044-1047.

Winston, W.M., Molodowitch, C., and Hunter, C.P. (2002). Systemic RNAi in C. elegans requires the putative transmembrane protein SID-1. Science *295*, 2456-2459.

Winston, W.M., Sutherlin, M., Wright, A.J., Feinberg, E.H., and Hunter, C.P. (2007). Caenorhabditis elegans SID-2 is required for environmental RNA interference. Proceedings of the National Academy of Sciences of the United States of America *104*, 10565-10570.

Wolters, D.A., Washburn, M.P., Yates, J.R. 3rd (2001) An automated multidimensional protein identification technology for shotgun proteomics. Analytical chemistry *73*, 5683-5690.

Woolcock, K.J., Gaidatzis, D., Punga, T., and Buhler, M. (2011). Dicer associates with chromatin to repress genome activity in Schizosaccharomyces pombe. Nature structural & molecular biology *18*, 94-99.

Wright, J.E., Gaidatzis, D., Senften, M., Farley, B.M., Westhof, E., Ryder, S.P., and Ciosk, R. (2011). A quantitative RNA code for mRNA target selection by the germline fate determinant GLD-1. The EMBO journal *30*, 533-545.

Wulczyn, F.G., Cuevas, E., Franzoni, E., Rybak, A. (2010) MiRNA need a TRIM regulation of miRNA activity by Trim-NHL proteins. Advances in experimental medicine and biology *700*, 85-105.

Xu, T., Venable, J.D., Park, S.K., Cociorva, D., Lu, B., Liao, L., Wohlschlegel, J., Hewel, J., Yates, J.R. 3rd (2006) ProLuCID, a fast and sensitive tandem mass spectra-based protein identification program. Molecular and cellular proteomics *5*, S174.

Yang, H., Zhang, Y., Vallandingham, J., Li, H., Florens, L., and Mak, H.Y. (2012). The RDE-10/RDE-11 complex triggers RNAi-induced mRNA degradation by association with target mRNA in C. elegans. Genes & development *26*, 846-856.

Yang, S., Tutton, S., Pierce, E., and Yoon, K. (2001). Specific double-stranded RNA interference in undifferentiated mouse embryonic stem cells. Molecular and cellular biology *21*, 7807-7816.

Yates, J.R., Ruse, C.I., and Nakorchevsky, A. (2009). Proteomics by mass spectrometry: approaches, advances, and applications. Annual review of biomedical engineering *11*, 49-79.

Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.C., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell *127*, 747-757.

Youngman, E.M., and Claycomb, J.M. (2014). From early lessons to new frontiers: the worm as a treasure trove of small RNA biology. Frontiers in genetics *5*, 416.

Yuan, Y., Li, D.M., and Sun, H. (1998). PIR1, a novel phosphatase that exhibits high affinity to RNA . ribonucleoprotein complexes. The Journal of biological chemistry *273*, 20347-20353.

Zambon, R.A., Vakharia, V.N., and Wu, L.P. (2006). RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster. Cellular microbiology *8*, 880-889.

Zamudio, J.R., Kelly, T.J., and Sharp, P.A. (2014). Argonaute-bound small RNAs from promoter-proximal RNA polymerase II. Cell *156*, 920-934.

Zhang, C., Montgomery, T.A., Fischer, S.E., Garcia, S.M., Riedel, C.G., Fahlgren, N., Sullivan, C.M., Carrington, J.C., and Ruvkun, G. (2012). The Caenorhabditis elegans RDE-10/RDE-11 complex regulates RNAi by promoting secondary siRNA amplification. Current biology *22*, 881-890.

Zhang, C., Montgomery, T.A., Gabel, H.W., Fischer, S.E., Phillips, C.M., Fahlgren, N., Sullivan, C.M., Carrington, J.C., and Ruvkun, G. (2011). mut-16 and other mutator class genes modulate 22G and 26G siRNA pathways in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America *108*, 1201-1208.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. The EMBO journal *21*, 5875-5885.

Zhang, L., Ding, L., Cheung, T.H., Dong, M.Q., Chen, J., Sewell, A.K., Liu, X., Yates, J.R., 3rd, and Han, M. (2007). Systematic identification of C. elegans miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. Molecular cell *28*, 598-613.

Zhang, Z.Y. (2003). Mechanistic studies on protein tyrosine phosphatases. Progress in nucleic acid research and molecular biology *73*, 171-220.

Zhuang, J.J., and Hunter, C.P. (2012). The Influence of Competition Among Small RNA Pathways on Development. Genes 3.

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. Nature *402*, 835-838.

# APPENDIX A

## List of *C. elegans* Strains Used in This Study

| short name | genotype | remarks | official CGC name[a] |
|---|---|---|---|
| ***alg-3; alg-4; gfp::alg-3*** | *alg-4(ok1041)* III*; alg-3(tm1155)* IV*; In[gfp::alg-3]* | | |
| ***avr-14; glc-1*** | *avr-14(ad1305)* I*; glc-1(pk54::Tc1)* V | | DA1384 |
| ***avr-15*** | *avr-15(ad1051)* V | | DA1051 |
| **avr3x DA1316** | *avr-14(ad1305)* I*; glc-1(pk54::Tc1) avr-15(ad1051)* V | | DA1316 |
| **avr3x outcrossed** | *avr-14(ad1305)* I*; glc-1(pk54::Tc1) avr-15(ad1051)* V | D.C.; outcrossed 6x against N2 | |
| ***dcr-1(mg375)*** | *dcr-1(mg375)* | temperature sensitive sterile and Eri allele of Dicer | YY470 |
| ***dcr-1; avr3x*** | *avr-14(ad1302)* I*; dcr-1(ok247)/qC1*[neIs ($P_{myo-2}$::avr-15, rol-6(su1006), unc-22(RNAi))]* III*; avr-15(ad1051) glc-1(pk54::Tc1)* V | | |
| ***drh-1; avr3x*** | *avr-14(ad1302)* I*; drh-1(tm1328)* IV; *avr-15(ad1051) glc-1(pk54::Tc1)* V | D.C. | |
| ***drh-1; pir-1; avr3x*** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1*(In[$P_{sur-5}$::gpf, $P_{myo-2}$::avr-15])* II*; drh-1(tm1328)* IV; *avr-15(ad1051) glc-1(pk54::Tc1)* V | D.C. | |
| ***eat-2*** | *eat-2(ad1113)* II | | DA1113 |
| ***eft-3; avr3x*** | *avr-14 (ad1302)* I*; eft-3(q145)/qC1*[neIs($P_{myo-2}$::avr-15, rol-6(su1006), unc-22(RNAi))]* III*; avr-15(ad1051) glc-1(pk54::Tc1)* V | D.C. | |
| ***eri-1*** | *eri-1(mg366)* IV | | GR1373 |
| **N2** | wild type | N2 Bristol; wild-type reference strain | |
| ***pir-1(tm1496)*** | *pir-1(tm1496) unc-4(e120)/ mnC1[dpy-10(e128), unc-52(e444)]* II | *unc-4* and *pir-1* are genetically linked; non-GFP balancer | WM170 |
| ***pir-1(tm1496);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm1496)/mnC1* [In[$P_{sur-5}$::gfp, $P_{myo-2}$::avr-15] dpy-10(e128) unc-52(e444)]* II*; avr-15(ad1051) glc-1(pk54::Tc1)* V | D.C. | |
| ***pir-1(tm3198)*** | *pir-1(tm3198)/mnC1** II | D.C. | |
| ***pir-1(tm3198)*** | *pir-1(tm3198)/mnC1* II | D.C. | |
| ***pir-1; pir-1(mut ATG1)::gfp*** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(M1A)::gfp]* | D.C.; mutated positions based on 261 aa protein | |
| ***pir-1; pir-1(mut ATG2)::gfp*** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1** II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(M29L)::gfp]* | D.C.; mutated positions based on 261 aa protein | |
| ***pir-1; pir-1(mut ATG3)::gfp*** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(M89L)::gfp]* | D.C.; mutated positions based on 261 aa protein | |
| ***pir-1; pir-1::gfp; drh-3*** | *avr-14(ad1302) drh-3(tm1217)/hT2[bli-4(e937) let-?(q782) qIs48($P_{myo-2}$::gfp; $P_{pes-10}$::gfp; $P_{ges-1}$::gfp)]* I*; pir-1(tm3198)/mnc1*(In[$P_{sur-5}$::gpf, $P_{myo-2}$::avr-15])* II*;* +/hT2 III*;; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; In[pir-1::gfp, Cb-unc-119(+)]* | D.C.; from bombardment line CG8E | |
| ***pir-1; pir-1::gfp; dcr-1*** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; dcr-1(ok247)/ qC1*[neIs($P_{myo-2}$::avr-15, rol-6(su1006), unc-22(RNAi))] unc-119(ed3)* III*; In [Cb-unc-119(+); pir-1::gfp]* IV*; avr-15(ad1051) glc-1 (pk54::Tc1)* V | D.C.; transgene integrated at *cxTi10882* locus (chr. IV) | |
| ***pir-1; pir-1::gfp; rde-4*** | *pir-1(tm3198) rde-4(ne337)* II*; unc-119(ed3)* III*; eri-1(mg366)* IV*; In[pir-1::gfp, Cb-unc-119(+)]* | D.C.; from bombardment line CG8E | |

| | | |
|---|---|---|
| ***pir-1;* avr3x** | *avr-14(ad1305)* I*; pir-1(tm3198)/mnC1* II; *avr-15(ad1051), glc-1(pk54::Tc1)* V | D.C. |
| ***pir-1; cep-1*** | *cep-1(gk138)* I*; pir-1(tm3198)/mnc1** II | D.C. |
| ***pir-1; cxTi10882;* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1** II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V | D.C. |
| ***pir-1; dcr-1; avr3x*** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1** II*; dcr-1(ok247)/ qC1*[neIs($P_{myo-2}$::avr-15, rol-6(su1006), unc-22(RNAi))]* III*; avr-15(ad1051) glc-1(pk54::Tc1)* V | D.C. |
| ***pir-1; gfp::alg-3*** | *pir-1(tm3198)/mnc1** II*; In[gfp::alg-3]* | D.C. |
| ***pir-1; gld-1*** | *gld-1(ar202)/hT2[bli-4(e937) let-?(q782) qIs48]* I; *pir-1(tm3198)/mnc1** II; *+/hT2 III* | D.C. |
| ***pir-1; glp-1*** | *pir-1(tm3198)/mnc1** II; *glp-1(ar202)* III | D.C.; *glp-1* temperature sensitive allele |
| ***pir-1; pgl-1::RFP*** | *pir-1(tm3198) )/mnC1** II; *In[pgl-1::rfp, Cb-unc-119]* | |
| ***pir-1; pir-1(R28A);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1** II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex[$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(R28A)::gfp]* | D.C.; non-integrated transgenic line |
| ***pir-1; pir-1(K76R);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(K76R)::gfp]* | D.C.; non-integrated transgenic line |
| ***pir-1; pir-1(K118A);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(K118A)::gfp]* | D.C.; non-integrated transgenic line |
| ***pir-1; pir-1(K118R);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; In [Cb-unc-119(+), pir-1(K118R)::gfp]* IV*; avr-15(ad1051) glc-1 (pk54::Tc1)* V | D.C.; transgene integrated at *cxTi10882* locus (chr. IV) |
| ***pir-1; pir-1(C150S);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1** II*; unc-119(ed3)* III*; In[Cb-unc-119(+); pir-1(C150S)::gfp]* IV*; avr-15(ad1051) glc-1 (pk54::Tc1)* V | D.C.; transgene integrated at *cxTi10882* locus (chr. IV) |
| ***pir-1; pir-1(T151S);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; In [Cb-unc-119(+), pir-1(T151S)::gfp]* IV*; avr-15(ad1051) glc-1 (pk54::Tc1)* V | D.C.; transgene integrated at *cxTi10882* locus (chr. IV) |
| ***pir-1; pir-1(T157A);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1(pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(T157A)::gfp]* | D.C.; non-integrated transgenic line |
| ***pir-1; pir-1(R184A);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1** II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex[$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(R184A)::gfp]* | D.C.; non-integrated transgenic line |
| ***pir-1; pir-1(T151S);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; In [Cb-unc-119(+), pir-1(T151S)::gfp]* IV*; avr-15(ad1051) glc-1 (pk54::Tc1)* V | D.C.; transgene integrated at *cxTi10882* locus (chr. IV) |
| ***pir-1; pir-1(T157A);* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(T157A)::gfp]* | D.C.; non-integrated transgenic line |
| ***pir-1; pir-1(K194A; K195A); avr3x*** | *avr-14(ad1302)* I*; pir-1(tm3198)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex[$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(K194A; K195A)::gfp]* | D.C.; non-integrated transgenic line |

| | | | |
|---|---|---|---|
| ***pir-1; pir-1(K194R; K195R); avr3x*** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(K194R; K195R)::gfp]* | D.C.; non-integrated transgenic line | |
| ***pir-1; pir-1(Δ194-208); avr3x*** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1** II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex[$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(Δ194-208)::gfp]* | D.C.; non-integrated transgenic line | |
| ***pir-1; pir-1(K204A; K205A); avr3x*** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(K204A; K205A)::gfp]* | D.C.; non-integrated transgenic line | |
| ***pir-1; pir-1(Δ204-208); avr3x*** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; cxTi10882* IV*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V*; Ex [$P_{myo-2}$::mCherry, $P_{myo3}$::mCherry, $P_{glh-2}$::transposase, $P_{hsp-16.48}$::transposase, Cb-unc-119(+), pir-1(Δ204-208)::gfp]* | D.C.; non-integrated transgenic line | |
| ***pir-1; pir-1::3xflag*** | *pir-1(tm3198)* II*; unc-119(ed3)* III*; In[pir-1::3xflag, Cb-unc-119(+)]* | D.C.; from bombardment line CF1C | |
| ***pir-1; pir-1::gfp*** | *pir-1(tm3198)* II*; unc-119(ed3)* III*; In[pir-1::gfp, Cb-unc-119(+)]* | D.C.; from bombardment line CG8E | |
| ***pir-1; pir-1::gfp*** | *pir-1(tm3198)* II*; unc-119(ed3)* III*; In [Cb-unc-119(+); pir-1::gfp]* IV | transgene integrated at *cxTi10882* locus (chr. IV) | |
| ***pir-1; pir-1::gfp;* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)* II*; unc-119(ed3)* III*; In [Cb-unc-119(+); pir-1::gfp]* IV*; avr-15(ad1051) glc-1 (pk54::Tc1)* V | D.C.; transgene integrated at *cxTi10882* locus (chr. IV) | |
| ***pir-1; pir-1::gfp; dcr-1; Ex[3xFlag::dcr-1]*** | *pir-1(tm3198)* II*; dcr-1(ok247) unc-119(ed3)* III*; In[pir-1::gfp, Cb-unc-119(+)]; Ex[10xhis::tev::3xflag::dcr-1, $P_{dpy-30}$::mCherry]* | D.C.; from bombardment line CG8E | |
| ***pir-1; pir-1::gfp; eri-1*** | *pir-1(tm3198)* II*; unc-119(ed3)* III*; eri-1(mg366)* IV*; In[pir-1::gfp, Cb-unc-119(+)]* | D.C.; from bombardment line CG8E | |
| ***pir-1; pir-1::gpf; pgl-1::RFP*** | *pir-1(tm3198)* II*; unc-119(ed3)* III*; In[pir-1::gfp, Cb-unc-119(+)]; In[pgl-1::rfp, C.b unc-119]* | D.C.; from bombardment line CG8E | |
| ***pir-1; seam cell gfp*** | *pir-1(tm3198)/mnc1** II; *In[$P_{seam\ cell}$::gfp]* | D.C. | |
| ***pir-1; unc-119*** | *pir-1(tm3198)/mnc1** II*; unc-119(ed3)* III | D.C. | |
| ***pir-1; unc-119;* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1** II*; unc-119(ed3)* III*; avr-15 (ad1051) glc-1 (pk54::Tc1)* V | D.C. | |
| ***rde-1*** | *rde-1(ne300)* V | | |
| ***rde-1; avr3x*** | *avr-14(ad1302)* I*; avr-15(ad1051) glc-1(pk54::Tc1) rde-1(ne300)* V | D.C. | |
| ***rde-1; pir-1; avr3x*** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1** II; *avr-15(ad1051) glc-1(pk54::Tc1) rde-1(ne300)* V | D.C. | |
| ***rrf-1*** | *rrf-1(pk1417)* I | | PD8488 |
| ***rrf-1; avr3x*** | *avr-14(ad1302) rrf-1(pk1417)* I*; avr-15(ad1051) glc-1(pk54::Tc1)* V | D.C. | |
| ***rrf-1; pir-1; avr3x*** | *avr-14(ad1302) rrf-1(pk1417)* I*; pir-1(tm3198)/mnc1** II; *avr-15(ad1051) glc-1(pk54::Tc1)* V | D.C. | |
| ***rrf-3*** | *rrf-3(pk1426)* II | | NL2099 |
| ***sec-5*** | *sec-5(tm1443)/mnC1** II | D.C. | |
| ***ergo-1*** | *ergo-1(tm1860)* V | | WM158 |
| ***rde-1; unc-119;* avr3x** | *avr-14(ad1302)* I*; unc-119(ed3)* III; *avr-15(ad1051) glc-1(pk54::Tc1) rde-1(ne300)* V | D.C. | |
| ***rde-1; pir-1; unc-119;* avr3x** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1*; unc-119(ed3)* III; *avr-15(ad1051) glc-1(pk54::Tc1) rde-1(ne300)* V | D.C. | |
| ***rde-1; Ex[gfp::rde-1]; avr3x Line 4*** | *avr-14(ad1302)* I*; unc-119(ed3)* III; *avr-15(ad1051) glc-1(pk54::Tc1) rde-1(ne300)* V; *Ex[Cb-119(+), gfp::rde-1]* | D.C.; non-integrated transgenic line | |

| | | | |
|---|---|---|---|
| ***pir-1; rde-1; Ex[gfp::rde-1]; avr3x Line 9*** | *avr-14(ad1302)* I*; pir-1(tm3198)/mnc1*(In[$P_{sur-5}$::gpf, $P_{myo-2}$::avr-15]); unc-119(ed3)* III; *avr-15(ad1051) glc-1(pk54::Tc1) rde-1(ne300)* V; *Ex[Cb-119(+), gfp::rde-1]* | D.C.; non-integrated transgenic line | |
| ***alg-1*** | *alg-1(tm1492)* X | | |
| ***alg-2*** | *alg-2(ok304)* II | | WM53 |
| ***pir-1(tm1496); rrf-1*** | *pir-1(tm1496)/mnc1*(In[$P_{sur-5}$::gpf, $P_{myo-2}$::avr-15])* II*; rrf-1(pk1417)* I | D.C. | WM97 |
| ***nst-1::gfp*** | *nst-1(vr6) I; unc-119(ed3) III; In[Cb-unc-119(+), nst-1::gfp]* | bombardment line | |
| ***pir-1; pir-1::3xflag; nst-1::gfp*** | *pir-1(tm3198)* II*; unc-119(ed3)* III*; In[pir-1::3xflag, Cb-unc-119(+)]; In[Cb-unc-119(+), nst-1::gfp]* | D.C.; from bombardment line CF1C | |
| ***F54C8.4*** | *+/mT1* II; *F54C8.4(tm2568)/mT1[dpy-10 (e128)]* III | D.C. | |
| ***rde-4*** | *rde-4(ne337)* III | | |
| ***dcr-1; dcr-1::HA::flag*** | *dcr-1(ok247)* III*; Ex[10xhis::tev::3xflag::dcr-1, $P_{dpy-30}$::mCherry]* | | BB92 |

[a] The CGC – Caenorhabditis Genetics Center, University of Minnesota – is the official repository for important *C. elegans* strains. Each name specifies the laboratory where the strain was created.

'*Ex*' denotes non-integrated extrachromosomal DNA that is stably transmitted at each generation (the rate of transmission varies with each particular line).

'*In*' refers to sequences integrated in the genome, regardless of the transgenesis method utilized.

The *mnC1** genetic balancer is modified from the original balancer *mnC1*, which contains *dpy-10(e128)* and *unc-52(e444*) recessive markers. *mnC1** contains an integrated transgene with $P_{sur-5}$*::gpf* and $P_{myo-2}$*::avr-15*. These markers were not displayed in front of every balancer for simplification. For an extensive list of genetic balancers and their characteristics refer to WormBook.org.

All strains marked with 'D.C' were made by Daniel Chaves.

# APPENDIX B

## List of Oligonucleotides Used in This Study

| oligo | targeted gene/sequence | sequence (5'-3') | purpose/remarks |
|---|---|---|---|
| **O1** | *pir-1(tm3198)* | CGCCCACGCGGATATGAAAG | genotyping |
| **O2** | *pir-1(tm3198)* | CCCGCTCCATTGGATGTCCA | genotyping |
| **O3** | *18S* rRNA | TTGCGTACGGCTCATTAGAGCA | qRT-PCR |
| **O4** | *18S* rRNA | GCCTTGCGTTGGGGTATAGTTG | qRT-PCR |
| **O5** | *gpd-2* mRNA | GGAGCCAAGAAGGTCATCATCTCT | qRT-PCR |
| **O6** | *gpd-2* mRNA | TAGTGGTGCAGGAAGCATTGGA | qRT-PCR |
| **O7** | *sec-5* mRNA | GATTGGAGGATATTAATCAAATGC | qRT-PCR |
| **O8** | *sec-5* mRNA | GCAAATTGATCTCCGTACTGC | qRT-PCR |
| **O9** | *pir-1* mRNA | GGAAAGAGACTCCCAGACAGATGG | qRT-PCR |
| **O10** | *pir-1* mRNA | CAGGCATATTTTTCCCATCGAAGA | qRT-PCR |
| **O11** | *unc-22* mRNA | GTCTTATCTCCATTTGTCAATTCTTGTCC | qRT-PCR |
| **O12** | *unc-22* mRNA | TGTAACAAAGCACTTGGGAGG A | qRT-PCR |
| **O13** | *pir-1* | TTTTAATACGACTCACTATAGGTCGTTGCACAG CAATCGTTA | with T7 promoter for *in vitro* dsRNA synthesis |
| **O14** | *pir-1* | TTTTAATACGACTCACTATAGGGAACAAATCGA GTCCCATCG | with T7 promoter for *in vitro* dsRNA synthesis |
| **O15** | *sec-5* | TTTTAATACGACTCACTATAGGTTGAAAATCAG CAGAGAATCAAT | with T7 promoter for *in vitro* dsRNA synthesis |
| **O16** | *sec-5* | TTTTAATACGACTCACTATAGGTCAAGCAGTGA AGAAATCAAAGA | with T7 promoter for *in vitro* dsRNA synthesis |
| **O17** | *pir-1* | TTTTAATACGACTCACTATAGGTCGTTGCACAG CAATCGTTA | with T7 promoter for dsRNA feeding vector |
| **O18** | *pir-1* | TTTTAATACGACTCACTATAGGAAATTGTGGAA TCGGCTGAG | with T7 promoter for dsRNA feeding vector |
| **O19** | *gfp* mRNA | GGTGATGTTAATGGGCACAA | qRT-PCR |
| **O20** | *gfp* mRNA | AGTAGTGACAAGTGTTGGCCATG | qRT-PCR |
| **O21** | *pir-1* | TTTTTGGATCCTTTGTCGCTATTCGTAC TTCACC | N-terminal tagging of *pir-1* with *Bam*HI site |
| **O22** | *pir-1* | TTTTTGGATCCGAGAAATCAATGTTT TTACACCAAATTG | N-terminal tagging of *pir-1* with *Bam*HI site |
| **O23** | *pir-1* | TTTTTGATATCATTTAGAAAAATCGAAAGTTTAT AGGAAAC | C-terminal tagging of *pir-1* with *Xba*I site |
| **O24** | *pir-1* | TTTTTGATATCTATTCTCACCATCCAC AAAACTTCC | C-terminal tagging of *pir-1* with *Xba*I site |
| **O25** | *pir-1* | GCGGCCGCGCCGGAGCCTCGTTGCACAGC | N-terminal tagging of *pir-1* with *Not*I site in frame with *gfp* cassette |
| **O26** | *pir-1* | CGCGGCCGCCATTCCGCTTAGCGCCTTGT | N-terminal tagging of *pir-1* with *Not*I site in frame with *gfp* cassette |
| **O27** | *pir-1* | GGCGGCCGCGCCGGAGCCTCGTTGCACAGC | N-terminal tagging of *pir-1* with *Not*I site in frame with *3xflag* cassette |
| **O28** | *pir-1* | CGCGGCCGCCCATTCCGCTTAGCGCCTTGT | N-terminal tagging of *pir-1* with *Not*I site in frame with *3xflag* cassette |
| **O29** | *pir-1* | GCGGCCGCGTAACACCCAATACTGTTGTT | C-terminal tagging of *pir-1* with *Not*I site in frame with *gfp* cassette |
| **O30** | *pir-1* | CGCGGCCGCTTGAGAATTATTTCGATGCA | C-terminal tagging of *pir-1* with *Not*I site in frame with *gfp* cassette |
| **O31** | *pir-1* | GGCGGCCGCGTAACACCCAATACTGTTGTT | C-terminal tagging of *pir-1* with *Not*I site in frame with *3xflag* cassette |
| **O32** | *pir-1* | CGCGGCCGCCTTGAGAATTATTTCGATGCA | C-terminal tagging of *pir-1* with *Not*I site in frame with *3xflag* cassette |
| **O33** | *pir-1* | TTTTTCCGCGGCTATTGATCAGCATTTTCTCCT TTTTCTAACG | TOPO cloning of tagged *pir-1* sequences |

| | | | |
|---|---|---|---|
| **O34** | *pir-1* | TTTTTCCGCGGCACAATAACTTTATGTCTGTTG GCAAAAAG | TOPO cloning of tagged *pir-1* sequences |
| **O35** | *pir-1* | GGGGACAAGTTTGTACAAAAAAGCAGGCTCTA TTGATCAGCATTTTCTCCTTTTTCTAACG | Cloning of tagged *pir-1* sequences into pDONR-201 gateway vector. Contains *att*B1 site. |
| **O36** | *pir-1* | GGGGACCACTTTGTACAAGAAAGCTGGGTCAC AATAACTTTATGTCTGTTGGCAAAAAG | Cloning of tagged *pir-1* sequences into pDONR-201 gateway vector. Contains *att*B2 site. |
| **O37** | *cxTi10882* | AGCTGCTTGAAGAACCCTGA | genotyping |
| **O38** | *cxTi10882* | TTGCATCCCGTTAGAAACAA | genotyping |
| **O39** | *pir-1* R28A mutant | TCCAGGAAAGAGACTCCCAGACGCGTTTTTTTA CAGTTAATTCC | site-directed mutagenesis |
| **O40** | *pir-1* R28A mutant | GGAATTAACTGTAAAAAAACGCGTCTGGGAGT CTCTTTCCTGGA | site-directed mutagenesis |
| **O41** | *pir-1* K76R mutant | GACTCGATTTGTTCCATTCAGAACACCACTAGA TTCATCAT | site-directed mutagenesis |
| **O42** | *pir-1* K76R mutant | ATGATGAATCTAGTGGTGTTCTGAATGGAACAA ATCGAGTC | site-directed mutagenesis |
| **O43** | *pir-1* K118A mutant | GAAAAATCGAAAGTTTATAGGGCACTCAATTGC C CCGGTCACGA | site-directed mutagenesis |
| **O44** | *pir-1* K118A mutant | TCGTGACCGGGGCAATTGAGTGCCCTATAA ACTTTCGATTTTTC | site-directed mutagenesis |
| **O45** | *pir-1* K118R mutant | AAATCGAAAGTTTATAGGAGACTCAATTGCCCC GGTCACGAG | site-directed mutagenesis |
| **O46** | *pir-1* K118R mutant | CTCGTGACCGGGGCAATTGAGTCTCCTATAAA CTTTCGATT T | site-directed mutagenesis |
| **O47** | *pir-1* C150S mutant | GGAAAACTGATTGGCGTACACAGCACTCATGG GCTCAACAGAAC | site-directed mutagenesis |
| **O48** | *pir-1* C150S mutant | GTTCTGTTGAGCCCATGAGTGCTGTGTACGCC AATCAGTTTTCC | site-directed mutagenesis |
| **O49** | *pir-1* T151S mutant | ACTGATTGGCGTACACTGCTCTCATGGGCTCA ACAGAA | site-directed mutagenesis |
| **O50** | *pir-1* T151S mutant | TTCTGTTGAGCCCATGAGAGCAGTGTACGCCA ATCAGT | site-directed mutagenesis |
| **O51** | *pir-1* T157A mutant | ACTCATGGGCTCAACAGAGCAGGATACCTT ATTTGC | site-directed mutagenesis |
| **O52** | *pir-1* T157A mutant | GCAAATAAGGTATCCTGCTCTGTTGAGCCC ATGAGT | site-directed mutagenesis |
| **O53** | *pir-1* R184A mutant | TCAGTGTTTGAATACTACGCTGGACATCCAATG GAGCGGG | site-directed mutagenesis |
| **O54** | *pir-1* R184A mutant | CCCGCTCCATTGGATGTCCAGCGTAGTATTCA AACACTGA | site-directed mutagenesis |
| **O55** | *pir*-1 R190K mutant | ACCGTGGACATCCAATGGAGAAGGAACACTAC AAAAAATCATTG | site-directed mutagenesis |
| **O56** | *pir*-1 R190K mutant | CAATGATTTTTTGTAGTGTTCCTTCTCCATTGGA TGTCCACGGT | site-directed mutagenesis |
| **O57** | *pir*-1 K194A; K195A mutant | ATCCAATGGAGCGGGAACACTACGCTGCATCA TTGTACGAAGCTGAAAGGAAG | site-directed mutagenesis |
| **O58** | *pir*-1 K194A; K195A mutant | CTTCCTTTCAGCTTCGTACAATGATGCAGCGTA GTGTTCCCGCTCCATTGGAT | site-directed mutagenesis |
| **O59** | pir-1 K194R; K195R mutant | ATCCAATGGAGCGGGAACACTACCGTCGATCA TTGTACGAAGCTGAAAGGAAG | site-directed mutagenesis |
| **O60** | pir-1 K194R; K195R mutant | CTTCCTTTCAGCTTCGTACAATGATCGACGGTA GTGTTCCCGCTCCATTGGAT | site-directed mutagenesis |
| **O61** | *pir*-1 Δ194-208 mutant | GTAGTGTTCCCGCTCCATTG | mutagenesis |
| **O62** | *pir*-1 Δ194-208 mutant and *pir-1* Δ204-208 mutant | AGCAGCGGAAAAAGCAGC | mutagenesis |
| **O63** | *pir*-1 K204A; K205A mutant | CATTGTACGAAGCTGAAAGGAAGGCTGCTTAC GGAAAAAGCAGCGGAAAAAG | site-directed mutagenesis |
| **O64** | *pir*-1 K204A; K205A mutant | CTTTTTCCGCTGCTTTTTCCGTAAGCAGCCTTC CTTTCAGCTTCGTACAATG | site-directed mutagenesis |
| **O65** | *pir*-1 Δ204-208 mutant | CTTCCTTTCAGCTTCGTACAATG | mutagenesis |
| **O66** | *pir*-1 M1A mutant (ATG1) | CAAGGCGCTAAGCGGAGCGCCGGAGCCT CGTTGC | site-directed mutagenesis (assuming 261 aa protein) |
| **O67** | *pir*-1 M1A mutant (ATG1) | GCAACGAGGCTCCGGCGCTCCGCTTAGC GCCTTG | site-directed mutagenesis (assuming 261 aa protein) |
| **O68** | *pir*-1 M29L mutant (ATG2) | CAATATTTTCAGTCAGTGTCCTGTCTAATTACC ATCATAATCAC | site-directed mutagenesis (assuming 261 aa protein) |
| **O69** | *pir*-1 M29L mutant (ATG2) | GTGATTATGATGGTAATTAGACAGGACACTGAC TGAAAATATTG | site-directed mutagenesis (assuming 261 aa protein) |

| | | | |
|---|---|---|---|
| **O70** | *pir*-1 M89L mutant (ATG3) | CTTCGATGGGAAAAATCTGCCTGTGGAGCTTC | site-directed mutagenesis (assuming 261 aa protein) |
| **O71** | *pir*-1 M89L mutant (ATG3) | GAAGCTCCACAGGCAGATTTTTCCCATCGAAG | site-directed mutagenesis (assuming 261 aa protein) |
| **O72** | SL1 | GGTTTAATTACCCAAGTTTGAG | mapping of SL1 splice site |
| **O73** | SL2 | GGTTTTAACCCAGTTACTCAAG | mapping of SL1 splice site |
| **O74** | *pir-1* mRNA | CGGCCGACGTTATCGTAAAT | mapping of SL1 splice site |
| **O75** | *sec-5* mRNA | TTGATGGGGATTTCCATTTC | mapping of SL1 splice site |
| **O76** | human *PIR1* | TTTTTGCGGCCGCAAGCCAGTGGCATCATCCC | cloning of human *PIR1* cDNA with *Not*I site |
| **O77** | human *PIR1* | TTTTTTCTAGACTGGGTCCATTCCCAACAGG | cloning of human *PIR1* cDNA with *Xba*I site |
| **O78** | *act-3* mRNA | GGCCCAATCCAAGAGAGGTATCC | qRT-PCR |
| **O79** | *act-3* mRNA | GGGCAACACGAAGCTCATTGTA | qRT-PCR |
| **O80** | *pos-1* mRNA | rArUrUrCrCrCrGrUrUrGrGrUrCrCrArU rUrCrArGrGrA | sense control for sRNA northern blotting |
| **O81** | *pos-1* mRNA | rUrCrCrUrGrArArUrGrGrArCrCrArArC rGrGrGrArArU | anti-sense control for sRNA northern blotting |
| **O82** | *pos-1* mRNA | CGGCTTCCAATGAACCCTCGTGGGAG | qRT-PCR |
| **O83** | *pos-1* mRNA | CTA GCT TCG CGG CAT TCC CAT CGA C | qRT-PCR |
| **O84** | *pos-1* mRNA | TCGGAGATTATTCCCGTTGGTCCATTCAGGAA GCCAAGAC/Starfire/ | starfire oligo for northern probe |
| **O85** | *unc-22* mRNA | rGrGrUrArUrCrGrUrArUrCrCrArGrUrG rArUrUrUrCrU | anti-sense control for sRNA northern blotting |
| **O86** | *unc-22* mRNA | AGAAATCACTGGATACGATACC/Starfire/ | starfire oligo for northern probe |
| **O87** | miR-66 | TCACATCCCTAATCAGTGTCATG/Starfire/ | starfire oligo for northern probe |
| **O88** | SL1 precursor | CTCAAACTTGGGTAATTAAACC/Starfire/ | starfire oligo for northern probe |
| **O89** | let-7 miRNA | AACTATACAACCTACTACCTCACC GGATCC/Starfire/ | starfire oligo for northern probe |
| **O90** | *F37D6.3* | AAGATCCCAATTTCGAATACAATCGGTAT CCAGTC/Starfire/ | starfire oligo for northern probe |
| **O91** | *Y47H10A.5* | GGAATCTGTCGCAGAGTTCCGTGACGTCG TCCGTC/Starfire/ | starfire oligo for northern probe |
| **O92** | 26G-RNA 1 | CGGAATCTCAAACTTTTCCATCTTGC/Starfire/ | starfire oligo for northern probe |
| **O93** | 26G-RNA 263 | TAGCATATGCATGCACCATAAACAAC/Starfire/ | starfire oligo for northern probe |
| **O94** | 21U-RNA 1 | GCACGGTTAACGTACGTACCA/Starfire/ | starfire oligo for northern probe |
| **O95** | *T01C8.5* | CAACCCAACATGGGGAAATC | qRT-PCR |
| **O96** | *T01C8.5* | GACAGACTTCTCTGGGGCTGA | qRT-PCR |
| **O97** | *Y47H10A.5* | CGTTTATCAGGTTGAAACACTCG | qRT-PCR |
| **O98** | *Y47H10A.5* | TCATCGGATCTGGTGAAAACG | qRT-PCR |
| **O99** | *E01G4.5* | CGGCAATTTATTCTAGAGCACAC | qRT-PCR |
| **O100** | *E01G4.5* | GTCAAACAACAGCTTTCCAACG | qRT-PCR |
| **O101** | *F39E9.7* | CCCAGTGGCCCAATTAAACG | qRT-PCR |
| **O102** | *F39E9.7* | CCCACGGCTTGTTCTTTGACA | qRT-PCR |
| **O103** | *Y43F8B.9* | CCCAGTGGCCCAATTAAACG | qRT-PCR |
| **O104** | *Y43F8B.9* | CCCACGGCTTGTTCTTTGACA | qRT-PCR |
| **O105** | *Y57G11C.51* | AATGATGCAGCGAAGGAGGTCG | qRT-PCR |
| **O106** | *Y57G11C.51* | AGGATGATGGCGTGGAAG | qRT-PCR |
| **O107** | *klp-16* mRNA | CGTCGAGGAGTTCATTCGGTCG | qRT-PCR |
| **O108** | *klp-16* mRNA | CGGGATGATTCCTTCCTCTT | qRT-PCR |
| **O109** | *bub-1* mRNA | AAAGCTGCATTTGGACCAAAAACC | qRT-PCR |
| **O110** | *bub-1* mRNA | CGGGATGATTCCTTCCTCTT | qRT-PCR |
| **O111** | *vig-1* mRNA | TCCATCCGACGACGACGATGAG | qRT-PCR |
| **O112** | *vig-1* mRNA | GGGCTTCACAGACTTTTCCTC | qRT-PCR |
| **O113** | *hcp-3* mRNA | GTCTATGGATTTCACGGTCGGC | qRT-PCR |
| **O114** | *hcp-3* mRNA | TTCTTCGTCGGAGCTATCGT | qRT-PCR |

| | | | |
|---|---|---|---|
| **O115** | cloning/deep-sequencing | AppCTGTAGGCACCATCAAT/ddC/ | 3'-end linker |
| **O116** | cloning/deep-sequencing | TCTACrArGrUrCrCrGrArCrGrArUrC-barcode | 5' linker (DNA/RNA hybrid); barcode is rNrNrNrN using diferent combinations of rA, rU, rC and rG. |
| **O117** | cloning/deep-sequencing | ATTGATGGTGCCTACAG | oligo for reverse transcription |
| **O118** | cloning/deep-sequencing | TCTACAGTCCGACGATCGGG | short 5' PCR primer |
| **O119** | cloning/deep-sequencing | GTTCTACAGTCCGACGATC | short 3' PCR primer |
| **O120** | cloning/deep-sequencing | AATGATACGGCGACCACCGACAGGTTCAGAGT TCTACAGTCCGACGATC | long 5' PCR primer (Illumina P5 sequence underlined) |
| **O121** | cloning/deep-sequencing | CAAGCAGAAGACGGCATACGAATTGATGGTGC CTACAG | long 3' PCR primer (Illumina P7 sequence underlined) |
| **O122** | M13 Forward | GTAAAACGACGGCCAG | PCR of inserts cloned into TOPO-TA vector |
| **O123** | M13 Reverse | CAGGAAACAGCTATGAC | PCR of inserts cloned into TOPO-TA vector |
| **O124** | T7 promoter | TAATACGACTCACTATAGGG | Sanger sequencing of inserts cloned into TOPO-TA vector |
| **O125** | pppRNA for dephosphorylation assay | TAATACGACTCACTATAGGATCCTTGAAATGGA ACATCTGAAT | sense oligo for T7 synthesis of RNA substrate for dephosphorylation assay |
| **O126** | pppRNA for dephosphorylation assay | ATTCAGATGTTCCATTTCAAGGATCCTATAGTG AGTCGTATTA | antisense oligo for T7 synthesis of RNA substrate for dephosphorylation assay |
| **O127** | Orsay virus RNA1 | ACCTCACAACTGCCATCTACA | qRT-PCR GW194 (Felix *et al*., 2011) |
| **O128** | Orsay virus RNA1 | GACGCTTCCAAGATTGGTATTGGT | qRT-PCR GW195 (Felix *et al*., 2011) |

# APPENDIX C

## List of Additional Potential Interactors with RNA-Related Functions Obtained through MudPIT

| protein | PIR-1::GFP IP | PIR-1::GFP IP (light) | GFP IP N2 (heavy) | PIR-1::3xFlag IP (light) | Flag IP N2 (heavy) | PIR-1::GFP IP in *drh-3* | short protein description adapted from WormBase |
|---|---|---|---|---|---|---|---|
| **small RNA pathway** | | | | | | | |
| **ALG-1** | | • 3 | | | | | alg-1 encodes an Argonaut ortholog; alg-1 is involved in RNA interference and affects developmental timing along with alg-2 and dcr-1 by regulating expression of the lin-4 and let-7 small temporal RNAs. |
| **ALG-2** | | • 4 | | • 3 | | | alg-2 encodes a PAZ and PIWI-domain containing protein that is a member of the highly conserved eukaryotic RDE-1/AGO1/PIWI family of proteins that regulate posttranscriptional gene silencing (PTGS); ALG-2 functions with ALG-1 to control specific developmental timing events by positively regulating expression of small temporal RNAs (stRNAs) encoded by lin-4 and let-7; ALG-2 and ALG-1 are also required for RNAi in the germ line, although they are not required for RNAi in the soma; ALG-2 is expressed in nearly all cells from embryogenesis through adulthood and localizes to the cytoplasm. |
| **CSR-1** | • 4 | | • 11 | | | • 3 | csr-1 encodes, by alternative splicing, two isoforms of an Argonaute protein required for chromosome segregation, embryonic viability, Slicer activity induced by secondary siRNAs, and (partially) for germline RNAi. |
| **NRDE-1** | | | | | | • 3 | nrde-1 encodes a novel protein conserved amongst nematodes; NRDE-1 activity is required, along with that of NRDE-2, NRDE-3, and NRDE-4, for regulation of gene expression via a nuclear RNAi pathway that inhibits RNA polymerase II elongation and deposits histone H3K9 methylation on genomic regions targeted by RNAi; an NRDE-1:GFP fusion protein localizes to the nucleus. |
| **RDE-8 (ZC477.5)** | • 5 | | | | | | rde-8 encodes a Zc3h12a-like endoribonuclease required for RNAi and localizes to mutator foci. |
| **VIG-1** | • 20 | • 15 | • 45 | • 7 | | | vig-1 encodes a predicted RNA-binding protein orthologous to Drosophila VIG (Vasa Intronic Gene). VIG-1 is a component of the 250 kDa RNA-induced silencing complex (RISC) complex and co-immunoprecipates with both TSN-1, the C. elegans Tudor-SN ortholog, and the let-7 miRNA. |
| **P-granule associated** | | | | | | | |
| **CAR-1** | • 22 | • 19 | • 24 | | | • 4 | car-1 encodes a putative RNA-binding protein orthologous to budding yeast Scd6p, fission yeast Sum2p, Drosophila TRAL, and human LSM14A and LSM14B; CAR-1 expressed in the germline; CAR-1 associates with CGH-1, DCAP-1, and CEY-2/3/4, in P granules and other cytoplasmic particles of the early embryo; CAR-1 also localizes to the mitotic spindle of dividing 1-cell embryos, and to ER; CAR-1 requires CGH-1 for normal localization in meiotic germ cells, and binds CGH-1 in an RNA-dependent manner. |
| **CEY-2** | • 18 | • 13 | • 14 | | | | cey-2 encodes a cold-shock/Y-box domain-containing protein; by homology, CEY-2 is predicted to function as either an RNA-binding protein involved in translation or RNA processing, or a DNA-binding protein involved in transcriptional regulation; CEY-2 associates with CGH-1 and CEY-3/4 in cytoplasmic particles of the gonad and early embryo. |
| **CEY-3** | • 7 | • 5 | • 8 | | | | cey-3 encodes a protein with a cold-shock/Y-box domain; CEY-3 associates with CGH-1, CEY-2, and CEY-4 in cytoplasmic particles of the gonad and early embryo. |
| **CEY-4** | | • 8 | • 9 | | | | cey-4 encodes a Y box-containing protein with no known function in vivo; CEY-4 associates with CGH-1 and CEY-2/3 in cytoplasmic particles of the gonad and early embryo. |
| **CGH-1** | • 10 | • 7 | • 6 | | | | cgh-1 encodes a putative DEAD-box RNA helicase, orthologous to budding yeast Dhh1p, fission yeast Ste13p, Drosophila ME31B, and human DDX6; CGH-1 is expected to enable decapping-dependent mRNA degradation. |
| **GLH-1** | • 2 | | | | | • 3 | glh-1 encodes a putative DEAD-box RNA helicase that contains four CCHC zinc fingers and is homologous to Drosophila VASA, a germ-line-specific, ATP-dependent RNA helicase; GLH-1 is a constitutive P granule component and thus, with the exception of mature sperm, is expressed in germ cells at all stages of development. |
| **mRNA processing** | | | | | | | |
| **CPF-1** | | | | • 13 | • 14 | | cpf-1 encodes an ortholog of human cleavage stimulation factor, 3' pre-RNA, subunit 1 (HGNC:CSTF1). |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **HRP-1 (F42A6.7)** | • 2 | | | | | | hrp-1 encodes a putative heterogeneous nuclear ribonucleoprotein (hnRNP); HRP-1 has two N-terminal RRM domains and a low-complexity C-terminal domain; HRP-1's homologs include HNRNPA0 (OMIM:609409), HNRNPA1 (HNRPA1; OMIM:164017; overexpressed in vitamin D resistance); HNRPA3 (OMIM:605372); and HNRNPA1L2. |
| **HRP-2** | | | | | | • 4 | hrp-2 encodes an ortholog of human heterogeneous nuclear ribonucleoprotein R (HGNC:HNRNPR); hrp-2 is predicted to have nucleotide binding activity and nucleic acid binding activity. |
| **PAB-1** | • 15 | • 10 | • 13 | • 2 | | | pab-1 encodes a polyadenylate-binding protein (i.e., poly(A)-binding protein, or PAB). |
| **PAB-2** | • 5 | | | | | | pab-2 encodes a polyadenylate-binding protein 1 homolog with high similarity to human PABP 1. |
| **SQD-1** | • 13 | • 10 | • 8 | | | | sqd-1 encodes an ortholog of human heterogeneous nuclear ribonucleoprotein D-like (HGNC:HNRNPDL); sqd-1 is predicted to have nucleotide binding activity and nucleic acid binding activity. |
| **SUF-1** | | | | • 26 | • 24 | | suf-1 encodes an ortholog of human cleavage stimulation factor, 3' pre-RNA, subunit 3 (HGNC:CSTF3). |
| **UAF-1** | • 2 | | | | | | uaf-1 encodes the large subunit of splicing factor U2AF (U2 Auxiliary Factor), orthologous to mammalian and Drosophila U2AF65; UAF-1 has been shown to bind RNA and this binding is enhanced by UAF-2. |
| **nucleolus-associated** | | | | | | | |
| **DAO-5** | • 2 | | | | | | dao-5 encodes a predicted nucleolar phosphoprotein related to Saccharomyces cerevisiae SRP40 and the vertebrate Nopp140 proteins that may play a role in rRNA gene transcription and nucleolar structural organization. |
| **FIB-1** | • 9 | • 9 | • 6 | | | | fib-1 encodes the C. elegans ortholog of fibrillarin and Saccharomyces cerevisiae Nop1p, an essential component of the U3 SnoRNP; FIB-1 localizes to the nucleolus. |
| **K07H8.10** | | | | | • 3 | • 2 | K07H8.10 encodes an ortholog of human nucleolin (HGNC:NCL); K07H8.10 is predicted to have nucleotide binding activity and nucleic acid binding activity. |
| **LPD-7** | • 3 | | | | | | lpd-7 encodes a BRCT domain-containing protein that is orthologous to Saccharomyces cerevisiae Nop7p and the vertebrate pescadillo proteins required for nucleolar assembly, ribosome biogenesis, and cell proliferation. |
| **NGP-1** | • 4 | | | | | | ngp-1 encodes an ortholog of human guanine nucleotide binding protein-like 2 (nucleolar) (HGNC:GNL2); ngp-1 is predicted to have GTP binding activity. |
| **NOL-1** | • 3 | | | | | | nol-1 encodes an ortholog of human NOP2 nucleolar protein (HGNC:NOP2); nol-1 is predicted to have RNA binding activity and S-adenosylmethionine-dependent methyltransferase activity. |
| **NOL-5** | • 10 | • 13 | • 8 | | | | nol-5 encodes an ortholog of human NOP58 ribonucleoprotein (HGNC:NOP58). |
| **NOLA-3** | • 3 | | | | • 3 | | nola-3 encodes an ortholog of human NOP10 ribonucleoprotein (HGNC:NOP10); nola-3 is predicted to have snoRNA binding activity. |
| **NST-1** | • 8 | | | | | | nst-1 encodes a homolog of human GNL3 (OMIM:608011, nucleostemin) and GNL3L, and of S. cerevisiae NUG1. |
| **T04A8.6** | • 2 | | | | | | T04A8.6 encodes an ortholog of S. cerevisiae NOP15 that may suppress tumorous growth in the germ line by ensuring robust larval germline proliferation. |
| **Y66H1A.4** | • 9 | • 10 | • 7 | | | • 5 | Y66H1A.4 encodes an ortholog of human GAR1 ribonucleoprotein (HGNC:GAR1). Nucleolar. Necessary for ribosome biogenesis and telomere maintenance. |
| **miscelaneous** | | | | | | | |
| **C44E4.4** | • 6 | | • 3 | | | | C44E4.4 encodes an ortholog of human Sjogren syndrome antigen B (autoantigen La) (HGNC:SSB); C44E4.4 is predicted to have nucleotide binding activity and RNA binding activity. The human protein has tRNA processing activities in the nucleus. |
| **CEY-1** | | • 7 | • 8 | | | | cey-1 encodes a protein with a cold-shock/Y-box domain that is expressed in early embryonic blastomeres (at the 15-cell stage, i.e., pre-gastrulation), but is normally repressed in early germline blastomeres by PIE-1. |
| **DDB-1** | • 10 | | | | | • 7 | ddb-1 encodes a protein orthologous to human DNA damage-binding protein 1 (DDB1); ddb-1 is part of a ubiquitin ligase E3 complex, CUL-4/DDB-1, and targets the replication licensing factor, cdt-1, for ubiquitin-mediated proteolysis. |
| **DDX-17 (F58E10.3)** | • 8 | • 4 | • 7 | | | | ddx-17 encodes an ortholog of human DEAD (Asp-Glu-Ala-Asp) box helicase 17 (HGNC:DDX17). |
| **F08B12.4** | | | | | | • 4 | F08B12.4 is localized to the nucleus. |
| **IMA-3** | | | | | | • 2 | ima-3 encodes one of three C. elegans importin alpha nuclear transport factors and the importin alpha that is most similar to the alpha3-subtype; ima-3 is required for normal embryonic, larval, and germline development. |
| **LAF-1** | | • 4 | | | | | laf-1 encodes a DEAD-box RNA helicase; laf-1 is required for embryonic development and sex determination. |
| **MADF-6** | | | | • 4 | | | MADF domain transcription factor. |
| **PLP-1** | • 7 | • 5 | • 4 | | | | plp-1 encodes a protein containing three PUR repeats that has similarity to the mammalian transcription factor pur alpha. |
| **PSF-1** | • 11 | | | | | | psf-1 encodes an ortholog of human GINS complex subunit 1 (Psf1 homolog) (HGNC:GINS1). DNA replication factor. |
| **PUF-12** | • 7 | | | | | | puf-12 encodes an ortholog of human KIAA0020 (HGNC:KIAA0020); puf-12 is predicted to have RNA binding activity. |
| **TCER-1** | • 2 | | | | | | tcer-1 encodes an ortholog of human transcription elongation regulator 1 (HGNC:TCERG1); tcer-1 is localized to the nuclear periphery. |

# APPENDIX D

## List of Two-Fold Down- and Upregulated Genes in *pir-1* Mutant Animals

This list is appended as an electronic Excel spreadsheet due to its large size. The list includes genes grouped by relevant biological categories which were used to build the graphs depicted on Figure 4.10, page 184. As mentioned previously, the data was collected by mRNA-seq of poly(A)-selected RNA from three-day and ten-day old counter-selected *pir-1* mutant animals. Only the genes that were consistently down- or upregulated in both sets were considered in the analysis, increasing our confidence that they are truly misregulated in the absence of PIR-1. We include the fold-misregulation relative to wild-type as well as normalized reads per million for each sample. The tables display functional information about each gene, where available, collected from the WormBase Gene Ontology Browser at WormBase.org. For the downregulated genes, we also added phenotypes obtained by RNAi (where available) to illustrate how silencing some of these genes can recapitulate phenotypes associated with the *pir-1(tm3198)* mutant.

# APPENDIX E

## List of Authored Papers and Respective Contributions

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., **Chaves, D.A.**, Gu, W., Vasale, J.J., Duan, S., et al. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Molecular Cell *31*, 67-78.

*Daniel Chaves performed tiling microarray experiments.*

Gu, W., Shirayama, M., Conte, D.Jr., Vasale, J., Batista, P,J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., Chen, C.C., **Chaves. D.A.**, Duan, S., Kasschau, K.D., Fahlgren, N., Yates, J.R. 3rd, Mitani, S., Carrington, J.C., Mello, C.C. (2009) Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. Molecular Cell *36*, 231-244.

*Daniel Chaves performed tiling microarray experiments.*

Claycomb, J.M., Batista, P.J., Pang, K.M., Gu, W., Vasale, J.J., van Wolfswinkel, J.C., **Chaves, D.A.**, Shirayama, M., Mitani, S., Ketting, R.F., et al. (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell *139*, 123-134.

*Daniel Chaves performed tiling microarray experiments.*

Conine, C.C., Batista, P.J., Gu, W., Claycomb, J.M., **Chaves, D.A.**, Shirayama, M., and Mello, C.C. (2010). Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America *107*, 3588-3593.

*Daniel Chaves performed tiling microarray experiments.*

Gu, W., Lee, H.C., **Chaves, D.**, Youngman, E.M., Pazour, G.J., Conte, D., Jr., and Mello, C.C. (2012). CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. Cell *151*, 1488-1500.

*Daniel Chaves provided samples and aided in optimization of experiments.*

Tsai, H.Y., Chen, C.C., Conte, D., Jr., Moresco, J.J., **Chaves, D.A.**, Mitani, S., Yates, J.R., 3rd, Tsai, M.D., and Mello, C.C. (2015). A Ribonuclease Coordinates siRNA Amplification and mRNA Cleavage during RNAi. Cell *160*, 407-419.

*Daniel Chaves performed Orsay virus infection experiments.*