Mechanisms of Small RNA Biogenesis in *Drosophila*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Benjamin Czech

aus Göppingen

Tübingen

2012

Tag der mündlichen Qualifikation:	30. 01. 2013
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Gregory J. Hannon
2. Berichterstatter:	Prof. Dr. Gerd Jürgens
3. Berichterstatter:	Prof. Dr. Erik J. Sontheimer

Table of contents

Abbreviations

A	adenine
AGO	Argonaute
bp	base pair(s)
C	cytosine
CDS	coding sequence
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
dsRBD	dsRNA-binding domain
endo-siRNA	endogenous small interfering RNA
FKBP	FK506-binding protein
FPKM	fragments per kilo base per million reads
G	guanine
HSP	heat shock protein
kb	kilo base(s)
kDa	kilo Dalton
LINE	long interspersed elements
LTR	long terminal repeats
miR*	microRNA-star
miRNA	microRNA
mRNA	messenger RNA
nt	nucleotide(s)
PCR	polymerase chain reaction
Pre-miRNA	precursor microRNA
Pri-miRNA	primary microRNA
piRNA	Piwi-interacting RNA
PIWI	*P*-element induced wimpy testis
qPCR	quantitative real time polymerase chain reaction
RdRP	RNA-dependent RNA polymerase
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
shRNA	short hairpin RNA
siRNA	small interfering RNA
ssRNA	single-stranded RNA
TAP	tandem affinity purification
TPR	tetratricopeptide repeat
U	uracil
UAS	upstream activation sequence
UTR	untranslated region

List of publications

Accepted manuscripts:

1. Zhou, R.*, **Czech, B.***, Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., Hannon, G.J. (2009). Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA* **15**(10): 1886-1895.
*** joint first authors**

2. **Czech, B.***, Zhou, R.*, Erlich, Y., Brennecke, J., Binari, R., Villalta, C., Gordon, A., Perrimon, N., Hannon, G.J. (2009). Hierarchical rules for Argonaute loading in *Drosophila*. *Molecular Cell* **36**(3): 445-456.
*** joint first authors**

3. **Czech, B.**, Hannon G.J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nature Reviews Genetics* **12**(1): 19-31. **[REVIEW ARTICLE]**

4. Ni, J.-Q.*, Zhou, R.*, **Czech, B.***, Liu, L.-P., Holderbaum, L., Yang-Zhou, D., Shim, H.-S., Tao, R., Handler, D., Karpowicz, P., Binari, R., Booker, M., Brennecke, J., Perkins, L.A., Hannon, G.J., Perrimon, N. (2011). A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nature Methods* **8**(5): 405-407.
*** joint first authors**

5. Preall, J.B.*, **Czech, B.***, Guzzardo, P.M., Muerdter, F., Hannon, G.J. (2012). *shutdown* is a component of the *Drosophila* piRNA biogenesis machinery. *RNA* **18**(8): 1446-1457.
*** joint first authors**

Submitted and prepared manuscripts:

6. **Czech, B.***, Preall, J.B.*, McGinn, J., Knott, S.R., Hannon, G.J. (*). A transcriptome-wide RNAi screen in *Drosophila* ovaries reveals novel factors of the germline piRNA pathway. (*in preparation for submission to Cell*).
*** joint first authors**

Additional publications (not included in thesis):

7. Haase, A.D., Fenoglio, S., Muerdter, F., Guzzardo, P.M., **Czech, B.**, Pappin, D.J., Chen, C., Gordon, A., Hannon, G.J. (2010). Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila. Genes & Development* **24**(22): 2499-2504.

8. Wang, J., **Czech, B.**, Crunk, A., Wallace, A., Mitreva, M., Hannon, G.J., Davis, R.E. (2011). Deep small RNA sequencing from the nematode *Ascaris* reveals conservation, functional diversification, and novel developmental profiles. *Genome Research* **21**(9): 1462-1477.

9. Muerdter, F., Olovnikov, I., Molaro, A., Rozhkov, N., **Czech, B.**, Gordon, A., Hannon, G.J., Aravin, A.A. (2012). Production of artificial piRNAs in flies and mice. *RNA* **18**(1): 42-52.

10. Toledano, H., D'Alterio, C., **Czech, B.**, Levine, E., Jones D.L. (2012). The *let-7*-Imp axis regulates ageing of the *Drosophila* testis stem-cell niche. *Nature* 485(7400): 605-10.

Contributions to publications

For all publications stated below, I contributed a great deal to the scientific ideas that were relevant to the project. Furthermore, I was significantly involved in the planning of experiments, the data generation, the analysis and interpretation of results, as well as the preparation of figures and the manuscript writing. Detailed contributions to each publication are stated below.

1. Zhou, R.*****, **Czech, B.***, Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., Hannon, G.J. (2009). Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA* **15**(10): 1886-1895.
*** joint first authors**

Specific contributions:

Julius Brennecke, Rui Zhou and I noticed the dependence of endo-siRNAs on Loquacious. Rui Zhou, Julius Brennecke, Greg Hannon and I planned all experiments. Rui Zhou prepared constructs, performed knockdown experiments and immunohistochemical analyses (Western blots, co-immunoprecipitations) (Fig.1 & Fig.2). Rui Zhou performed luciferase assays (Fig.S2). Rui Zhou (Fig.1) and I (Fig.3 & Fig.S1) performed small RNA Northern blots; fly crosses were carried out by myself (Fig.S1). I cloned and analysed small RNA libraries, computed length profiles, heat maps and density plots (Fig.3 & Fig.4). Ravi Sachidanandam contributed bioinformatic tools for library analysis. James Wohlschlegel provided proteomic data that helped validating the novel Loqs isoform Loqs-PD. Results were discussed between Rui Zhou, Julius Brennecke, Greg Hannon and myself, with suggestions from other authors. Norbert Perrimon and Greg Hannon supervised the project. Rui Zhou and I wrote the draft of the manuscript, which was edited by Greg Hannon, Rui Zhou and myself with comments from all authors.

2. **Czech, B.***, Zhou, R.*****, Erlich, Y., Brennecke, J., Binari, R., Villalta, C., Gordon, A., Perrimon, N., Hannon, G.J. (2009). Hierarchical rules for Argonaute loading in *Drosophila*. *Molecular Cell* **36**(3): 445-456.
*** joint first authors**

Specific contributions:

Julius Brennecke and I noted the association of miR* strands with AGO2. Rui Zhou, Greg Hannon and I planned all experiments. I performed small RNA cloning, library analysis and generated heat maps from total RNAs (Fig.1). I cloned and

analysed small RNA libraries from AGO1 and AGO2 immunoprecipitates and computed heat maps (Fig.2). I performed β-elimination on immunoprecipitated and total RNAs and carried out Northern blots (Fig.2 & Fig.4). I performed bioinformatic analyses to extract nucleotide biases and base pairing within duplexes preferentially bound by AGO1 or AGO2 (Fig.2). I prepared small RNA libraries from knockdown material, analysed sequences and computed heat maps and bar diagrams (Fig.4 & Fig.S4). I generated the scheme showing analysed duplex structures (Fig.S2). Rui Zhou generated a stable S2 cell line expressing FLAG/HA-AGO2 and analysed the transgene expression (Fig.S1). Rui Zhou performed Argonaute loading assays (Fig.3 & Fig.S3). Rui Zhou carried out knockdown experiments followed by luciferase assays (Fig.5). Rui Zhou prepared sensor constructs, and generated transgenic flies with help from Richard Binari and Christians Villalta. Rui Zhou carried out fly genetics, generated and analysed mitotic clones (Fig.6 & Fig.S5-10). Yaniv Erlich analysed thermodynamic properties of AGO2-associated RNAs (Fig.7). Assaf Gordon contributed bioinformatic tools for library analysis. All results were discussed between Rui Zhou, Greg Hannon and myself, with suggestions from other authors. Norbert Perrimon and Greg Hannon supervised the project. I wrote the draft of the manuscript, which was edited by Rui Zhou, Greg Hannon and myself incorporating comments from other authors.

3. **Czech, B.**, Hannon G.J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nature Reviews Genetics* **12**(1): 19-31. **[REVIEW ARTICLE]**

Specific contributions:

Greg Hannon and I developed the review concept and decided on content. I wrote the draft of the manuscript, which was edited by Greg Hannon and myself. Figures were re-drawn by a journal editor.

4. Ni, J.-Q.*, Zhou, R.*, **Czech, B.***, Liu, L.-P., Holderbaum, L., Yang-Zhou, D., Shim, H.-S., Tao, R., Handler, D., Karpowicz, P., Binari, R., Booker, M., Brennecke, J., Perkins, L.A., Hannon, G.J., Perrimon, N. (2011). A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nature Methods* **8**(5): 405-407.
*** joint first authors**

Specific contributions (adopted from published paper):

Jian-Quan Ni, Rui Zhou and Benjamin Czech performed major experiments. Lu-Ping Liu, Laura Holderbaum, Donghui Yang-Zhou, Hye-Seok Shim, Richard Binari, Rong Tao, Matthew Booker, and Lizabeth Perkins produced TRiP lines. Phillip

Karpowicz performed luciferase experiments in ovaries. Dominik Handler and Julius Brennecke analysed effects of shRNAs on the piRNA pathway during oogenesis. Greg Hannon and Norbert Perrimon supervised the project. Rui Zhou, Benjamin Czech, Jian-Quan Ni, Dominik Handler, Julius Brennecke, Greg Hannon and Norbert Perrimon wrote the manuscript.

Note:

I was mainly involved in optimizing the design of shRNAs and in testing the processing accuracy and abundance of shRNAs compared to other endogenous small RNAs. In addition, I amplified and generated the pooled shRNA libraries. I also identified and cherry-picked individual correct shRNA clones.

5. Preall, J.B.*, **Czech, B.***, Guzzardo, P.M., Muerdter, F., Hannon, G.J. (2012). *shutdown* is a component of the *Drosophila* piRNA biogenesis machinery. *RNA* **18**(8): 1446-1457.
*** joint first authors**

Specific contributions:

Jonathan Preall and I identified *shutdown* as piRNA pathway component, with supporting data provided by Paloma Guzzardo and Felix Muerdter. Jonathan Preall, Greg Hannon and I planned all experiments. Jonathan Preall performed homology analyses and computed evolutionary trees (Fig.1A). Jonathan Preall performed qPCR experiments, scored patterning defects and female fertility, and carried out immunofluorescent analyses upon knockdown (Fig.2). Jonathan Preall studied the subcellular localization of piRNA pathway components in cell culture (Fig.S2). Jonathan Preall and I analysed transposon de-repression and fertility upon germline knockdown (Fig.1D). Jonathan Preall, Paloma Guzzardo and Felix Muerdter carried out and analysed RNAseq experiments (Fig.1C). The domain structures of *Drosophila* FKBP family members was derived by myself (Fig.1B). I prepared and analysed small RNA libraries, computed density plots for siRNAs targeting depleted genes (Fig.S1), generated bar diagrams and heat maps for miRNAs (Fig.S3), and created density plots and length profiles for piRNAs derived from clusters (Fig.3). I computed heat maps and density plots for transposon-derived piRNAs (Fig.4). Felix Muerdter contributed a bioinformatic tool for ping-pong analysis and generated ping-pong scores for cluster and transposon derived piRNAs (Fig.3 & Fig.4). Jonathan Preall, Greg Hannon and I discussed all results, with suggestions from other authors. Greg Hannon supervised the project. Jonathan Preall and I wrote the

manuscript draft, which was edited by Greg Hannon, Jonathan Preall and myself with comments from other authors.

6. **Czech, B.***, Preall, J.B.*, McGinn, J., Knott, S.R., Hannon, G.J. (*). A transcriptome-wide RNAi screen in *Drosophila* ovaries reveals novel factors of the germline piRNA pathway. (*in preparation for submission to Cell*).
*** joint first authors**

Specific contributions:

Based on my preliminary data, Greg Hannon and I formed the concept to carry out an RNAi screen in germline cells of the female ovary. Jonathan Preall, Greg Hannon and I planned all experiments. I performed pilot experiments to optimize the RNA isolation procedure and to establish the multiplexed qPCR protocol, with help from Jonathan Preall. Jonathan Preall built fly stocks that facilitate virgin collection. Jonathan Preall and I performed and analysed RNAseq experiments to identify ovary-expressed genes (Fig.1). Jonathan Preall and I carried out the primary screen (i.e., fly husbandry and crosses, material collection, RNA isolation, reverse transcription, qPCR followed by subsequent data analysis), with help from Jon McGinn (Maria Mosquera and Steven Sau also contributed to fly husbandry and qPCR) (Fig.1). Simon Knott (heat maps) and I (box plots) performed bioinformatic data analyses to present screen hits (Fig.2 & Fig.S2). Jonathan Preall, Jon McGinn and I conducted the secondary screen (Fig.3). Jonathan Preall carried out immunofluorescent analyses (Fig.4). I prepared and analysed small RNA libraries (Fig.5). Jonathan Preall and I performed data analyses and identified specific requirements for silencing of different transposon classes (Fig.6). Jonathan Preall, Greg Hannon and I discussed all results, with suggestions from other authors. Greg Hannon supervised the project. I wrote the draft of the manuscript, which was edited by Jonathan Preall, Greg Hannon and myself to incorporate comments from other authors.

The manuscript is not published, thus the final version might vary from the one presented in this thesis.

Zusammenfassung

Eukaryonten steuern eine Vielzahl biologischer Prozesse wie die Regulation der Genexpression, die Änderung der Chromatinstruktur und die Sicherstellung der Genomstabilität mittels kleiner regulatorischer RNAs. Kleine RNAs werden anhand ihrer Biogenese-Mechanismen und dem Argonaute Protein, an welches sie binden, in drei Hauptklassen unterteilt: MicroRNAs (miRNAs), kleine interferierende RNAs (engl. small interfering RNAs, siRNAs) und Piwi-interagierende RNAs (engl. Piwi-interacting RNAs, piRNAs).

Die etwa 22 bis 23 Nukleotid (nt) langen miRNAs sind ubiquitär exprimiert, werden von Drosha/Pasha und Dcr1/Loquacious Protein-Komplexen aus Vorläufer-Molekülen mit lokaler doppelsträngiger RNA (dsRNA) prozessiert, und binden in *Drosophila* überwiegend an AGO1. Dagegen werden die ungefähr 21 nt langen siRNAs aus langer dsRNA, die endogenen oder exogenen Ursprungs sein kann, von Dcr2 hergestellt und binden an AGO2. Obwohl die Biogenese von miRNAs und siRNAs unterschiedliche Faktoren erfordert, besitzen sie ähnliche Zwischenprodukte. Ihnen gemeinsam sind kleine RNA Duplexe. Da sich beide Klassen kleiner RNAs jedoch deutlich in ihren biologischen Wirkungsweisen unterscheiden, muss die Assoziation von miRNAs und siRNAs mit den spezifischen Argonaute Proteinen strikt reguliert sein. Dieser Prozess wird auch als „Sortierung“ (engl. „sorting“) bezeichnet. Aufgrund früherer Daten wurde vorgeschlagen, dass die Duplexstruktur die Sortierung entscheidend beeinflusst und Duplexe mit perfekter dsRNA mit AGO2 assoziieren, wohingegen Duplexe mit mehreren ungepaarten Basen überwiegend in AGO1 geladen werden. Der erste Teil dieser Arbeit belegt, dass die Biogenese von miRNAs und siRNAs entgegen früherer Annahmen stärker verknüpft ist als vermutet und bestehende Modelle für die Sortierung von kleinen RNAs stark vereinfacht sind. Die vorliegende Arbeit charakterisiert eine neuartige Isoform des dsRNA-Bindeproteins Loquacious, welches bisher nur für die Produktion von miRNAs bekannt war. Diese Isoform ist entscheidend an der Herstellung aller untersuchter Typen von endogenen siRNAs beteiligt.

Als nächstes wurde mittels Hochdurchsatz-Sequenzierung von AGO1- und AGO2-Immunoprezipitaten gezeigt, dass miR Stränge an AGO1 binden, während miR* Stränge, die ursprünglich für reine Nebenprodukte der miRNA Biogenese gehalten wurden, völlig unerwartet mit AGO2 assoziieren. Da beide Stränge eines miRNA Vorläufermoleküls in unterschiedliche Argonaute Proteine geladen werden, kann die Duplexstruktur jedoch nicht alleine für die Sortierung verantwortlich sein. Unsere Daten

belegen vielmehr, dass kleine RNAs auf Grund mehrerer Parameter auf beide Argonaute Proteinen verteilt werden. Basierend auf unseren Ergebnissen haben wir diese Parameter extrahiert und eine Rangordnung aufgestellt, welche schließlich experimentell getestet wurde. Zuletzt wurden diese validierten Sortierungs-Parameter verwendet, um effiziente shRNAs (engl. short hairpin RNAs) – künstliche molekulare Werkzeuge zur gezielten Abschaltung jedes beliebigen Gens – für *Drosophila* zu entwickeln. Ein weiteres Ziel war die Herstellung einer genomweiter shRNA-Bibliothek zur Durchführung von Hochdurchsatz-Screenings.

Die dritte Klasse kleiner RNAs, die piRNAs, sind in Zellen der Keimbahn exprimiert, etwa 23 bis 28 nt lang und assoziieren mit PIWI Proteinen. *Drosophila* besitzt drei Proteine der PIWI-Familie: Piwi, Aub, und AGO3. Im Gegensatz zu miRNAs und siRNAs werden piRNAs aus einzelsträngigen Vorläufermolekülen produziert. Trotz intensiver Bemühungen in den letzten Jahren wurden nur begrenzte Fortschritte gemacht, um die Biogenese und Wirkmechanismen von piRNAs genau aufzuklären. Insbesondere über die Herstellung der primären piRNAs ist wenig bekannt. Der zweite Teil dieser Arbeit zielt darauf ab, dieses Problem anzugehen. Zuerst wurde *Shutdown* als neuer Faktor für die Biogenese von primären und sekundären piRNAs identifiziert. Um weitere, bisher in *Drosophila* nicht bekannte Faktoren zu finden wurde ein innovativer auf RNA Interferenz basierendes Screening durchgeführt. Dabei wurde untersucht, welchen Effekt die gezielte Abschaltung („Gen-Knockdown“) jedes einzelnen der circa 8,400 in den Keimzellen der Ovarien exprimierten Gene auf die Repression von Transposons oder die weibliche Fruchtbarkeit haben. Dieser Screen führte zur Identifizierung Dutzender neuer Gene welche die Biogenese von piRNAs und ihre Wirkmechanismen beeinflussen.

Summary

Eukaryotic organisms utilise small RNAs to control many biological processes including gene regulation, chromatin modifications and genome protection from threats like viruses or transposons. Small RNAs are distinguished based on their biogenesis mechanisms and associating Argonaute protein partners into three major classes: microRNAs (miRNAs), small interfering RNAs (siRNAs), and Piwi-interacting RNAs (piRNAs).

In *Drosophila*, miRNAs and siRNAs derive from double-stranded RNA and associate with AGO-clade Argonaute proteins, AGO1 and AGO2. These two small RNA classes share similar processing intermediates, small RNA duplexes, but differ significantly in their biological mode of action. Therefore, pairing of miRNAs and siRNAs with specific Argonaute proteins must be tightly controlled through a process known as "sorting". Previous models suggested independent pathways to direct miRNAs and siRNAs into AGO1 and AGO2, respectively. The first part of this thesis provides evidence contrary to those models. First, a novel isoform of the dsRNA-binding protein Loqs (normally associated with miRNAs) was identified that is crucial for processing of endogenous siRNAs. Second, an unexpected association of AGO2 with miR* strands was revealed through profiling of bound small RNAs. Taken together, our data indicates a greater flexibility and inter-connection between the two pathways, with small RNAs being sorted based on their intrinsic duplex structures. Based on these results, we extracted and experimentally tested a hierarchy of determinants that governs small RNA sorting in *Drosophila*. These sorting rules were then applied to design efficient short hairpin RNAs (shRNAs) for artificial gene silencing, with the goal of generating a genome-wide shRNA library for *Drosophila*.

The third *Drosophila* small RNA class, piRNAs, associate with members of the PIWI subfamily of Argonaute proteins: Piwi, Aub and AGO3. Despite intense efforts, limited progress has been made in recent years to uncover the biogenesis and effector mechanisms of piRNAs. In particular, the production of primary piRNAs is poorly understood. The second part of this thesis aims to address this question. First, *shutdown* was identified as novel biogenesis factor of primary and secondary piRNAs. To uncover missing piRNA pathway components, an innovative, transcriptome-wide RNAi screen was carried out in germ cells of the *Drosophila* ovary. The screen revealed dozens of new genes that impact piRNA biogenesis, transposon silencing and female fertility.

1 Introduction

A diverse repertoire of small RNA molecules has emerged in eukaryotes, which fulfil manifold functions and act in a myriad of biological processes. This chapter will introduce key concepts and components of small RNA pathways, followed by a detailed description of the three major small RNA classes, microRNAs (miRNAs), small interfering RNAs (siRNAs), and Piwi-interacting RNAs (piRNAs), with an emphasis on their biogenesis mechanisms and effector functions. Although this introduction is focused mainly on *Drosophila* small RNAs, noteworthy commonalities and differences in other organisms are also highlighted. Parts of the figures and text in this chapter are modified from a review article published in *Nature Reviews Genetics* (Czech & Hannon, 2011).

1.1 Overview of small RNA pathways

For decades, RNA was regarded largely as a carrier of genetic information that bridges DNA and protein synthesis. Subsequent RNA profiling revealed that only a minor fraction of RNAs possesses protein-coding capacity, and it is now well established that cellular organisms also express a massive repertoire of non-protein-coding transcripts, also known as non-coding RNAs (ncRNAs). Prominent types of ncRNAs include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), and various types of regulatory RNAs. These regulatory RNAs can be subdivided into small RNAs that employ RNA interference (RNAi) related machineries (i.e. associate with proteins of the Argonaute family) and those that do not. Up to now, little is known about most groups of regulatory RNAs that are not associated with RNAi (e.g., long intergenic non-coding RNAs, lincRNAs). In contrast, our knowledge of the biogenesis mechanisms and functional significance of RNAi-related small RNAs is expanding rapidly.

Since the discovery of RNAi in the late 1990s (Fire *et al.*, 1998), several small RNA types have been identified in eukaryotes. In *Drosophila*, three major classes are distinguished: microRNAs (miRNAs), small interfering RNAs (siRNAs), and Piwi-interacting RNAs (piRNAs). Common to all aforementioned RNAi-related classes is their typical length (19- to 30-nt) and association with members of the Argonaute protein family for their effector functions, with a single-stranded small RNA bound to an Argonaute protein forming the core of the so-called RNA-induced silencing complex (RISC). These small RNAs guide RISC via complementary base pairing to target transcripts, and enable RISC to carry out its effectors function. RISC is engaged in a myriad of biological processes, ranging from repression of transposable elements to

ensure reproductive success, over precisely controlling gene expression via post-transcriptional silencing, to altering chromatin states during development. A failure to correctly assemble a functional RISC can severely perturb normal growth and development, often with fatal consequences (Bohmert *et al*., 1998; Cox *et al*., 1998; Moussian *et al*., 1998; Liu *et al*., 2004). In order to avoid this, the generation and loading of small RNAs into Argonaute proteins must be tightly controlled.

The type of small RNA and the properties of the associated Argonaute protein (e.g., catalytic activity or subcellular protein localization) define their corresponding effector mechanism. Distinct biogenesis events upstream result in small RNAs with unique properties that can affect their sorting and association with different Argonaute proteins. Most miRNAs and siRNAs directly originate from double-stranded RNA (dsRNA) that can form through intramolecular (within a single RNA molecule) or intermolecular (between two RNA molecules) interactions, respectively. Small RNAs derived from dsRNAs require the action of type III Ribonuclease (RNase) family proteins (Ghildiyal & Zamore, 2009). These are sequence-independent, dsRNA-specific endonuclease enzymes that are found throughout all kingdoms from prokaryotes and yeast, to plants and animals (MacRae & Doudna, 2007). In contrast, piRNAs are generated from single-stranded precursors without distinct secondary structures (Brennecke *et al.*, 2007). The biogenesis of piRNAs engages alternative mechanisms that are independent of RNase III enzymes (Vagin *et al.*, 2006; Houwing *et al.*, 2007), though the molecular components remain to be identified.

Following their production, small RNAs are assigned to specific proteins of the Argonaute family, which can be divided into three subgroups based on sequence relationships: The AGO-clade, the PIWI-clade and the worm-specific WAGO subfamily (Yigit *et al.*, 2006; Tolia & Joshua-Tor, 2007; Hutvagner & Simard, 2008). AGO-clade members usually pair with RNase III-derived small RNA duplexes (miRNAs and siRNAs), whereas PIWI-clade proteins load small RNAs derived from single-stranded precursors (piRNAs) (Ghildiyal & Zamore, 2009). WAGO-clade proteins associate with secondary siRNAs that are direct RNA-dependent RNA polymerase (RdRP) products (Yigit *et al.*, 2006; Pak & Fire, 2007; Sijen *et al.*, 2007). The *Drosophila* genome encodes five Argonaute proteins: AGO1, AGO2 (both ubiquitous AGO-clade members), and Piwi, Aub and AGO3 (germline-specific PIWI-clade members).

While in mature RISC a single-stranded RNA is associated with an Argonaute protein, its precursor (in the case of miRNAs and siRNAs) exists as short dsRNA, also known as small RNA duplex. In *Drosophila*, AGO-clade proteins are initially loaded with such a duplex, from which only one strand (termed the guide or miR strand) is retained,

whereas the other strand (the passenger or miR* strand) is removed (Matranga *et al.*, 2005; Miyoshi *et al.*, 2005; Rand *et al.*, 2005; Kawamata *et al.*, 2009). As loading of passenger strands would misdirect RISC to a different set of targets, selection and stabilization of the desired guide strand is of key importance, and suggested to be influenced by molecular and thermodynamic parameters (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003). Following correct assembly of mature RISC, the complex is guided to targets via complementary base pairing, and in most cases leads to target repression (Bartel, 2009; Brodersen & Voinnet, 2009; Fabian *et al.*, 2010).

1.2 MicroRNA biogenesis and function

MicroRNAs (miRNAs) impact various developmental processes by regulating the expression of target genes at the post-transcriptional level (Bartel, 2009; Fabian *et al.*, 2010). Mature miRNAs, 21- to 23-nt in size, originate from long precursor transcripts with local hairpin structures. Although encoded in most eukaryotic genomes, such as protozoa, plants and animals, there are many variations in the biogenesis and regulatory mechanism utilized, which suggests independent evolutions of miRNAs as small regulators in plant and animal kingdoms (Bartel, 2004; Cuperus *et al*., 2011). The first miRNAs – *lin-4* and *let-7* – were uncovered in worms to be responsible for developmental heterochrony (Lee *et al.*, 1993; Wightman *et al.*, 1993). To date, thousands of miRNAs have been identified in other species either by sequencing (accelerated by application of next generation sequencing technologies) or through predictive computational approaches (Kozomara & Griffiths-Jones, 2011). Numerous miRNAs exist in families that are comprised of miRNAs sharing sequence similarities while originating from distinct precursor transcripts. Differential temporal and spatial expression patterns of individual miRNA family members allows for increased complexity in regulating their target genes. While some miRNAs show deep evolutionary conservation between many species from different clades of the phylogenetic tree, others are species-specific indicating an on-going emergence and evolution of novel miRNAs. Giving this evolutionary flexibility, it is no surprise that microRNAs have been co-opted into numerous biological pathways, including cell proliferation and apoptosis in animals and control of leaf and flower development in plants (Brennecke *et al*., 2003; Palatnik *et al*., 2003; Chen, 2004).

Canonical miRNA genes are transcribed by DNA-dependent RNA polymerase II (Pol II) to generate primary transcripts (pri-miRNAs) containing at least one region of imperfect dsRNA known as the stem-loop, which harbours the future mature miRNA (Fig.

1-1) (Lee *et al.*, 2004a). Pri-miRNAs range in size from a few hundred base pairs to several kilobases, feature 5' cap structures and polyA tails, and may be subject to splicing (Cai *et al.*, 2004). Animal miRNAs are typically organized in polycistronic transcription units or within large introns (Rodriguez *et al.*, 2004), whereas most plant miRNAs originate from independent transcription units (Jones-Rhoades *et al.*, 2006). The concatenation of several miRNAs into clusters in animals allows coordinated and robust target regulation through increased silencing trigger levels.

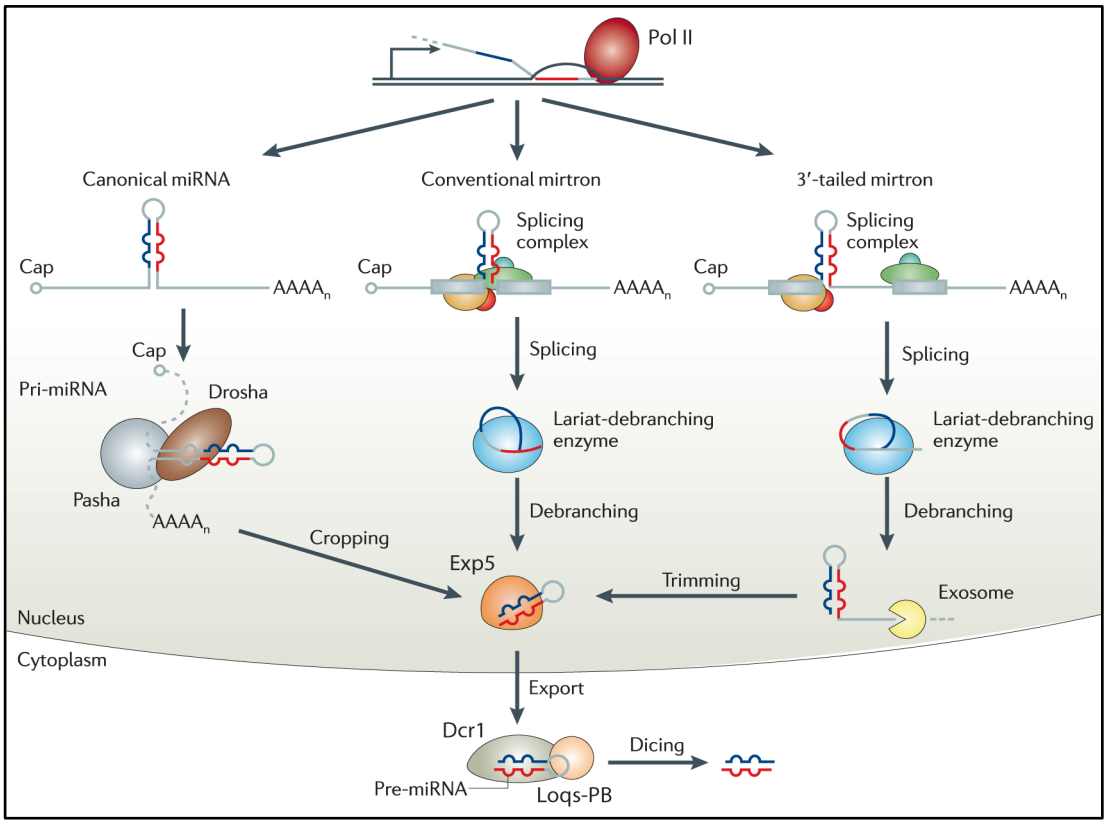


Fig. 1-1. MicroRNA biogenesis in *Drosophila melanogaster*. Most miRNAs are transcribed by RNA polymerase II (Pol II) to yield pri-miRNAs. Processing by the Drosha/Pasha complex releases the shorter pre-miRNAs (left). Mirtrons are miRNAs that reside within short introns of protein-coding genes. They are excised by the splicing machinery, and following debranching fold into pre-miRNAs (middle). Some 3' tailed mirtrons require further trimming by the exosome (right). Pre-miRNAs are exported to the cytoplasm by Exp5, where further processing by Dcr1/Loqs-PB complexes takes place to liberate miR:miR* duplexes. (Image and legend modified from: Czech & Hannon, 2011)

Animal miRNA biogenesis is initiated by the nuclear double-stranded RNA-binding (dsRBD) protein, Pasha, which binds the pri-miRNA and recruits the RNase III enzyme, Drosha (Lee *et al.*, 2003; Denli *et al.*, 2004; Gregory *et al.*, 2004; Han *et al.*, 2004a; Landthaler *et al.*, 2004). Drosha/Pasha complexes (also known as microprocessor) cleave pri-miRNAs at the base of the hairpin region, leading to the release of ~60- to 70-nt long stem loop intermediates known as precursor miRNAs (pre-miRNAs) (Lee *et al.*, 2003; Han *et al.*, 2006). Pre-miRNAs contain 2-nt single-stranded 3' overhangs, a

hallmark of RNase III cleavage products, which are recognized by the nuclear export protein Exportin-5 (Exp5), and actively transported to the cytoplasm (Yi *et al.*, 2003; Bohnsack *et al.*, 2004; Lund *et al.*, 2004). Additional factors including Arsenite-resistance protein 2 (Ars2), nuclear export receptor Exportin 1 (Xpo1), and the cap-binding complex were also implicated in the conversion of pri-miRNAs to pre-miRNAs, but their precise molecular function is still unknown (Sabin *et al.*, 2009; Gruber *et al.*, 2009; Bussing *et al.*, 2010).

Following nuclear export, a second processing step takes place in the cytoplasm, where pre-miRNAs are cleaved into ~22- to 23-nt miR:miR* duplexes by the RNase III enzyme Dicer-1 (Dcr1) (Bernstein *et al.*, 2001; Grishok *et al.*, 2001; Hutvagner *et al.*, 2001; Ketting *et al.*, 2001). In *Drosophila*, Dcr1 interacts with a specific isoform of the dsRBD protein Loquacious (Loqs), Loqs-PB (Förstemann *et al.*, 2005; Jiang *et al.*, 2005; Saito *et al.*, 2005; Park *et al.*, 2007). Mutations in either Dcr1 or Loqs result in reduced levels of mature miRNAs accompanied by pre-miRNA accumulation. Dcr1 cleavage generates small RNA duplexes that exhibit RNase III-characteristic 2-nt single-stranded 3' overhangs at both ends, a necessary signal for downstream RISC loading.

In contrast to animals, processing of primary miRNA transcripts in plants is initiated directly by the nuclear RNase III family protein DICER-LIKE 1 (DCL1) (Park *et al.*, 2002; Reinhart *et al.*, 2002; Kurihara & Watanabe, 2004; Henderson *et al.*, 2006). In the absence of a Drosha ortholog, DCL1 fulfils Drosha and Dicer functions by performing consecutive cleavage events that generate miR:miR* duplexes.

Recently, a number of unconventional miRNAs have been identified in various animals that use alternative maturation strategies. These miRNAs bypass either Drosha (mirtrons) or Dicer processing (e.g., *miR-451*). Mirtrons originate from short intronic stem loops, and utilize the splicing machinery for debranching and refolding into pre-miRNAs (Berezikov *et al.*, 2007; Okamura *et al.*, 2007; Ruby *et al.*, 2007). Some mirtrons with extended 3' tails additionally require trimming by the exosome to generate pre-miRNAs that can be further processed (Flynt *et al.*, 2010). Exp5 then exports pre-miRNAs in a similar fashion as canonical miRNAs (Fig. 1-1). While mirtrons are common in *Drosophila*, miRNAs escaping Dicer processing have only been described in fish and mouse: here, Drosha/Pasha complexes conventionally process primary *miR-451* transcripts, but the resulting pre-*miR-451* is not processed by Dicer. Instead, pre-*miR-451* is directly loaded into the catalytically active Argonaute 2 (AGO2), and further processed by AGO2-mediated intramolecular cleavage of the hairpin, followed by resection by an unknown enzymatic activity to generate the mature *miR-451* (Cheloufi *et al.*, 2010; Cifuentes *et al.*, 2010; Yang *et al.*, 2010).

Increasing evidence supports the notion that miRNA biogenesis itself is subject to substantial regulation. For example, Lin28 in worms and mammals promotes the uridylation of the 3' end of pre-miRNAs (Heo *et al.*, 2008; Lehrbach *et al.*, 2009), which prevents processing by Dicer and instead leads to degradation. MicroRNA precursors can also be edited by dsRNA-specific adenosine deaminases (ADARs) (Yang *et al.*, 2006), changing their processing efficiencies and potentially also their target spectrum. Lastly, increasing numbers of post-translational modifications of components of microRNA biogenesis have been uncovered (Johnston & Hutvagner, 2011).

In *Drosophila*, following production by Dcr1, miR:miR* duplexes are assembled into Argonaute proteins, typically AGO1 with a few exceptions (Förstemann *et al.*, 2007; Czech *et al.*, 2008; Ghildiyal *et al.*, 2008). During this process – termed 'loading' – the duplex is released from Dicer and, assisted by other yet to be determined factors, incorporated into AGO1. To yield mature AGO1-RISC, which is programmed with a single-stranded RNA, one strand of the duplex must be removed. The strand most commonly retained is named the miR strand, whereas the complementary miR* strand is discarded. The selection of miR strands must be accurate and precise, as loading of the miR* strand could erroneously silence different targets. Determinants for strand choice are encoded in the miR:miR* duplex and include the identity of the 5' nucleotide as well as the duplex structure (see Results and Discussion). Interestingly, some *Drosophila* miRNAs are modified subsequent to their incorporation into AGO1. The exonuclease Nibbler was shown to resect several AGO1-bound miR strands at their 3' termini (Han *et al.*, 2011; Liu *et al.*, 2011), with the significance of this additional processing event yet to be resolved.

In general, once paired with an Argonaute family protein, miRNAs guide RISC to its target transcripts by base pairing with complementary sequences. The types of target regulation depend on the amount of sequence complementarity between small RNA and target, and the involved Argonaute protein, which differ in their biochemical properties (Bartel, 2009; Brodersen & Voinnet, 2009; Fabian *et al.*, 2010), with major differences between animals and plants. Animal miRNAs regulate their targets without requirement of Argonaute cleavage activity through interactions with limited base pairing between their 'seed' region and the 3' UTRs of targeted transcripts, which is insufficient to mediate cleavage and typically results in reduced protein synthesis (Guo *et al.*, 2010). RISC-targeted mRNAs are destabilized by AGO1 and its partner Gawky (also known as GW182) via disruption of interactions between the polyA tail and mRNA cap (Eulalio *et al.*, 2007; Iwasaki *et al.*, 2009). This causes deadenylation, decapping and subsequent

degradation of the transcript. Many targets regulated by miRNAs also localize to processing bodies (P-bodies), cytoplasmic foci thought as sites of mRNA storage and/or degradation (Liu *et al.*, 2005). Although both *Drosophila* AGO-clade proteins are capable of cleaving their targets (Okamura *et al.*, 2004), the activity of AGO1 – the principal acceptor of miRNAs – is limited due to rapid dissociation from its target transcripts (Förstemann *et al.*, 2007). MiRNA-directed target cleavage is rare in animals and has only been reported in cases with extensive target complementarity (Yekta *et al.*, 2004; Davis *et al*., 2005). In contrast, plant miRNA directly cleave their target transcripts by binding with perfect or near-perfect sequence complementarity (Llave *et al.*, 2002; Rhoades *et al.*, 2002). However, recent work indicates the ubiquity of cleavage-independent repression modes in plants, even for targets with extensive sequence complementarity (Brodersen *et al.*, 2008).

1.3 Processing and roles of small interfering RNAs

The first small interfering RNAs (siRNAs) were reported in plants as result of the introduction of dsRNA either through transgene copies or viral replication intermediates (Hamilton & Baulcombe, 1999). Numerous types of siRNAs have since been characterized in a variety of organisms, with functions ranging from regulation of endogenous gene expression, to defence against parasitic nucleic acids like viruses and transposons. We have also co-opted the siRNA pathway using artificial dsRNA triggers to efficiently silence genes, which is an increasingly important tool for biological studies and potentially therapeutics. There are currently two major siRNA categories, distinguished by their dependence on (or independence of) the action of RNA-dependent RNA polymerases (RdRPs) for their biogenesis. Since the genomes of *Drosophila* and mammals do not encode RdRPs, they only generate RdRP-independent siRNAs, with the dsRNA triggers arising via different mechanisms.

In flies, endogenous siRNAs result from the processing of dsRNA generated from extensive hairpin structures, units of convergent transcription, or the annealing of sense and antisense RNAs from unlinked loci (e.g., hybridization of sense transposon transcript with antisense cluster RNA) (Chung *et al.*, 2008; Czech *et al.*, 2008; Ghildiyal *et al.*, 2008; Kawamura *et al.*, 2008; Okamura *et al.*, 2008a). Fly endo-siRNAs play a role in various processes ranging from control of gene expression, over transposon silencing (mainly in somatic tissues), to defence against viruses (Chung *et al.*, 2008; Czech *et al.*, 2008; Ghildiyal *et al.*, 2008; Kawamura *et al.*, 2008; Okamura *et al.*, 2008a).

The conversion of long dsRNA into siRNAs is well understood in *Drosophila*. Long dsRNAs – either endogenously or exogenously introduced – are processed through sequential dicing into 21-nt siRNA duplexes by the RNase III enzyme Dicer-2 (Dcr2) (Liu *et al.*, 2003; Lee *et al.*, 2004b), assisted by Loqs (Fig. 1-2a) (Czech *et al.*, 2008; Okamura *et al.*, 2008a). Like miRNAs, *Drosophila* siRNAs are generated as dsRNA duplexes, which must be loaded accurately into Argonaute (Fig. 1-2b). As a consequence of perfect or near-perfect dsRNA within duplexes, fly siRNAs typically load into Argonaute 2 (AGO2) (see Results and Discussion) (Okamura *et al.*, 2004; Tomari *et al.*, 2007), and require a designated loading machinery that comprises Dcr2 and the dsRBD protein R2D2 (Liu *et al.*, 2003; Lee *et al.*, 2004b; Pham *et al.*, 2004; Tomari *et al.*, 2004; Liu *et al.*, 2006). Recent work suggests that AGO2 is conformationally altered into an 'open' configuration by the Hsp90/Hsc70 chaperone, which makes AGO2 amenable to loading with small RNA duplexes (Iki *et al.*, 2010; Iwasaki *et al.*, 2010; Johnston *et al.*, 2010; Miyoshi *et al.*, 2010b). R2D2 then senses thermodynamic asymmetry of small RNA duplexes and binds to the more stable end, while Dcr2 is positioned to the less stable end, thus orientating the duplex for loading into 'open' AGO2 (Tomari *et al.*, 2004). Once primed with a duplexed siRNA, AGO2 cleaves the passenger strand (Matranga *et al.*, 2005; Miyoshi *et al.*, 2005; Rand *et al.*, 2005), which is rapidly degraded by an endonuclease complex, C3PO (component 3 promoter of RISC), consisting of several copies of Translin and Trax (Liu *et al.*, 2009). Following removal of passenger strand fragments, the remaining AGO2-bound guide strand is methylated at the 3' terminus by the 2'-*O*-methyltransferase Hen1 to yield mature AGO2-RISC (Horwich *et al.*, 2007; Saito *et al.*, 2007). Mature AGO2-RISC typically regulates targets through mRNA cleavage (Okamura *et al.*, 2004; Czech *et al.*, 2008), although translational repression modes have also been reported from *in vitro* studies (Iwasaki *et al.*, 2009).

Contrary to flies and mammals, worms and plants possess RdRPs and consequently produce siRNAs via different mechanisms. RdRPs either copy single-stranded precursors into long dsRNAs (plants) or *de novo* synthesize the siRNA (secondary siRNAs in nematodes). RdRPs are not only important factors for biogenesis, but also allow a specific and robust amplification of substrates to generate small RNA silencing triggers. For instance, *Caenorhabditis elegans* produces two major types of endogenous siRNAs: primary siRNAs that are conventionally processed from long dsRNA via DCR-1 (Grishok *et al.*, 2001; Ketting *et al.*, 2001; Knight & Bass, 2001), and secondary siRNAs as a product of direct RNA synthesis by RdRPs (Fig. 1-2c) (Pak & Fire, 2007; Sijen *et al.*, 2007). The production of secondary siRNAs, however, depends on the availability of primary siRNAs that associate with the AGO-clade protein RDE-1.

Primary siRNAs guide RDE-1 to target RNAs via complementary base pairing, which causes the recruitment of an RdRP. Using the target transcript as template, the RdRP then synthesizes 22- to -24-nt secondary siRNAs that associate with WAGO-clade proteins. As a result of *de novo* synthesis, secondary siRNAs contain triphosphate groups at their 5' termini (Pak & Fire, 2007; Sijen *et al.*, 2007), which might be a contributing determinant for loading into WAGO proteins.

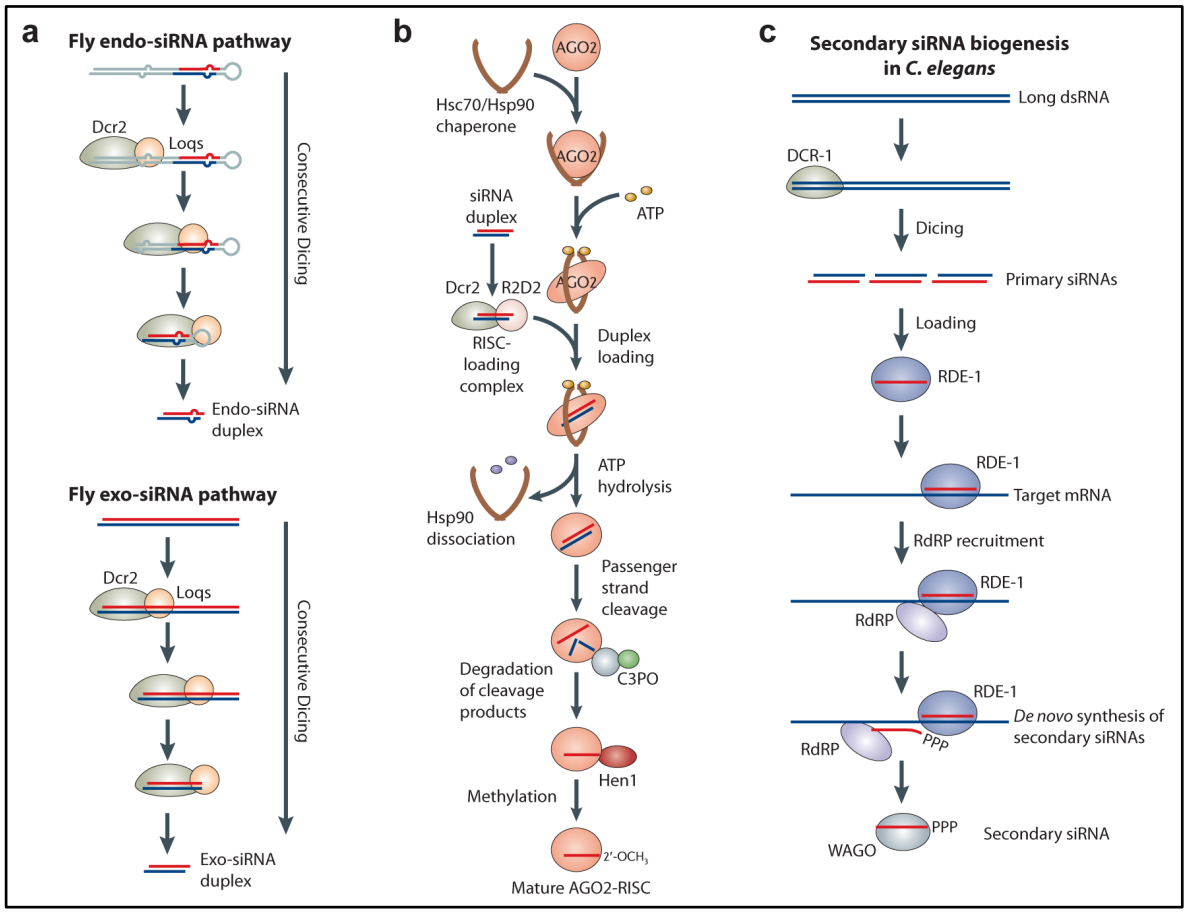


Fig. 1-2. Production and loading of small interfering RNAs. (a) In flies, perfect or nearly perfect dsRNA precursors of various origins are processed into siRNA duplexes by Dcr2/Loqs complexes. **(b)** In *Drosophila*, unloaded AGO2 is recognized by the chaperone complex consisting of Hsc70/Hsp90. ATP binding induces conformational changes with AGO2 adopting an "open" state. Loading-competent AGO2 receives an siRNA duplex from the RISC-loading machinery (Dcr2/R2D2 heterodimers). ATP hydrolysis causes the dissociation of the chaperone complex, which is followed by AGO2-mediated passenger strand cleavage ("slicing"). Subsequently, C3PO degrades cleavage products and the guide strand is methylated by Hen1 to yield mature AGO2-RISC. **(c)** Example of RdRP-derived siRNAs. In *C. elegans*, primary siRNAs are produced from long dsRNA by DCR-1, and loaded into the Argonaute RDE-1. SiRNA-primed RDE-1 is guided to its target transcripts, resulting in RdRP recruitment. Using the target RNA as template, secondary siRNAs are *de novo* synthesized by the RdRP and associate with Argonautes of the WAGO-clade. (Image and legend modified from: Czech & Hannon, 2011)

In plants, three major types of endogenous siRNAs can be distinguished: *trans*-acting siRNAs (ta-siRNAs), natural antisense transcript-derived siRNAs (nat-siRNAs), and heterochromatic siRNAs (hc-siRNAs). Each of these small RNA types utilizes the

action of RdRPs during their biosynthesis to convert single-stranded RNA into dsRNA. In addition, they are generated by specific Dicer family members and preferentially loaded into distinct Argonaute proteins. As a result of different Dicer and Argonaute proteins being involved, distinct siRNA types feature characteristic lengths and nucleotide biases (e.g., 21-nt ta-siRNAs, 24-nt hc-siRNAs) (see excellent reviews for details) (Chapman & Carrington, 2007; Ghildiyal & Zamore, 2009).

1.4 Piwi-interacting RNAs and transposon control

Piwi-interacting RNAs (piRNAs) are the most recent addition to the major small RNA classes. They are typically 24- to 30-nt in length and associate exclusively with Argonaute proteins of the PIWI-clade, whose expression is restricted to the gonads of animals. Consequently, piRNAs have only been found in reproductive tissues of mammals, fish, worms, and flies (Aravin *et al.*, 2006; Girard *et al.*, 2006; Grivna *et al.*, 2006; Lau *et al.*, 2006; Ruby *et al.*, 2006; Saito *et al.*, 2006; Vagin *et al.*, 2006; Watanabe *et al.*, 2006; Brennecke *et al.*, 2007; Carmell *et al.*, 2007; Gunawardane *et al.*, 2007; Houwing *et al.*, 2007; Tam *et al.*, 2008; Watanabe *et al.*, 2008). Defects in Piwi proteins or other core components of the piRNA pathway lead to sterility (Deng & Lin, 2002; Carmell *et al.*, 2007; Li *et al.*, 2009), often accompanied by increased mobility of transposable elements (Sarot *et al.*, 2004; Savitsky *et al.*, 2006; Vagin *et al.*, 2006; Aravin *et al.*, 2007). Thus, piRNAs ensure successful reproduction by protecting germline genomes from deleterious effects of uncontrolled transposon activity.

The piRNA pathway is best characterized in *Drosophila*, which possesses three PIWI-clade proteins: *P*-element induced wimpy testis (Piwi), Aubergine (Aub), and Argonaute 3 (AGO3), each of which show specific expression patterns and bind different subsets of small RNAs. PiRNA populations in flies are rich in sequences matching repeats and transposable elements (Brennecke *et al.*, 2007). This is in contrast to murine piRNAs, which are variable depending on the PIWI-clade protein and developmental time point (Aravin *et al.*, 2007; Aravin *et al.*, 2008).

Recent studies identified a complex conceptual network underlying the piRNA system, where one pathway produces primary piRNAs that serve as initial silencing triggers, and another, the ping-pong amplification cycle, builds on the availability of primary piRNAs to produce secondary piRNAs against highly active transposon threats. Remarkably, contrary to other small RNA systems, the ping-pong loop and consequently piRNA amplification takes place without the action of an RdRP.

The presence of two different ovarian cell types in flies adds further complexity to the piRNA system (Malone *et al.*, 2009). The germ cells co-express all three PIWI-clade proteins, Piwi, Aub, and AGO3, and have operating primary biogenesis and ping-pong cycle (Harris & Macdonald, 2001; Brennecke *et al.*, 2007; Gunawardane *et al.*, 2007; Li *et al.*, 2009; Malone *et al.*, 2009). In contrast, somatic follicle cells, which surround and protect the germ cells, solely express Piwi, resulting in primary biogenesis as the only active pathway in these cells (Brennecke *et al.*, 2007; Lau *et al.*, 2009; Saito *et al.*, 2009). The subcellular localization of PIWI-clade proteins also differs. Piwi is nuclear-localized in somatic and germline cells (Cox *et al.*, 1998; Cox *et al.*, 2000; Saito *et al.*, 2006; Brennecke *et al.*, 2007), whereas Aub and AGO3 express only in germ cells, where they localize to perinuclear, cytoplasmic structures called “nuage” (French for “cloud”) (Brennecke *et al.*, 2007; Gunawardane *et al.*, 2007; Lim & Kai, 2007; Nishida *et al.*, 2007).

In contrast to other small RNA classes (miRNAs and siRNAs), piRNA biogenesis is independent of Dicer (Vagin *et al.*, 2006; Houwing *et al.*, 2007). The main sources of piRNAs are either discrete genomic loci, named piRNA clusters, or transcripts of active transposons (Aravin *et al.*, 2006; Girard *et al.*, 2006; Brennecke *et al.*, 2007). In *Drosophila*, piRNA clusters mostly contain truncated transposon fragments that are unable to transpose, and may serve as graveyards to provide a genetic memory of previous transposon activity (Brennecke *et al.*, 2007). Cytologically, piRNA clusters often reside in pericentromeric or subtelomeric heterochromatin (Brennecke *et al.*, 2007). Although the genomic location and overall organisation of clusters is conserved between related species, their sequence content is not (Malone *et al.*, 2009). Some piRNA clusters, like the follicle-cell-specific *flamenco* (*flam*) cluster, are unidirectional (i.e. transcribed from one genomic strand), while others produce piRNAs from both strands (bidirectional cluster), such as the germline-specific *42AB* cluster (Brennecke *et al.*, 2007; Malone *et al.*, 2009). Genetic data suggest that cluster transcription originates from single promoters, with transcripts of extreme size (up to 250 kb) synthesized (Brennecke *et al.*, 2007).

The production of mature piRNAs from primary transcripts is poorly understood. Although primary piRNA biogenesis begins with cluster transcription, the factors involved in regulating and initiating cluster transcription remains largely unknown. A recent report implicated the chromatin status as a major regulator of piRNA transcription, with the histone 3 lysine 9 (H3K9) methyltransferase Eggless (Egg) shown to be required for piRNA production (Rangan *et al.*, 2011). Egg converts dimethylated H3K9 (H3K9me2) to

trimethylated H3K9 (H3K9me3), which serves as a repressive mark that could initiate heterochromatin formation via recruitment of heterochromatin protein 1 (HP1) to piRNA clusters (Rangan *et al.*, 2011). The link between heterochromatin and piRNA cluster transcription is further strengthened by the characterization of Rhino (Rhi), a HP1 homologue (also recognizing H3K9me3), and Cutoff (Cuff), a germ cell-specific protein interacting with Rhi, in promoting the transcription of the bidirectional *42AB* cluster (Chen *et al.*, 2007b; Klattenhoff *et al.*, 2009; Pane *et al.*, 2011). However, the molecular mechanism by which H3K9 methylation regulates cluster transcription remains unknown.

A current model, mostly built upon data from the somatic piRNA pathway in *Drosophila*, proposes that primary processing of piRNAs takes place at specific sites in the cytoplasm (Fig. 1-3). Following cluster transcription in the nucleus, either full-length cluster transcripts or pre-processed intermediates are exported to the cytoplasm through unknown machineries, where they are cleaved into piRNA intermediates by an unidentified endonuclease specific for single-stranded, unstructured RNAs. A single precursor transcript is thereby parsed into several piRNA intermediates, containing mature 5' ends but possessing elongated 3' tails. Zucchini (Zuc), a homolog of phospholipase D and a putative nuclease, has been implicated in primary piRNA processing (Pane *et al.*, 2007; Malone *et al.*, 2009; Saito *et al.*, 2009; Haase *et al.*, 2010; Olivieri *et al.*, 2010; Saito *et al.*, 2010), and is a promising candidate for 5' end formation. However, its precise molecular function remains a matter of debate.

In somatic cells of the fly ovary, these piRNA intermediates are then loaded into the sole PIWI-clade protein, Piwi, which is predicted to take place at specific sites called Yb-bodies (Szakmary *et al.*, 2009; Olivieri *et al.*, 2010; Saito *et al.*, 2010; Qi *et al.*, 2011). The loading process requires the proteins Fs(1)Yb (Yb), Sister of Yb (SoYb), Vreteno (Vret), and Armitage (Armi) (Cook *et al.*, 2004; Haase *et al.*, 2010; Olivieri *et al.*, 2010; Saito *et al.*, 2010; Handler *et al.*, 2011; Qi *et al.*, 2011; Zamparini *et al.*, 2011). Piwi-associated piRNAs show a prominent nucleotide bias for 5' uridine (1U bias) (Saito *et al.*, 2006; Brennecke *et al.*, 2007; Yin & Lin, 2007), but it is unclear whether this bias stems from upstream processing or is a consequence of preferential loading. Once bound to Piwi, piRNA-intermediates are further trimmed at the 3' end by an enzymatic activity yet to be uncovered (Kawaoka *et al.*, 2011), and subsequently methylated at their 3' termini by Hen1 to yield mature Piwi-RISC (Horwich *et al.*, 2007; Saito *et al.*, 2007). Upon maturation, Piwi-RISC enters the nucleus to carry out its effector function (Saito *et al.*, 2009; Saito *et al.*, 2010; Klenov *et al.*, 2011), although the underlying transport factors and effector mechanisms are yet to be identified. The nuclear localization of Piwi is dispensable for piRNA loading, as amino-terminal truncated Piwi lacking the nuclear

localization signal is still loaded with piRNAs, supporting the hypothesis that piRNA loading occurs in the cytoplasm (Saito *et al.*, 2009; Saito *et al.*, 2010).

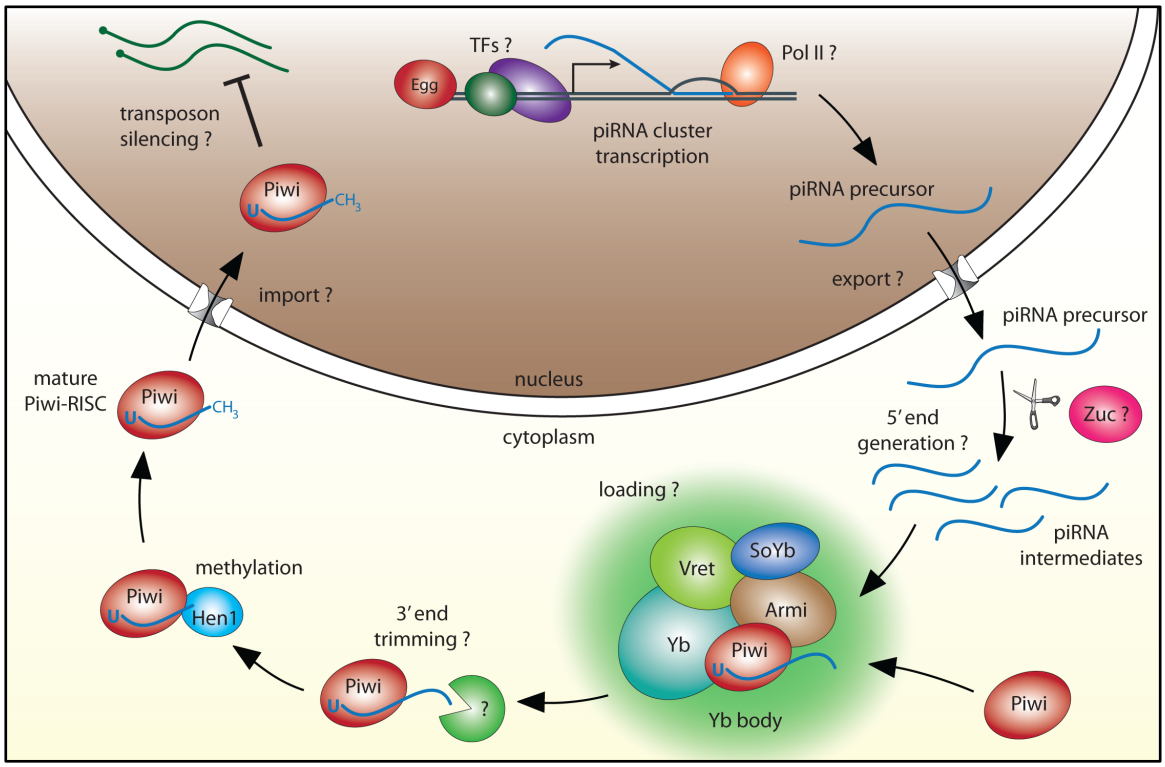


Fig. 1-3. Model for primary piRNA biogenesis in somatic tissues. PiRNA cluster transcription is promoted by unknown transcription factors and heterochromatin. The precursors are exported to the cytoplasm, where they are then processed into several piRNA intermediates and loaded into Piwi at Yb-bodies aided by additional factors. Following loading, piRNAs are matured by trimming through an unknown exonucleolytic activity and methylated by Hen1. Mature Piwi-RISC enters the nucleus to carry out transposon silencing.

Primary piRNA biogenesis in germline cells is thought to follow similar routes, with primary piRNAs loaded into Piwi and Aub, but not AGO3 (Brennecke *et al.*, 2007; Malone *et al.*, 2009). In contrast to somatic cells, piRNA production and/or loading in germ cells is predicted to take place at nuage granules, which is paralleled by the replacement of the soma-specific Yb protein with two germline-expressed homologues, SoYb and Brother of Yb (BoYb) (Handler *et al.*, 2011). Despite these differences, primary piRNA biogenesis likely occurs through similar mechanisms and uses a similar set of factors (e.g., Vret, Armi, Zuc, SoYb, BoYb, Hen1).

Although primary piRNAs enable animals to detect active transposon threats and initiate their repression, eminent transposition events necessitate an amplification of silencing triggers, especially in germline cells where transposable elements are more active. Towards this end, an elegant, RdRP-independent amplification mechanism, the ping-pong cycle, has evolved in flies with similar concepts found in mice (Aravin *et al.*,

2007; Brennecke *et al.*, 2007, Gunawardane *et al.*, 2007; Aravin *et al.*, 2008; Kuramochi-Miyagawa *et al.*, 2008). The PIWI-clade proteins themselves are integral components of the ping-pong cycle, as they create the 5' ends of future secondary piRNAs (Brennecke *et al.*, 2007, Gunawardane *et al.*, 2007; Li *et al.*, 2009). The ping-pong model was first proposed based on observations in flies, where deep sequencing of RNAs immunoprecipitated from the three PIWI-clade proteins revealed prominent biases in their sequence composition and orientation to transposons (Brennecke *et al.*, 2007, Gunawardane *et al.*, 2007). It was shown that AGO3 associates predominantly with piRNAs in sense orientation to transposons, whereas Aub- and Piwi-bound sequences were mainly antisense to mobile elements. Furthermore, piRNAs associated with Aub and Piwi showed a strong 1U bias, while AGO3-bound sequences featured a preference for an adenine nucleotide at position 10 (counting from the 5' end), also known as 10A bias. Intriguingly, Aub- and AGO3-associated sequences exhibited a prominent overlap of the first ten nucleotides (counting from the 5' end of each piRNA), with perfect sequence complementarity.

To explain these observations, the ping-pong model (Fig. 1-4) proposes that Aub complexes, containing piRNAs antisense to actively transposing elements, detect and cleave transposon transcripts through sequence complementarity (Brennecke *et al.*, 2007, Gunawardane *et al.*, 2007). The cleavage event generates a sense-orientated fragment containing the 5' end of the future secondary piRNA, which is then loaded into AGO3. The AGO3-bound fragment is further trimmed at the 3' end (by an unknown exonuclease termed "trimmer") and then methylated by Hen1 (Horwich *et al.*, 2007; Saito *et al.*, 2007), resulting in mature AGO3-RISC with associated secondary piRNA. Mature AGO3-complexes in turn cleave antisense transcripts derived from piRNA clusters, which also function as relay stations in the amplification loop (in addition to serving as source of primary piRNAs), to generate additional piRNA precursors that associate with Aub and, following maturation, resemble the primary silencing triggers. These piRNA-Aub-complexes perpetuate the ping-pong cycle, thereby efficiently producing and amplifying silencing triggers against eminent threats, while consuming or destroying their transcripts (Brennecke *et al.*, 2007, Gunawardane *et al.*, 2007). Mutant studies in *Drosophila* lacking specific piRNA pathway components (e.g., *ago3*, *qin* or *aub*) further supported the ping-pong model (Li *et al.*, 2009; Malone *et al.*, 2009; Zhang *et al.*, 2011), and efforts – including those presented within this thesis – are underway to uncover additional factors required in the amplification cycle.

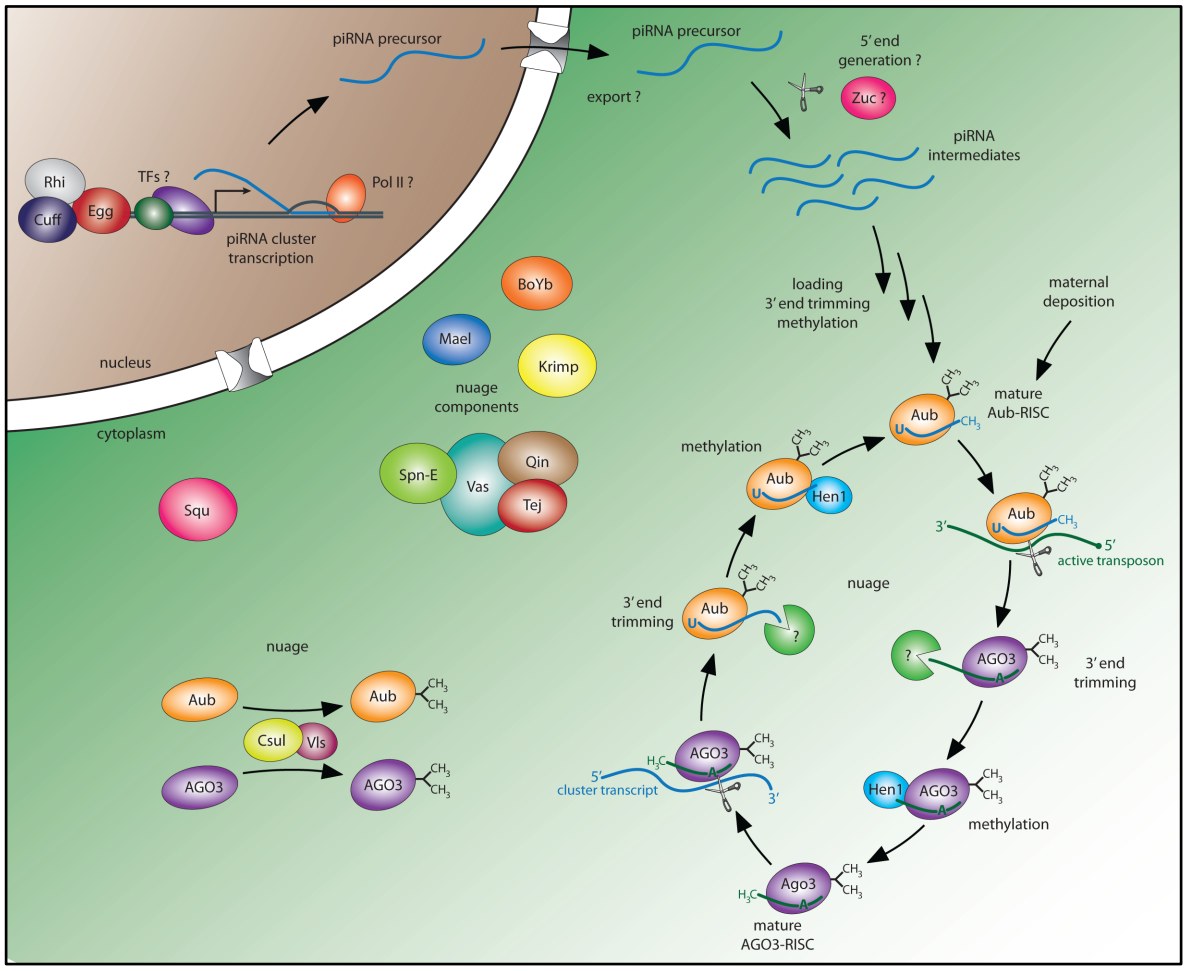


Fig. 1-4. The ping-pong cycle and secondary piRNA generation in the germline. Aub is primed with cluster-derived piRNAs antisense to transposons originating from primary biogenesis or maternal deposition. Aub detects and cleaves active transposons, generating a cleavage fragment that is loaded into AGO3, trimmed, and methylated to yield a mature secondary piRNA. AGO3 associated with a sense-orientated piRNA cleaves additional antisense cluster transcripts, which are incorporated into Aub and re-enter the cycle. Ping-pong takes place at nuage, with several factors playing important roles. Symmetrical dimethylarginine modifications of Aub and AGO3 are critical for proper transposon silencing and mediated by Csul/Vls complexes. For an improved overview, primary biogenesis is not shown.

The ping-pong amplification loop is not only important for the development of germ cells of the current generation, but also of the subsequent offspring. Maternally loaded Aub and Piwi complexes deposited during oogenesis have been shown to localize to future germ cells of the developing embryo (Brennecke *et al.*, 2007; Brennecke *et al.*, 2008; Malone *et al.*, 2009), where Aub-bound maternal piRNAs can feed into the ping-pong cycle (Brennecke *et al.*, 2008; Malone *et al.*, 2009). The importance of maternally deposited piRNAs can be seen in the phenomena called “hybrid dysgenesis” (Kidwell *et al.*, 1977; Brennecke *et al.*, 2008). Female offspring flies lacking maternal piRNAs against recently acquired transposons are incapable of silencing these mobile elements, despite possessing an intact primary piRNA biogenesis pathway. As a consequence of insufficient transposon silencing, these females feature

severely reduced fertility, and are often completely sterile (Brennecke *et al.*, 2008; Khurana *et al.*, 2011). In contrast, F1 females from reciprocal crosses, which are genetically identical but received maternal piRNAs against acquired transposons, efficiently silence mobile elements and are fertile. Although detailed molecular mechanisms remain elusive, it has been proposed that maternal piRNAs provide initial silencing triggers or kick-start the ping-pong cycle to produce piRNAs at levels sufficient for robust silencing.

The localization of piRNA-primed Aub and AGO3 to nuage implies this structure as the site of active ping-pong amplification and transposon silencing (Lim & Kai, 2007). The proper assembly and function of the nuage requires several proteins, such as Vasa (Vas), Spindle-E (Spn-E), Maelstrom (Mael), Krimper (Krimp), Tejas (Tej), BoYb, Qin (also known as Kumo), and Tudor (Tud) (Boswell & Mahowald, 1985; Gillespie & Berg, 1995; Clegg *et al.*, 1997; Findley *et al.*, 2003; Vagin *et al.*, 2004; Lim & Kai, 2007; Malone *et al.*, 2009; Patil & Kai, 2010; Handler *et al.*, 2011; Zhang *et al.*, 2011; Anand & Kai, 2012). Various mutants were identified that compromise Aub and AGO3 localization patterns, but details on their molecular functions and interactions with the piRNA pathway have yet to be elucidated. Intriguingly, Tudor domain proteins are prominent amongst regulators of Aub and AGO3 localization (Lim & Kai, 2007; Kirino *et al.*, 2009; Nishida *et al.*, 2009; Kirino *et al.*, 2010; Patil & Kai, 2010; Handler *et al.*, 2011; Zamparini *et al.*, 2011). They can interact either directly or via protein modifications with piRNA pathway components, particularly through symmetrical dimethylarginine (sDMAs) at their amino-termini of PIWI-clade proteins, and potentially serve as scaffolding for larger protein complexes. The sDMA modification is highly conserved in PIWI-clade proteins across species (Reuter *et al.*, 2009; Vagin *et al.*, 2009; Huang *et al.*, 2011), and is promoted in *Drosophila* by the arginine N-methyltransferase Capsuleen (Csul) and its co-factor Valois (Vls) (Anne & Mechler, 2005; Anne *et al.*, 2007; Kirino *et al.*, 2009; Nishida *et al.*, 2009; Kirino *et al.*, 2010).

Additional factors were reported to affect the piRNA pathway, though their roles are yet to be fully characterized. As described previously, mutants of *rhi* and *cuff* impair transcription of piRNA clusters, leading to an absence of primary piRNA biogenesis, and mislocalization of Aub and AGO3 from nuage (Chen *et al.*, 2007b; Klattenhoff *et al.*, 2009; Pane *et al.*, 2011). Unlike other protein factors described, the putative RNase H, Squash (Squ), does not appear to impact biogenesis of primary or secondary piRNAs, but mutants display defects in transposon suppression (Pane *et al.*, 2007; Malone *et al.*, 2009; Haase *et al.*, 2010), suggesting a potential role as component of the piRNA effector complex responsible for regulating mobile element activity.

1.5 Thesis aims

Small RNAs play crucial roles in *Drosophila*, regulating gene expression (miRNAs and siRNAs) and controlling transposable elements (piRNAs and siRNAs). As outlined in the introduction, many components of the miRNA and siRNA pathways have been identified, leading to a model that these pathways have little overlap. In the first part of my thesis, I aim to provide insights into whether the miRNA and siRNA pathways, particularly at the biogenesis and sorting level, operate in distinct networks. Prior studies have uncovered an unexpected dependence of a subset of endo-siRNAs on the dsRBD protein Loqs (Czech *et al.,* 2008; Okamura *et al.*, 2008a). Loqs was previously only implicated in miRNA biogenesis, with additional isoforms reported that have unresolved functions (Förstemann *et al.*, 2005; Jiang *et al.*, 2005; Saito *et al.*, 2005; Park *et al.*, 2007). Recognizing the gap in our knowledge, my first aim is to identify and characterize the specific Loqs isoform(s) involved in the biogenesis of endo-siRNAs.

Although earlier models of miRNA and siRNA biogenesis suggest that they are directly coupled to their respective downstream effectors, conflicting studies have claimed that precursor intermediates are redistributed after biogenesis based on the characteristics of their dsRNA intermediates (Förstemann *et al.*, 2007; Tomari *et al.*, 2007). Due to a limited number of sequences analysed, neither hypothesis has been conclusive. Therefore, my second aim is to use high throughput sequencing technologies to profile the whole repertoire of AGO1- and AGO2-bound small RNAs. This will allow me to identify and extract novel determinants that control the sorting of small RNAs to distinct Argonaute proteins, and aid in the development of improved artificial silencing triggers for research and potential therapeutics.

In contrast to miRNAs and siRNAs, the piRNA pathway is less well understood, with many components involved in the biogenesis of piRNAs yet to be uncovered. In the second part of my thesis, I aim to expand our knowledge of piRNA-mediated transposon silencing by carrying out an innovative, transcriptome-wide RNAi screen in germ cells of the female ovary. This unbiased approach will reveal novel genes that participate in piRNA production or piRNA-mediated effector mechanisms, thus providing new insights into the germline piRNA pathway as evidenced by the identification and characterization of *shutdown* as a new piRNA biogenesis factor.

2 Results

The results presented in this chapter summarize my thesis research and is divided into two parts for clarity. The first part will focus on discoveries made on the miRNA and endo-siRNA pathways, where I report the identification of a novel isoform of Loquacious that functions in the biogenesis of endo-siRNAs. I also uncovered the association of miR* strands with AGO2, the principal host of siRNAs, and revealed that partitioning of small RNAs between AGO1 and AGO2 complexes depends on intrinsic duplex properties. Based on these results, a hierarchy of determinants governing small RNA sorting in flies was extracted, and applied in designing novel shRNA-based tools for artificial gene silencing. The second part describes my contributions towards a better understanding of the *Drosophila* piRNA pathway, particularly the biogenesis and effector mechanisms of piRNAs. First, I report the characterization of *shutdown* as a new biogenesis factor. I will also describe my efforts towards uncovering missing piRNA pathway components via a transcriptome-wide RNAi screen in germ cells of the *Drosophila* ovary.

My work resulted in the papers stated in the publication list (individual manuscripts can be found in the appendix).

2.1 Part I: The AGO2 pathway and sorting of small RNAs

2.1.1 Endo-siRNA biogenesis requires a specific isoform of Loqs

Background

MicroRNAs are ubiquitously expressed small RNAs that are 22- to 23-nt in length. In *Drosophila*, miRNAs are processed by Drosha/Pasha and Dcr1/Loqs complexes, and associate with AGO1 to control gene expression (Lee *et al.*, 2003; Denli *et al.*, 2004; Lee *et al.*, 2004b; Förstemann *et al.*, 2005; Jiang *et al.*, 2005; Saito *et al.*, 2005; Bushati & Cohen, 2007; Eulalio *et al.*, 2008). Recently, several types of endogenous small interfering RNAs (endo-siRNAs) were discovered, which are ~21-nt in length, bind to AGO2, and function in gene regulation and transposon silencing (Czech *et al.*, 2008; Ghildiyal *et al.*, 2008; Kawamura *et al.*, 2008; Okamura *et al.*, 2008a). The processing of endo-siRNAs derived from structured loci and viruses requires Dcr2, but not its co-factor R2D2 (Liu *et al.*, 2003; Lee *et al.*, 2004b; Czech *et al.*, 2008; Okamura *et al.*, 2008a). Surprisingly, the production of these two endo-siRNA classes depends on the dsRNA-

binding domain (dsRBD) protein Loqs, previously only implicated as a co-factor of Dcr1 in miRNA processing (Förstemann *et al.*, 2005; Jiang *et al.*, 2005; Saito *et al.*, 2005; Czech *et al.*, 2008; Okamura *et al.*, 2008a). These findings were also supported by physical interactions of Loqs with both fly Dicer proteins, Dcr1 and Dcr2, as shown by quantitative proteomics (Czech *et al.*, 2008). Previous studies reported three distinct Loqs isoforms (Loqs-*RA*, *-RB*, and *-RC* transcripts, with the corresponding protein variants Loqs-PA, -PB, -PC) with specific expression patterns (Förstemann *et al.*, 2005; Jiang *et al.*, 2005). Loqs-*RA* is the major isoform in males, with Loqs-*RB* transcripts dominating in ovaries, and Loqs-*RC* only detected in S2 cells. Genetics demonstrated that isoform Loqs-PB alone is responsible for miRNA biogenesis (Park *et al.*, 2007), while the functions of Loqs-PA and Loqs-PC remain elusive.

Given the unexpected role of Loqs in the production of endo-siRNAs, its interaction with Dcr1 and Dcr2, and the existence of distinct transcript/protein variants, we sought to identify Loqs isoforms that are required for endo-siRNA biogenesis. This work was published as Zhou, Czech *et al.*, 2009 in *RNA*.

Results

To identity the Loqs variants that participate in interactions with Dicer proteins, we re-analysed our quantitative proteomic data (Czech *et al.*, 2008), and detected isoform-specific peptides for Loqs-PA and Loqs-PB, but not Loqs-PC. We also found peptides corresponding to a novel isoform (Loqs-PD) recently discovered by Förstemann and colleagues (Fig. 2-1a) (Hartig *et al.*, 2009). All four Loqs isoforms share an identical amino-terminal region that includes two dsRNA-binding domains. Compared to Loqs-PA (44.9 kDa) and Loqs-PB (50.1 kDa), isoforms Loqs-PC (41.1 kDa) and Loqs-PD (38.5 kDa) are truncated and therefore lack the third carboxy-terminal dsRNA-binding domain (Fig. 2-1b).

Next, we determined if individual Loqs isoforms show preferential interactions with specific Dicer proteins. Co-immunoprecipitation experiments were carried out in S2 cells, where a tagged protein complex is immunopurified and probed by Western blot for specific components. We were able to show strong interactions between FLAG-tagged Dcr2 and Loqs-PD (Fig. 2-1c), which was identified based on size (38.5 kDa) using an antibody specific to the amino-terminus of Loqs that is shared by all isoforms. In contrast, Dcr1 showed a strong interaction with TAP-tagged Loqs-PB isoform (Fig. 2-1d), consistent with previous reports (Förstemann *et al.*, 2005; Park *et al.*, 2007). These interactions were resistant to RNase treatment and to over-expression of endo-siRNA or miRNA precursors, thus suggesting direct protein-protein interactions independent of

RNA substrates. These results point towards specific roles of Loqs-PB and Loqs-PD in the production of miRNAs and siRNAs, respectively.

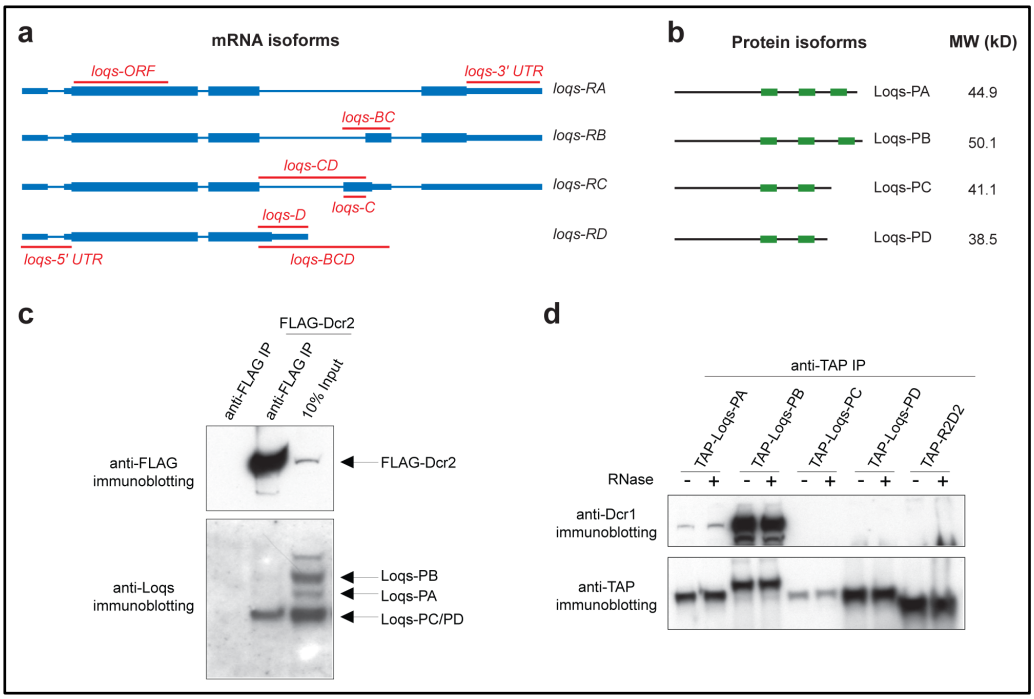


Fig. 2-1. Overview of Loqs isoforms and interaction of Loqs variants with Dicer proteins. (a) A scheme of the genomic structure (introns, exons, UTRs) of the different Loquacious isoforms is shown. Regions targeted by isoform-specific dsRNAs are indicated in red. **(b)** Schematic drawing showing domain structures (green boxes indicate dsRNA-binding domains) and molecular weights of distinct Loqs isoforms. **(c)** Immunoprecipitates of FLAG-tagged Dcr2 were subject to Western blots using an antibody against the amino-terminus of endogenous Loqs, and revealed specific interaction between FLAG-Dcr2 and Loqs-PD. **(d)** Coimmunoprecipitation of endogenous Dcr1 with TAP-tagged Loqs isoforms demonstrating strong interaction between Loqs-PB and Dcr1, independent of RNase treatment. (Image and legend modified from: Zhou, Czech *et al.*, 2009)

We next investigated the contribution of individual Loqs isoforms to the biogenesis of miRNAs and endo-siRNAs using isoform-specific knockdowns in S2 cells (Fig. 2-1a). Western blots confirmed reduced levels of the targeted proteins upon dsRNA-mediated depletion of specific isoforms (either alone or in combination) (Fig. 2-2a). The impact of isoform-specific knockdowns on steady-state levels of a miRNA (*miR-bantam*) and an endo-siRNA of structured origin (*esi-2.1*) was assessed by Northern blots (Fig. 2-2b). Coordinated depletion of all Loqs isoforms resulted in severely reduced endo-siRNA levels, while miRNA precursors accumulated. Knockdown of individual isoforms or the combinatorial depletion of several isoforms showed a strong dependence of endo-siRNAs on Loqs-PD, while reduced levels of Loqs-PB correlated with compromised miRNAs biogenesis. Knockdown of Loqs-PA or Loqs-PC did not impact either small RNA class.

To confirm our findings, we re-expressed individual RNAi-resistant isoform variants following dsRNA-mediated knockdown of all four Loqs isoforms in cell culture. Endo-siRNA levels were only restored upon Loqs-PD expression, whereas only Loqs-PB expression resulted in normal miRNA levels. Interestingly, R2D2 expression failed to restore endo-siRNA levels in *loqs* knockdowns, confirming previous results (Czech *et al.*, 2008; Okamura *et al.*, 2008a). Consistent with results from cell culture, expression of Loqs-PA in homozygous mutant *loqs*^{f00791} flies failed to restore miRNA and siRNA processing, while those expressing Loqs-PB had normal miRNA levels, but were still defective in endo-siRNA biogenesis.

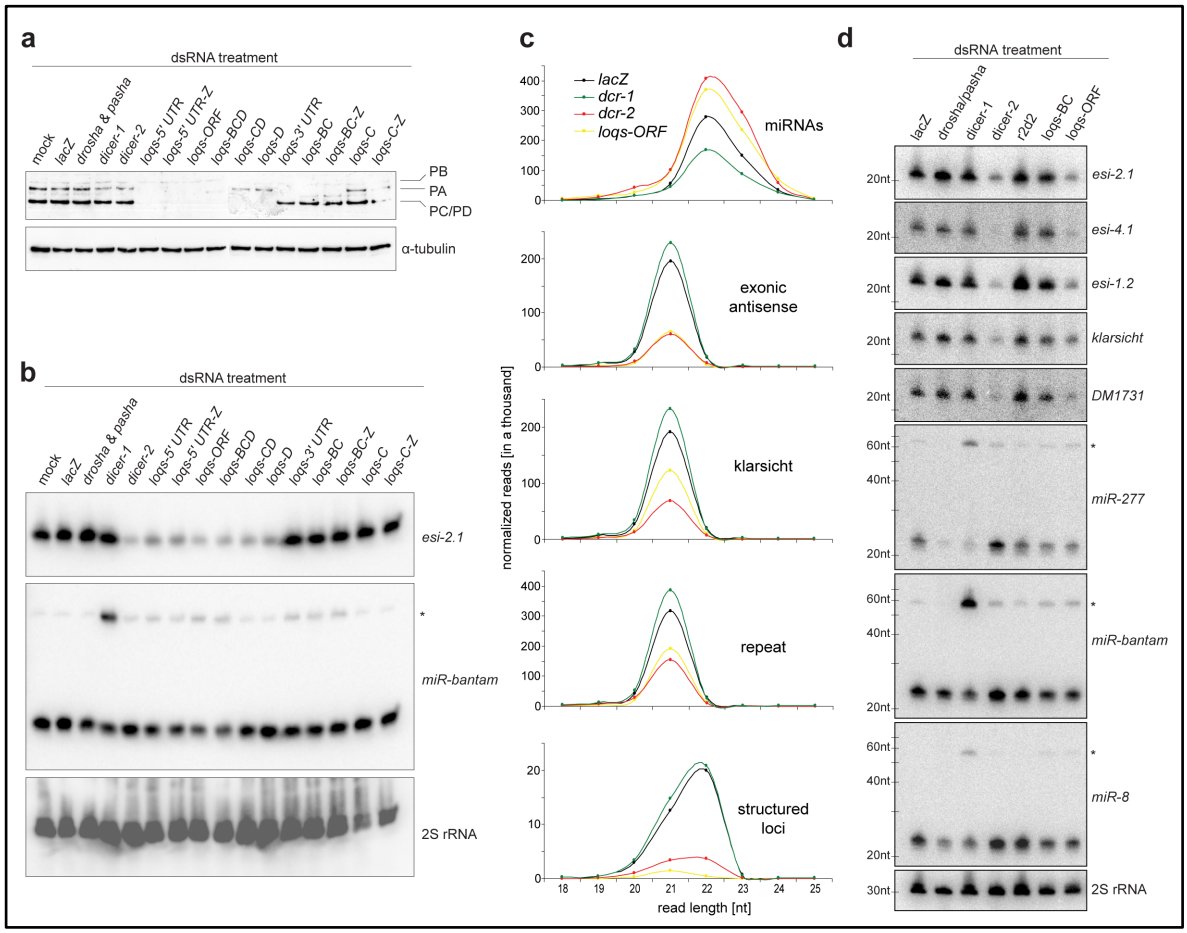


Fig. 2-2. Isoform Loqs-PD is required for the production of several endo-siRNA classes. (a) Depletion of specific isoforms upon targeting by the indicated dsRNAs is shown by Western blots ("-Z" marks *lacZ*-fused carrier dsRNA to increase knockdown efficiency). α-tubulin served as loading control. **(b)** Northern blots showing *esi-2.1* and *miR-bantam* levels in cells treated with indicated dsRNAs. 2S rRNA served as loading control. **(c)** Size profiles of miRNAs and different endo-siRNA types from cells treated with indicated dsRNAs are shown. **(d)** Levels of miRNAs and different types of endo-siRNAs upon depletion of indicated factors were probed by Northern blots. 2S rRNA served as loading control. (Image and legend modified from: Zhou, Czech *et al.*, 2009)

To test the impact of Loqs on endo-siRNA classes other than those derived from viruses and structured loci, we depleted Loqs and other siRNA/miRNA pathway components (*dcr1*, *dcr2*) in S2 cells using dsRNAs, and prepared small RNA libraries.

Following bioinformatic removal of presumed degradation products, libraries were normalized, and sequenced small RNA reads were distributed based on annotations to different categories. Size profiles for miRNAs, endo-siRNAs from (1) structured loci, (2) repeats, (3) genes, and (4) the *klarsicht* locus were obtained by plotting the cloning counts against the read length (Fig. 2-2c). Endo-siRNAs were predominantly 21-nt in length, with those derived from structured loci being slightly longer (~22-nt). MiRNAs showed the expected size distribution of 22- and 23-nt. Knockdown of *dcr1* severely reduced miRNA levels without changes in other small RNA types, whereas *dcr2* knockdown resulted in strongly reduced levels of all endo-siRNA classes, but not miRNAs. Depletion of all Loqs isoforms significantly affected endo-siRNAs derived from structured loci, repeats, and genes, while levels of *klarsicht* siRNAs were only moderately reduced. These results were validated using Northern blots (Fig. 2-2d), with knockdown of *dcr2* and *loqs* decreasing steady-state levels of endo-siRNAs from structured loci (*esi-2.1*, *esi-4.1*, *esi-1.2*), and from the transposon *DM1731*, while *klarsicht* siRNAs showed moderate reduction. Knockdowns that only depleted Loqs-PB, did not affect the abundance of any siRNA, but showed the expected miRNA precursor accumulation.

To assess whether Loqs depletion causes a global or element-specific reduction of transposon-derived siRNAs, we generated heat maps showing the relative abundance of endo-siRNAs matching to individual transposable elements. As predicted, knockdown of *dcr1* (involved in the miRNA pathway) showed no significant changes in siRNA levels, while depletion of Dcr2 caused a severe siRNA reduction for the majority of elements. Knockdown of all Loqs isoforms had similar, but slightly weaker effects, with levels of several transposon siRNAs being reduced, suggesting a global requirement of Loqs for siRNA-mediated transposon silencing.

Conclusions

Our proteomic analyses identified a novel isoform of the dsRBD protein Loqs, Loqs-PD, which predominantly interacts with Dcr2, and confirmed previously reported interactions between Dcr1 and a second isoform, Loqs-PB (Förstemann *et al.*, 2005; Park *et al.*, 2007). Molecular and genetic studies demonstrate that these isoforms act in different small RNA biogenesis pathways through binding to their preferential protein partners. Dcr1/Loqs-PB complexes process miRNAs, whereas Dcr2/Loqs-PD complexes produce siRNAs. The functions of two additional isoforms, Loqs-PA and Loqs-PC remain elusive, with Loqs-PC not detectable at the protein level. Deep sequencing and Northern blots reveal that Loqs-PD impacts the biogenesis of various endo-siRNA classes, in

addition to viral siRNAs and those derived from structured loci as reported previously (Czech *et al.*, 2008; Okamura *et al.*, 2008a). Taken together, our results add complexity to the landscape of dsRNA-binding domain proteins in small RNA biogenesis, and highlight the importance of co-factors to substrate specificity and production of different small RNA types.

2.1.2 Hierarchical rules for Argonaute loading in *Drosophila*

Background

In *Drosophila*, miRNAs and siRNAs originate from dsRNA precursors through processing by RNase III family proteins, and associate with two AGO-clade proteins: AGO1 and AGO2. MicroRNA duplexes are processed by Drosha/Pasha and Dcr1/Loqs-PB complexes (Lee *et al.*, 2003; Denli *et al.*, 2004; Lee *et al.*, 2004b; Förstemann *et al.*, 2005; Jiang *et al.*, 2005; Saito *et al.*, 2005), and following loading into AGO1, the miR strands are selectively stabilized without further modification at their 3' termini (Horwich *et al.*, 2007; Saito *et al.*, 2007). AGO1 primed with a mature miRNA causes repression of targets through cleavage-independent mechanisms. In contrast, siRNA duplexes are produced from long dsRNAs by Dcr2/Loqs-PD complexes (Hartig *et al.*, 2009; Zhou *et al.*, 2009), and loaded into AGO2 by Dcr2/R2D2 complexes (Liu *et al.*, 2003; Lee *et al.*, 2004b; Tomari *et al.*, 2004). Only one strand of the duplex is retained in AGO2 (termed guide strand), which is subsequently methylated at its 3' end by Hen1 (Horwich *et al.*, 2007; Saito *et al.*, 2007), and directs AGO2 to cleaving target transcripts with high sequence complementarity.

As outlined above, miRNAs and siRNAs share similar processing intermediates (small RNA duplexes) but differ fundamentally in their biological outputs. As a result, pairing of these small RNA classes with specific Argonaute proteins must be tightly controlled through a process known as "sorting". An early model suggested coupled processing ('dicing') and loading with small RNA intermediates directly assembled into Argonaute effector complexes upon biogenesis. However, exceptions like *miR-277*, which is produced by Dcr1/Loqs-PB but loaded into AGO2, point to more complex mechanisms (Förstemann *et al.*, 2007). Subsequent models have proposed that duplexes with diverging properties must be directed towards specific Argonaute proteins, with perfect match duplexes being biased towards AGO2 and duplexes with mismatches preferring AGO1 (Tomari *et al.*, 2007). Furthermore, individual strands of each duplex have different probabilities of being retained in Argonaute complexes. For duplexes with perfect dsRNA character, strand selection was proposed to depend on thermodynamic

properties of duplexes, with the strand featuring the less stable 5' end being retained (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003).

Due to technical limitations, previous work relied on only a constricted number of sequences. With the availability of high throughput sequencing technologies, we can now profile the full repertoire of endogenous binding partners of AGO1 and AGO2. Using these data, determinants that contribute to the sorting behaviour of dsRNA-derived small RNAs were extracted, validated experimentally, and compared with existing sorting models. This work was published as Czech, Zhou *et al.*, 2009 in *Molecular Cell*.

Results

We generated a set of small RNA libraries to profile the association of dsRNA-derived small RNAs with specific *Drosophila* Argonaute proteins. Small RNAs (19- to 24-nt in size) were isolated and cloned from wild-type S2 cells using the standard cloning strategy ("standard"), and a protocol modified to enrich for small RNAs with 2'-*O*-methylated 3' termini ("oxidized") (Seitz *et al.*, 2008). Sequence reads were then categorized into miRNAs and five classes of endo-siRNAs, corresponding to their genomic origin (genes, structured loci, repeats, viruses, or regions without annotation). Sequences matching to miRNAs were then split into miR and miR* strands (Fig. 2-3a). The majority of reads in the standard library corresponded to endo-siRNA (62.6%), with the remaining sequences matching to miRNAs (37.4%), while the oxidized library showed a significant bias against miRNA (2.3%) compared to endo-siRNA (97.7%). Consistent with previous reports, endo-siRNA classes were enriched in the oxidized library (Chung *et al.*, 2008; Kawamura *et al.*, 2008; Okamura *et al.*, 2008a), with the exception of partially methylated virus-derived siRNAs (Aliyari *et al.*, 2008; Flynt *et al.*, 2009). As predicted by the lack of modified 3' termini, the miRNA fraction was significantly reduced by the modified cloning strategy. A comparison of ratios of miR and miR* strands between the standard (~33:1) and oxidized (~2:1) libraries revealed a relative enrichment of miR* strands by ~16-fold using the modified cloning protocol. To assess the relative association of small RNAs with AGO1 or AGO2, we calculated the ratios of sequence counts between standard and oxidized libraries for the 40 most abundant sequences (matching to miR or miR* strands or to endo-siRNAs from structured loci) (Fig. 2-3b), taking advantage of the 2'-*O*-methylation as a characteristic of AGO2-loaded sequences. We observed a bias of miR strands associating with AGO1, whereas ratios for miR* strands and endo-siRNAs from structured loci indicated enrichment in AGO2.

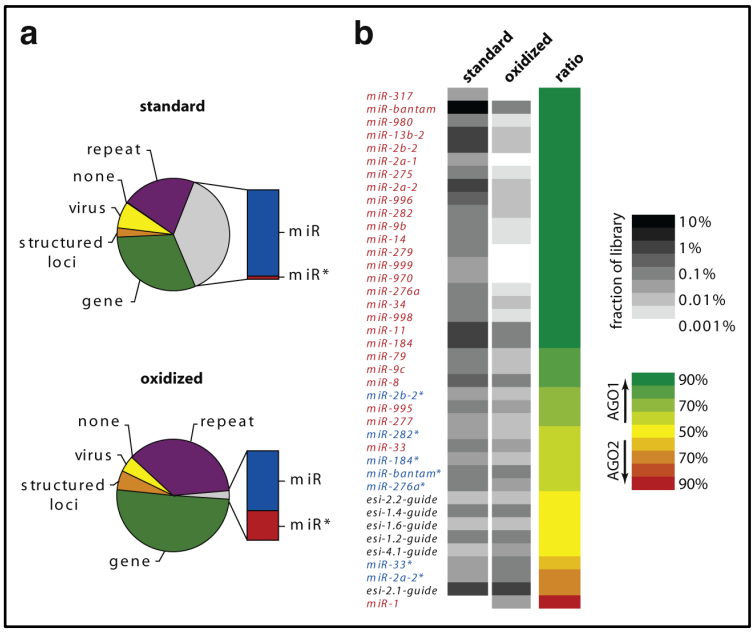


Fig. 2-3. MiR* strands have 2'-*O*-methylated 3' termini. (a) Cake diagrams showing the relative abundance of miRNAs (miR and miR* strands highlighted as bars) and indicated endo-siRNA classes in libraries from S2 cells. Results from standard (upper) and modified ("oxidized") cloning protocols are shown. **(b)** Relative abundances of miRs, miR*s and endo-siRNAs from structured loci are shown as heat maps (grey-scale), with the ratio between both libraries indicating biased association with AGO1 or AGO2 (green/red-scale). (Image and legend modified from: Czech, Zhou *et al.*, 2009)

We then directly analysed the small RNA populations by profiling RNAs from AGO1 and AGO2 immunoprecipitates in S2 cells. Using the same annotation criteria (Fig. 2-4a), we found ~98% of all AGO1-bound sequences matching to miRNAs, with the vast majority (~99%) of these corresponding to the miR strand, contradicting previous reports that detected significant miR* levels in AGO1 (Okamura *et al.*, 2008b). AGO2, in contrast, was predominantly associated with all classes of endo-siRNAs (~92%), with the remainder (~8%) matching to miRNAs. Sub-annotation revealed that almost 60% of those corresponded to miR* strands.

We confirmed our deep-sequencing results with Northern blots by probing for specific small RNAs isolated from AGO1 and AGO2 immunoprecipitates (Fig. 2-4b). Using β-elimination prior to PAGE, we were also able to determine if the 3' termini were methylated, as non-methylated RNAs were sensitive to the reaction and migrated as a lower molecular weight species. As expected, AGO1-bound RNAs were sensitive to β-elimination, and showed a strong signal for the miR strands of three miRNAs (*miR-bantam*, *miR-184*, *miR-276a*). AGO2-associated RNAs, in contrast, were completely resistant to β-elimination, and showed enrichment of the endo-siRNA, *esi-2.1*, as well as the corresponding miR* strands of the miRNAs investigated. Taken together, these results reveal an unexpected association of miR* strands with AGO2 that suggests bi-functionality of miRNA precursors.

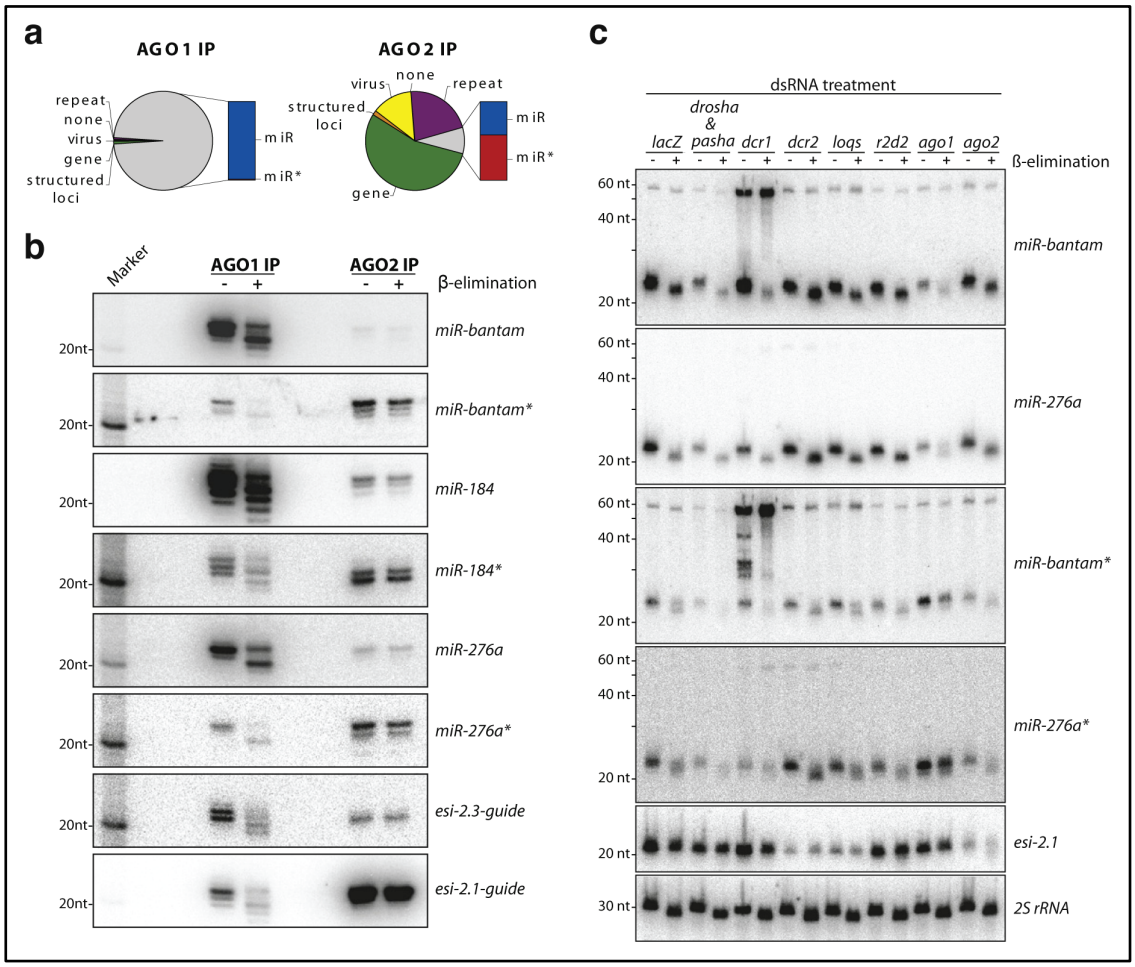


Fig. 2-4. MiR* strands preferentially associate with AGO2 and require the siRNA machinery for loading. (a) Cake diagrams showing the relative abundance of miRNAs (miR and miR* strands are highlighted as bar diagram) and indicated endo-siRNA classes in libraries from immunoprecipitates of AGO1 (left) and AGO2 (right) from S2 cells. **(b)** Northern blots of AGO1- and AGO2-bound RNAs from S2 cells probing the levels of the indicated miRs, miR*s and endo-siRNAs. Prior to gel electrophoresis, immunoprecipitated RNA were subjected to β-elimination (+) or untreated (-). **(c)** Northern blots probing for the levels of the indicated miRs, miR*s and endo-siRNAs. 2S rRNA served as loading control. Prior to gel electrophoresis, RNAs from the indicated RNAi knockdowns were subjected to β-elimination (+) or untreated (-). (Image and legend modified from: Czech, Zhou *et al.*, 2009)

We next analysed the processing and loading requirements of miR and miR* strands by depleting S2 cells of core components of the miRNA and siRNA pathways by RNAi. Using Northern blots, we probed the effects of our knockdowns on miRNA (both strands of *miR-bantam* and *miR-276a*) and endo-siRNA levels (*esi-2.1*), and assessed the modifications of the 3' termini with β-elimination prior to PAGE (Fig. 2-4c). Knockdown of canonical miRNA pathway components (Drosha & Pasha, Dcr1, Loqs) resulted in consistent reduction of miR and miR* strand levels, accompanied by slight precursor accumulation upon Dcr1 or Loqs depletion. As expected, endo-siRNA levels were not affected when miRNA factors were depleted. In contrast, knockdown of siRNA pathway components had differential effects on miR and miR* strands. RNAi against

dcr2 or *loqs* resulted in severely reduced endo-siRNA levels, but did not impact miR or miR* strand abundance. However, a knockdown of *dcr2* and *r2d2* resulted in miR* strands sensitive to β-elimination, indicating a lack of 2'-*O*-methylation. Depletion of AGO2 did not alter the levels of miR strands, but instead caused reduced levels of endo-siRNAs and miR* strands, with both now also sensitive to β-elimination. Intriguingly, knockdown of AGO1 had opposing consequences on miR and miR* strands: while miR strands were strongly reduced, a significant accumulation of miR* strands resistant to β-elimination was observed. Small RNA sequencing of *ago1* and *ago2* knockdowns confirmed these results.

With AGO2-bound miR* strands detected at comparable abundance to functional endo-siRNAs, we speculated that miR* strands could repress target transcripts. *Renilla* luciferase sensor constructs were generated containing target sites for either the miR or the miR* strand, with the binding sites either in perfect-match or bulged configuration. The reporter constructs were introduced into S2 cells in combination with knockdown of miRNA/siRNA pathway components. Depletion of Drosha caused consistent de-repression of miR and miR* sensors of two independent miRNAs (*miR-bantam* and *miR-276a*), with weaker effects observed upon Pasha and Dcr1 depletion. MiR* sensors with perfect target sites were de-repressed upon *ago2* knockdown, and consistent with their role in loading miR* strands, similar results were obtained for *dcr2* and *r2d2* knockdowns. Strikingly, RNAi against *ago1* enhanced the repression of perfect-match target sites for miR* strands, consistent with increased miR* levels previously observed by Northern blots. The miR* silencing abilities were confirmed *in vivo* using transgenic flies expressing similar sensor constructs (perfect-match or bulged target sites for miR or miR* strands located within the 3' UTR of an *EGFP* reporter) combined with clonal analysis in wing discs.

Our results have identified a strong bias of miR strands towards AGO1, while miR* strands and endo-siRNAs predominately associate with AGO2. We also observed a tendency of *miR-277* to be bound by AGO2, consistent with a previous report (Förstemann *et al.*, 2007). Given the observed bi-functionality of miRNA precursors, loading must rely on more complex parameters than previously presumed (Förstemann *et al.*, 2007; Tomari *et al.*, 2007). To investigate the underlying characteristics of specific AGO associations, we analysed the properties of small RNAs highly biased for either AGO1 or AGO2, and extracted positional nucleotide preference and base-pairing patterns (including mismatch distribution) within the duplexes (Fig. 2-5a). As previously observed for most miRNAs, AGO1-associated RNAs showed a strong nucleotide bias for terminal uracil (~90% contain 5' U). In contrast, sequences bound to AGO2 showed

moderate enrichment for a 5' cytosine (~50% of sequences show 5' C). Overall base pairing appears to be a minor determinant for sorting, as AGO1-biased duplexes showed only slightly higher mismatch frequencies than AGO2-associated sequences. Intriguingly, prominent base-pairing differences between AGO1 and AGO2-associated small RNAs were detected for the central duplex region. The strand loaded into AGO1 was often unpaired at positions 9 and 10, whereas more than 90% of AGO2-bound strands were paired. These patterns were observed regardless of origin of duplexes, whether from miRNAs and endo-siRNAs. Furthermore, the guide strand of *esi-2.3*, an endo-siRNA with central mismatches highly reminiscent of miRNA duplexes, was preferentially loaded into AGO1 instead of AGO2. Taken together, our results suggest that intrinsic duplex properties affect sorting and strand selection independent of upstream processing pathways.

The significance of the observed duplex properties for sorting and strand selection was evaluated through Argonaute loading assays (Fig. 2-5b). Artificially modified small RNA duplexes (based on *miR-276*a and siRNA *let-7*) were transfected into S2 cells, where they associated with Argonaute proteins. Subsequently, AGO1 and AGO2 complexes were immunoprecipitated, and the two strands of the artificial small RNAs were detected independently by Northern blots to determine their sorting behaviour. As expected, both strands of a duplex with perfect dsRNA character (siRNA *let-7*) were strongly bound by AGO2. Upon insertion of central bulges, both strands shifted towards loading into AGO1, with the strand carrying the bulge at position 9 and 10 (measured from 5' end) predominantly detected. The combination of unpaired terminal nucleotides (to change the thermodynamic asymmetry of duplexes) with central bulges highlighted the dominance of the central region over mismatches at the duplex ends. The wild-type *miR-276a* duplex showed similar patterns as observed from high throughput sequencing and Northern blots: *miR-276a* associates with AGO1, whereas *miR-276a** shows a preference for AGO2. Reversing the central mismatches of the *miR-276a* duplex resulted in the miR strand shifting towards AGO2 and the miR* strand preferring AGO1, while only minor changes were observed when terminal nucleotides were altered. Our results suggest that bulges and mismatches in the central duplex region have the most prominent effects on sorting, with positions 9 and 10 (from the 5' end) able to determine the fate of a duplex. Furthermore, central mismatches heavily impact strand selection, overriding the effects of thermodynamic asymmetry at the duplex ends.

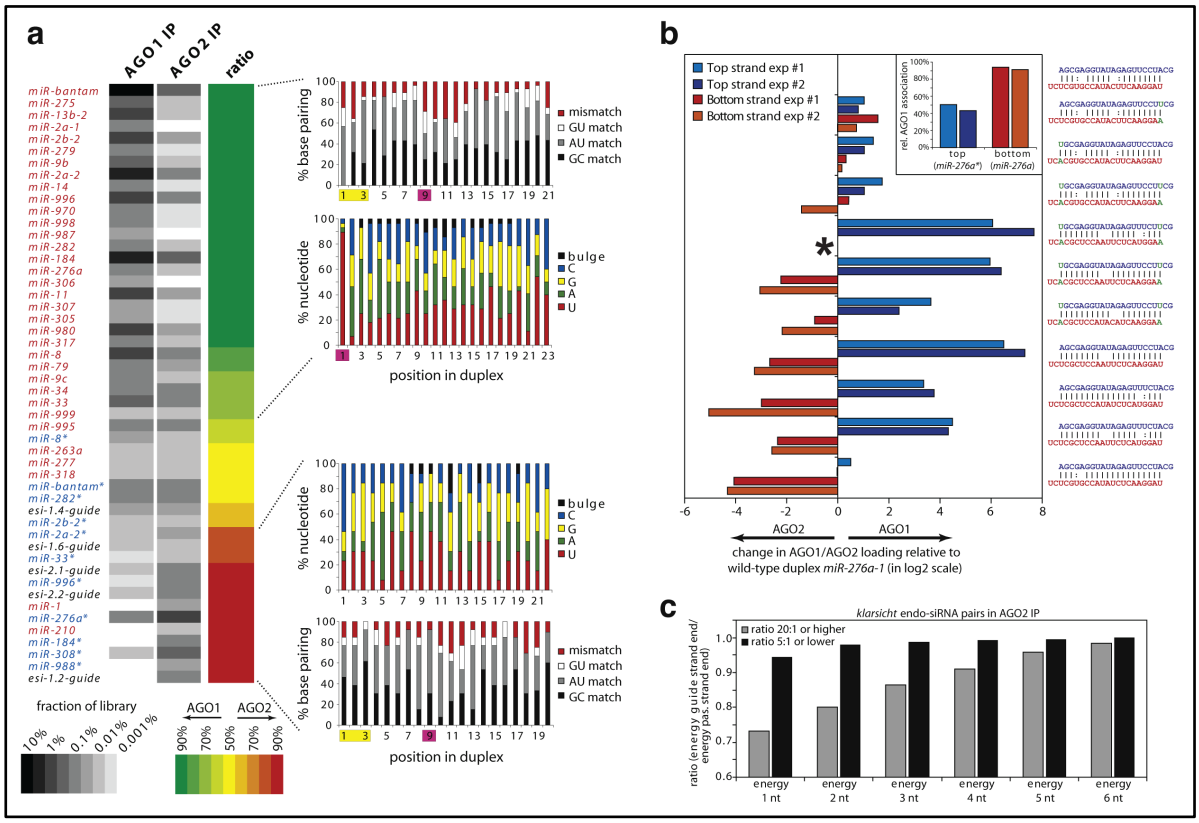


Fig. 2-5. Extraction and experimental validation of determinants for small RNA sorting. (a) Relative abundances of miR, miR* and endo-siRNAs in libraries from AGO1 and AGO2 immunoprecipitates are shown as heat maps (grey-scale), with the ratio between both libraries indicating preferential association with AGO1 or AGO2 (green/red-scale). The nucleotide composition and median base pairing (bulges were counted as mismatches) are shown for highly biased sequences (70% or more). **(b)** Argonaute loading assays for artificial *miR-276a* duplexes (shown to the right). Preferential association of each strand to AGO1 (positive values) or AGO2 (negative values) was determined by calculating Argonaute loading indices (log_2 scale). Partitioning of wild-type *miR-276a* duplexes is shown in the inset. **(c)** Thermodynamic properties of AGO2-bounds endo-siRNAs matching the *klarsicht* locus are shown. SiRNA duplexes with both strands detected were split into those with strong strand bias (20:1 or higher) and weak strand bias (5:1 or lower), with differences in their thermodynamic properties shown as ratios (energy of guide strand end versus those of the passenger strand end). (Image and legend modified from: Czech, Zhou *et al.*, 2009)

To further assess the impact of thermodynamic asymmetry on strand selection of duplexes with perfect dsRNA character, we analysed the energetic properties of endo-siRNAs derived from the *klarsicht* locus (Fig. 2-5c) and viral siRNAs. As expected by their duplex structure, these sequences were heavily biased towards AGO2. Sequences for which guide and passenger strands were detected in AGO2 immunoprecipitates were divided into two categories: duplexes that show strong strand bias (guide to passenger ratio of 20:1 or higher) and weak strand bias (ratio of 5:1 or lower). The terminal energies of both duplex ends were calculated using UNAfold for up to six nucleotides (Markham & Zuker, 2008) and the average energies were computed for all sequences of each category. The comparison of guide-strand ends and passenger-strand ends confirmed previous reports (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003), with endo-siRNAs

showing strong strand bias also exhibiting high thermodynamic asymmetric of the duplex ends. In contrast, endo-siRNAs with no or weak strand bias show only little, if any, energy differences between the duplex ends, indicating the contribution of thermodynamic asymmetry to strand selection for duplexes with perfect dsRNA.

Conclusions

Profiling of Argonaute-bound small RNAs revealed consistent results with previous reports that miRNAs principally occupy AGO1, whereas endo-siRNAs preferentially associate with AGO2. The characterization of AGO2-bound small RNAs identified a surprising association with miR* strands, which were previously believed to be mere by-products of miRNA biogenesis. Reporter assays suggest that miR*s can silence targets both *in vitro* and *in vivo*, though it remains unknown if they have biologically relevant endogenous targets.

The bi-functionality of miRNA precursors, potentially funnelling different strands into distinct Argonaute complexes, reveals the limitations of existing models (Tomari *et al.*, 2007), as the degree of base pairing within a given duplex can not explain different fates of miR and miR* strands with respect to AGO1 and AGO2 loading. Furthermore, as each strand of a precursor appears to be assessed separately, strand choice has to depend on additional properties. Our analysis of AGO1- or AGO2-biased sequences revealed determinants that enable us to predict the fate of small RNA strands during sorting. For duplexes with mismatches, pairing at the central region is the dominant determinant, with each strand being assessed individually. While thermodynamic asymmetry is a minor determinant for duplexes with imperfect dsRNA, it remains the principal determinant of strand choice for perfectly paired duplexes (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003). We were able to manipulate the sorting machinery *in vitro* through artificial small RNAs, and redirect them to different AGO complexes depending on the small RNA duplex parameters. Thus, by incorporating previously proposed and newly identified determinants, we propose a hierarchy of sorting rules that predict the fate of small RNA strands during Argonaute loading. Taken together, we demonstrate that small RNA sorting pathways are more integrated than previously thought.

2.1.3 Development of shRNAs as tools for efficient gene silencing

Background

The establishment of reverse genetic approaches has changed the pace and scope of studies of gene function in both tissue culture and *in vivo*. Currently available tools for transgenic RNAi in *Drosophila* typically utilize ~500 bp long double-stranded hairpin RNAs (Perrimon *et al.*, 2010). Dcr2 processes these dsRNAs into several siRNAs that load into AGO2 and cleave target transcripts. While dsRNA-mediated RNAi has proven to be a powerful tool for silencing in somatic cells, it is ineffective in the female germline, preventing similar studies in oogenesis or the germline piRNA pathway. Furthermore, the production of multiple siRNA molecules from a single dsRNA could lead to unwanted off-target suppression of transcripts that share sufficient sequence complementarity, a phenomenon previously observed (Kulkarni *et al.*, 2006). Alternative approaches using short hairpin RNAs (shRNAs) and artificial microRNAs have been shown to trigger efficient gene silencing in mammals and plants (Paddison *et al.*, 2004; Schwab *et al.*, 2006), with some reports demonstrating efficacy in flies (Chen *et al.*, 2007a; Haley *et al.*, 2008). Using information gained from studies of small RNA sorting in flies (Section 2.1.2), we designed and generated an optimized, genome-wide shRNA library for *Drosophila*. This work was published as Ni, Zhou, Czech *et al.*, 2011 in *Nature Methods*.

Results

Two previous studies demonstrated the ability of miRNA mimics (constructs modelled on a miRNA precursor) in *Drosophila* to repress target genes in somatic cells and, more importantly, in the female germline (Chen *et al.*, 2007a; Haley *et al.*, 2008). Our shRNA constructs are based on a similar design, with sequences targeting the gene of interest embedded into the *miR-1* scaffold (Fig. 2-6a). Based on our knowledge from previous studies (Tomari *et al.*, 2007; Czech *et al.*, 2009; Okamura *et al.*, 2009; Czech & Hannon, 2011), we expect Drosha/Pasha and Dcr1/Loqs complexes to process these shRNA precursors to yield small RNA duplexes, with the perfect duplex structure without mismatches ensuring predominant loading of shRNAs into AGO2. In addition, we also designed the shRNA duplexes to feature thermodynamic asymmetry that would favour the desired strand for loading (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003), further reducing putative off-target effects. To facilitate cloning and to ensure robust expression of shRNAs, a new vector, named pVALIUM20, was constructed by inserting the *miR-1* backbone and unique restriction sites for introduction of desired sequences. This vector

also contained features of the UAST plasmid (10x UAS sites, *hsp70* core promoter, SV40 polyadenylation signal), which allows spatially and temporally regulated expression via the GAL4 transcription activator protein (Brand & Perrimon, 1993).

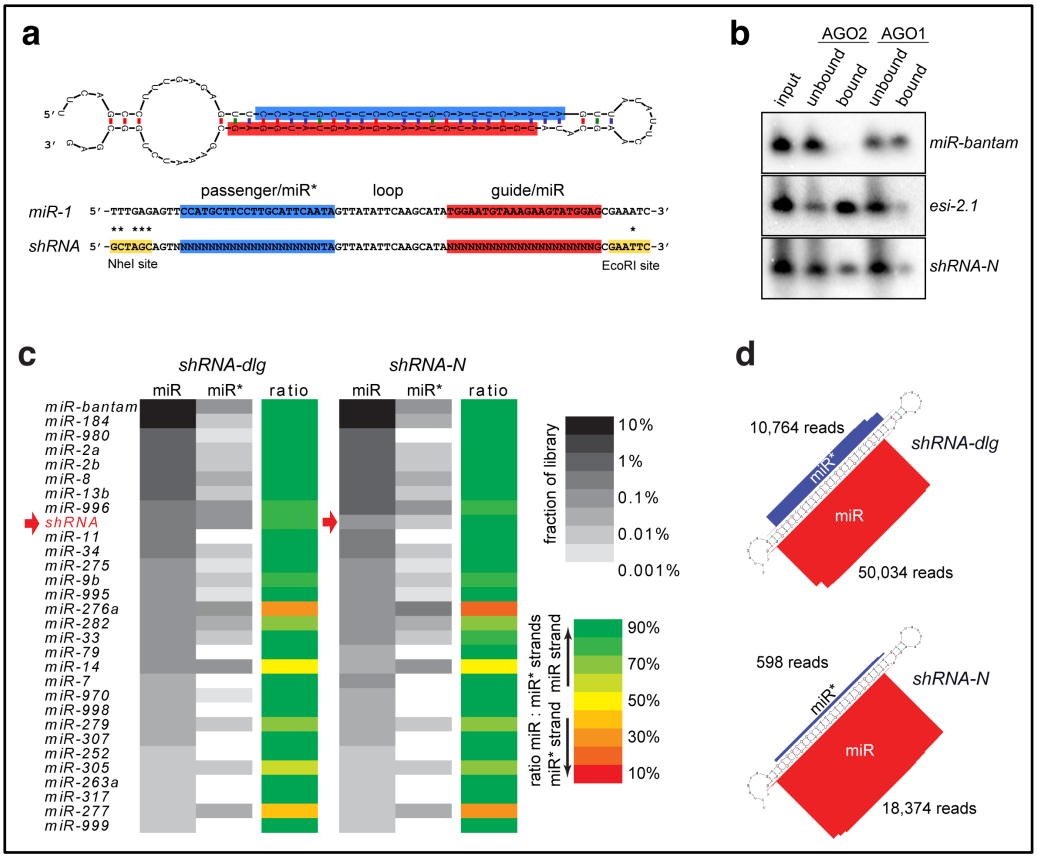


Fig. 2-6. Design, loading, abundance and processing accuracy of shRNAs. (a) Schematic drawing of shRNA and *Drosophila miR-1* hairpins, with guide/miR strands in red and passenger/miR* strands in blue (sequence against the gene of interest are indicated by N). **(b)** Northern blots of AGO1- and AGO2-associated RNAs from S2 cells transfected with a construct expressing *shRNA-N*. The membrane was sequentially probed for *miR-bantam*, *esi-2.1* and *shRNA-N* **(c)** Small RNA sequencing from S2 cells independently transfected with the indicated shRNA constructs was carried out. Heat maps showing the relative levels of the 30 most abundant miRNAs and the indicated shRNA in each library were calculated. MiR and miR* strands are shown separately (grey-scale) with their ratio calculated (green-red-scale) **(d)** ShRNA reads within the indicated libraries were mapped onto their originating precursor transcripts, with the guide/miR strand in red and the passenger/miR* strand in blue bars. Sharp peaks at the intended cleavage sites indicate accurate 5' and 3' processing. (Image and legend modified from: Ni, Zhou, Czech *et al.*, 2011)

To ensure shRNAs are generated and loaded as predicted, several shRNA constructs were expressed in S2 cells depleted of components of the miRNA and siRNA pathways. The effects of these knockdowns on shRNAs were compared to endo-siRNAs (*esi-2.1*) and miRNAs (*miR-bantam*) by Northern blots. Depletion of the miRNA pathway factors Drosha and Pasha (combined), Dcr1, Loqs or AGO1 resulted in reduced mature *miR-bantam* levels, which was accompanied by precursor accumulation when Loqs or Dcr1 were depleted. Knockdown of the siRNA pathway factors *dcr2*, *r2d2*, or *ago2* substantially reduced the levels of *esi-2.1*, without affecting miRNAs. Consistent with

previous reports (Czech *et al.*, 2008; Okamura *et al.*, 2008a; Zhou *et al.*, 2009), knockdown of *loqs* also reduced the levels of *esi-2.1*. As expected from their miRNA scaffolds, the processing of shRNAs depended on canonical miRNA pathway components, while the substantial but not complete resistance of shRNAs to β-elimination suggests that the majority of shRNAs associates with AGO2. Immunoprecipitation of AGO1 and AGO2 complexes followed by Northern blots also revealed that the major shRNA fraction is bound to AGO2, though a small portion associated with AGO1 (Fig. 2-6b). Taken together, shRNAs are processed and loaded *in vitro* as the design intended.

Deep sequencing of small RNAs from S2 cells expressing shRNAs identified expression levels similar to abundant microRNAs (Fig. 2-6c). Mapping of shRNA sequences onto their corresponding precursors confirmed their accurate processing, with the 5' ends of shRNAs perfectly matching to the predicted RNase III cleavage sites (Fig. 2-6d). As intended, the dominant length of shRNAs is 22-nt, and the ratios between guide and passenger strands of shRNAs mirror the patterns predicted by thermodynamic differences (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003), with guide strands significantly more abundant (Fig. 2-6d). Dominant guide strand loading is particularly important, as loading of passenger strands would trigger the suppression of different targets. Our results demonstrate that shRNAs from our miRNA-mimics are (1) abundantly expressed, (2) precisely processed, and (3) predominantly produce guide strands. Therefore, we believe that they are capable to reliably silence their intended targets.

To test the efficacy of our shRNA design, we generated several shRNAs against *white* (*shRNA-w*) or *Notch* (*shRNA-N*), cloned them into pVALIUM20 vectors, and established transgenic fly lines. Somatic expression of *shRNA-w* using the eye-specific *GMR*-GAL4 driver (Freeman, 1996) resulted in an eye colour phenotype highly reminiscent of reported *white* null mutants, while lines expressing traditional, long dsRNA constructs show less severe phenotypes (Fig. 2-7a). Forced expression of *shRNA-N* by the wing-specific driver *C96*-GAL4 (Presente *et al.*, 2002) produced notched wings with abnormal veins (Fig. 2-7b), similar to the wing phenotypes described for mutant *Notch* alleles, with shRNA-mediated phenotypes more penetrant than those from long hairpin RNAs. Additional shRNA and long hairpin lines against other targets with known phenotypes confirmed these results, indicating the higher effectiveness of shRNAs in silencing their targets.

To confirm that our shRNA constructs are functional in the female germline, we generated shRNAs against the piRNA pathway components Piwi, Aub, Armi, and Spn-E (expressed from pVALIUM20) induced by the germline driver *MTD*-GAL4 (bearing three

strong promoters active in *Drosophila* germ cells) (Petrella *et al.*, 2007). These flies exhibit the expected oogenesis defects (e.g., transposon de-repression) in the female ovary, though the effects are not as robust. To optimize the shRNA expression in germline cells, we generated pVALIUM22, which contains features of UASp (10x UAS sites, *P-element* transposase minimal promoter, K10 polyadenylation signal) (Rorth, 1998). Depletion of the same piRNA pathway genes (Piwi, Aub, Armi, Spn-E) with the improved construct resulted in stronger phenotypes, similar to those reported in null mutants. Immunofluorescence staining shows strong, germline-specific depletion of the targeted proteins (Fig. 2-7c), and consequently resulted in complete sterility (Fig. 2-7d) accompanied by dramatic de-repression of transposable elements as shown by quantitative polymerase chain reaction (qPCR) (Fig. 2-7e). Transposon mobilization was confirmed by immunofluorescence staining for the ORF1 of the *I-element*, a LINE-like transposon, which showed strong accumulation in early egg chambers upon depletion of Spn-E (Fig. 2-7f).

Our results demonstrate the superior silencing effectiveness of shRNAs in somatic and germline cells compared to long dsRNA hairpins. We therefore started to generate a genome-wide library of shRNA constructs for use in cell culture and the production of transgenic fly lines (Transgenic RNAi Project (TRiP), Harvard). Using the DSIR (Designer of Small Interfering RNA) prediction algorithm (Vert *et al.*, 2006), 83,256 unique shRNA constructs were designed, targeting all 14,208 annotated protein-coding genes (*Drosophila* genome release 5, April 2006), with an average coverage of ~6 shRNAs per gene. Oligonucleotides corresponding to the designed shRNA sequences were synthesized *in situ* on custom Agilent microarrays (Cleary *et al.*, 2004), amplified, and cloned in pools into pVALIUM20 and pVALIUM22 vectors. Individual clones were picked and sequenced (by Open Biosystems and iXpress), with positive clones transferred into 96-well plates and re-confirmed by in-house sequencing. The library construction is still in progress, with transgenic flies continuously being made available through the TRiP homepage.

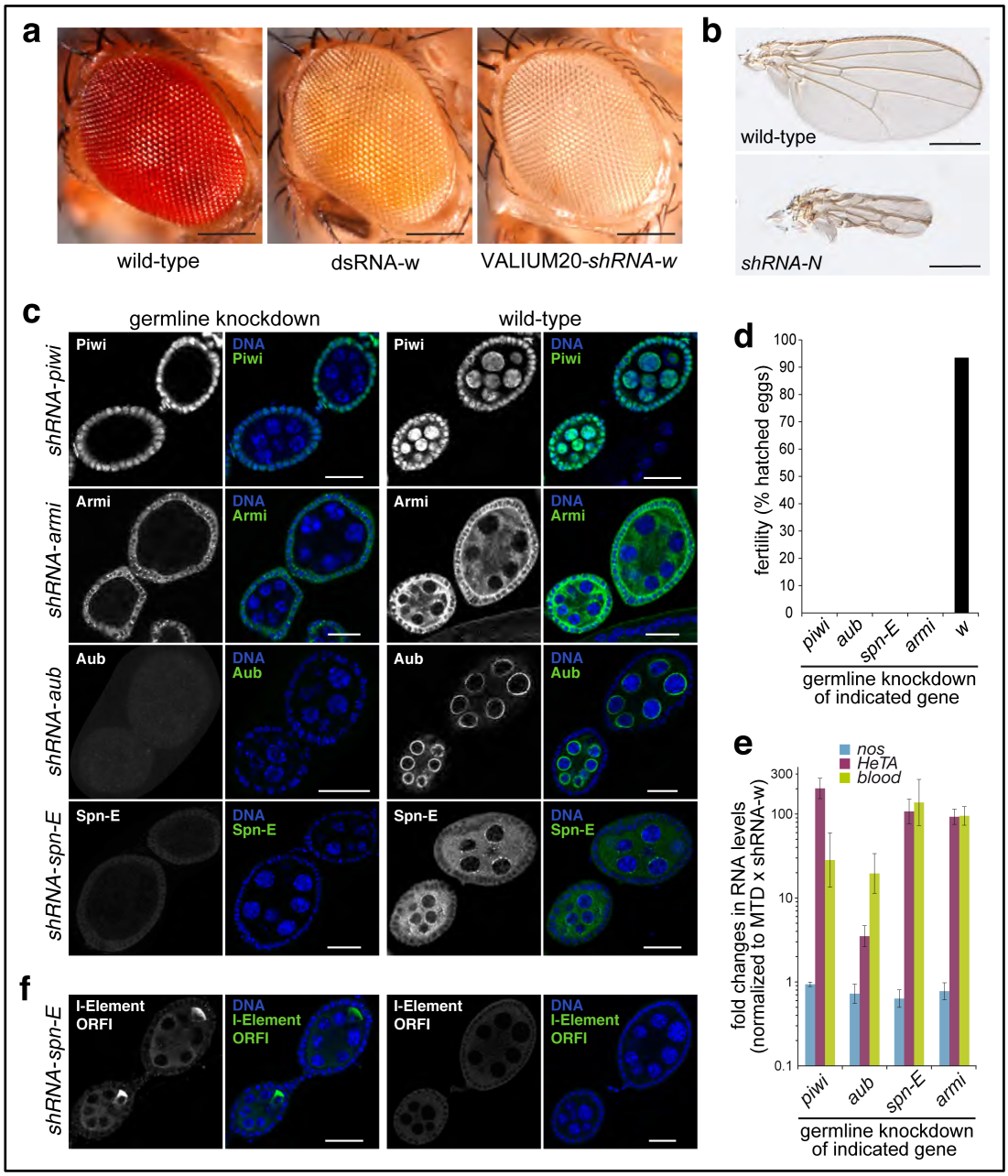


Fig. 2-7. Phenotypes of shRNA-mediated gene silencing. (a) Comparison of shRNA- (right) and dsRNA-mediated (middle) knockdown of white using the *GMR*-GAL4 driver. Wild-type is shown to the left (Scale bars equal 100 μm). **(b)** *C96*-GAL4-mediated depletion of *Notch* in the wing compared to wild-type (Scale bars equal 400 μm). **(c)** Immunofluorescence staining of piRNA pathway components in early egg chambers showing germline-specific depletion of indicated factors upon shRNA-mediated knockdown with the *MTD*-GAL4 driver. DAPI is used to stain DNA (blue) and specific factors are coloured in green (Scale bars equal 20 μm). **(d)** Female fertility rates upon germline knockdown of indicated piRNA pathway factors (using *MTD*-GAL4 and VALIUM22-expressed shRNAs). Between 300 and 500 eggs were counted for each knockdown. **(e)** Steady-state levels of two transposable elements (*HeTA* and *blood*) and the germline-expressed *nanos* control upon germline-specific depletion of the indicated piRNA pathway factors. Data is compared to white knockdowns and normalized to *rp49*. Shown are average data of three independent biological replicates. Error bars indicate SD (n = 3). **(f)** Immunofluorescence staining of the ORF1 (green) of the *I*-element transposon in early egg chambers upon knockdown of Spn-E (left) compared to wild-type (right). DAPI was used to stain DNA (blue). Scale bars equal 20 μm. (Image and legend modified from: Ni, Zhou, Czech *et al.*, 2011)

Conclusions

Our work has demonstrated the superior performance of shRNAs compared to traditional silencing triggers generated from long dsRNAs. We show that shRNAs generate more potent knockdowns than dsRNA hairpins, comparable to genetic ablation in null mutants. Furthermore, shRNAs have been shown to function in the female germline, thus enabling reverse genetics studies of oogenesis and the piRNA pathway. As per our design, the biogenesis of shRNAs involves Drosha/Pasha and Dcr1/Loqs complexes, but – due to their perfect duplex structure – they are predominantly loaded into AGO2 via the canonical siRNA loading machinery. Whether shRNAs are less prone to mediate off-target effects has yet to be addressed, though with only two sequences produced, one of which is strongly favoured for loading, the theoretical potential of off-target effects is much lower. Unfortunately, residual loading of shRNAs into AGO1 might have an unforeseen effect, and could lead to unwanted suppression of targets through seed matches similar to silencing via the miRNA pathway. The production of the genome-wide shRNA library is underway, and we hope that these shRNAs will find broad applications, especially in resolving the germline piRNA pathway.

2.2 Part II: Biogenesis and function of piRNAs in *Drosophila*

2.2.1 Identification of *shutdown* as a piRNA biogenesis factor

Background

Reproductive fitness of individuals strongly depends on the integrity of their gametic genomes. In animals, the piRNA pathway has evolved to protect the genetic material of germ cells against the deleterious effects of transposon mobilization, and to maintain genomic integrity (Khurana & Theurkauf, 2010; Senti & Brennecke, 2010; Siomi *et al.*, 2011). The pathway utilizes piRNA clusters, which contains remnants of recent and past transposition events, to generate silencing triggers (primary piRNAs) capable of detecting active mobile elements (Aravin *et al.*, 2006; Girard *et al.*, 2006; Brennecke *et al.*, 2007). Upon identification and targeting of transposon transcripts, the ping-pong cycle further amplifies the silencing triggers via the generation of secondary piRNAs, thus enabling an enhanced response to active transposon threats (Brennecke *et al.*, 2007; Gunawardane *et al.*, 2007; Li *et al.*, 2009; Malone *et al.*, 2009). While several genes were linked to primary piRNA biogenesis or the ping-pong cycle (Chen *et al.*, 2007b; Klattenhoff *et al.*, 2007; Pane *et al.*, 2007; Li *et al.*, 2009; Malone *et al.*, 2009; Szakmary *et al.*, 2009; Haase *et al.*, 2010; Olivieri *et al.*, 2010; Saito *et al.*, 2010; Handler *et al.*, 2011; Zamparini *et al.*, 2011), their precise molecular functions remain largely unknown. Many were originally reported in a study aimed to identify genes affecting female fertility in *Drosophila* (e.g., aub, squ, zuc, cuff) (Schüpbach & Wieschaus, 1989; Schüpbach & Wieschaus, 1991), and their associated mutant phenotypes became the hallmarks of piRNA pathway perturbations.

A recent study from our laboratory has identified members of the FK506-binding protein (FKBP) family as protein partners of mammalian PIWI-clade proteins (Vagin *et al.*, 2009). In addition to their characteristic FK506-binding protein domain, FKBP members often also feature tetratricopeptide repeat (TPR) domains, which mediate protein-protein interactions with heat shock proteins (HSPs) (Pratt, 1998; Pratt *et al.*, 2004; Allan & Ratajczak, 2011), with the latter shown to participate in RISC loading in mammals and *Drosophila* (Iki *et al.*, 2010; Iwasaki *et al.*, 2010; Miyoshi *et al.*, 2010b; Iki *et al.*, 2012). Intriguingly, the proteins and domain structures of some FKBP members are conserved between mammals and flies.

To address a potential role of *Drosophila* FKBP family members in the piRNA pathway, we depleted each protein and probed for transposon de-repression and sterility

phenotypes. Out of those investigated, only *shutdown* showed defects reminiscent of known piRNA factors. Analyses of small RNA profiles identified *shu* as important factor for the biogenesis of primary and secondary piRNAs. This work was published as Preall, Czech *et al.*, 2012 in *RNA*.

Results

Encouraged by the interaction studies of PIWI-clade proteins (Vagin *et al.*, 2009), we investigated the impact of FKBP proteins on the *Drosophila* piRNA pathway. *Drosophila* encodes eight members of the FKBP family (Fig. 2-8a) that are all expressed in ovaries and OSS cells based on RNAseq (Fig. 2-8b). However, only *shu* showed significantly enriched ovarian expression compared to other tissues (Fig. 2-8b) (Chintapalli *et al.*, 2007).

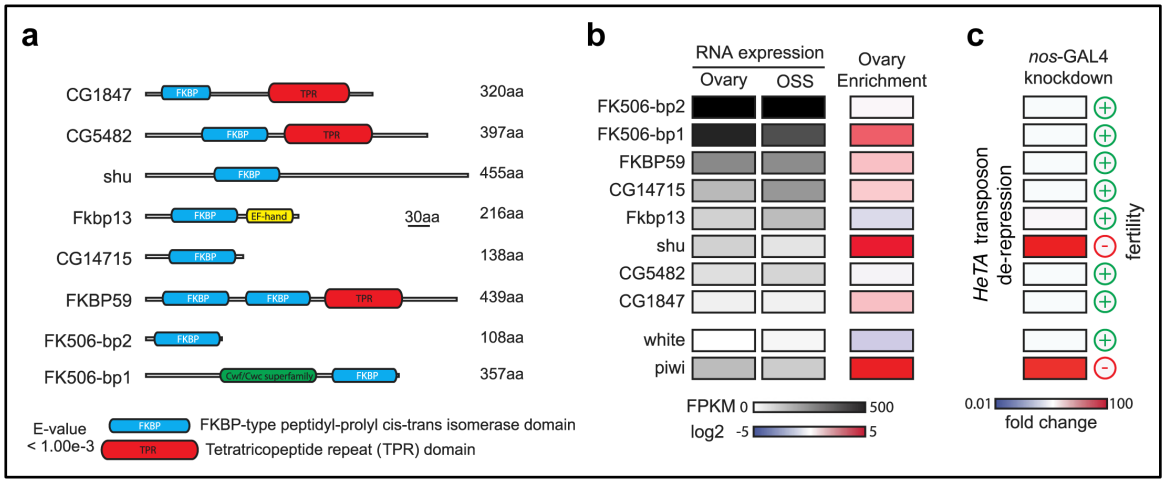


Fig. 2-8. The FKBP-family protein Shu is required for transposon silencing. (a) Schematic drawing of the domain structures of the eight *Drosophila* FKBP-family proteins. **(b)** RNAseq data showing relative expression of fly FKBP-family members in ovaries and OSS cells. The enrichment for ovarian expression (compared to other tissues) is shown to the right. The expression of white and piwi is shown as comparison. **(c)** De-repression of the *HeTA* transposon upon knockdown of each individual FKBP-family member in germ cells was detected by qPCR. White and Piwi depletion served as controls. Effects of germline knockdowns on fertility, fertile (+) or sterile (-), are shown to the right (Image and legend modified from: Preall, Czech *et al.*, 2012)

We tested the participation of individual FKBP family members in piRNA silencing by knockdown experiments using the GAL4-UAS system (Brand & Perrimon, 1993). *In vivo* depletion of FKBP proteins was performed by crossing virgin females expressing the germline-specific driver (*nos*-GAL4) (Tracey *et al.*, 2000) to males carrying a hairpin against our gene of interest under the control of the UAS promoter (UAS-dsRNA-*x*) (from the Vienna *Drosophila* RNAi Center, VDRC) (Dietzl *et al.*, 2007). The females also possess a UAS-driven construct to overexpress Dcr2, which has been shown to overcome previously observed limitations in efficiency of dsRNAs in germ cells of the female ovary (Handler *et al.*, 2011; Wang & Elgin, 2011). Two hallmarks of canonical

piRNA mutants, transposon de-repression and female sterility, were used to score the requirement of FKBP proteins in the piRNA pathway. Steady-state levels of the germline-specific transposon *HeTA* was assessed by qPCR in RNA isolated from ovaries depleted of individual FKBP family members (Fig. 2-8c). Only Shu depletion resulted in *HeTA* de-repression, similar to what was observed with *piwi* knockdowns. Accordingly, female sterility was only observed in *shu* and *piwi* knockdowns, with other FKBP knockdowns showing no effects (Fig. 2-8c). Interestingly, mutations in *shu* were previously reported to affect female fertility and proper germline development (Schüpbach & Wieschaus, 1991; Munn & Steward, 2000).

The impact of germline depletion of Shu on other transposons was analysed by qPCR and compared with knockdowns of *armi* and *piwi*, both of which are key components of the piRNA pathway (Fig. 2-9a). Depletion of Shu de-repressed several germline-specific elements (e.g., *TAHRE*, *TART*, *HeTA*, *burdock*, *transpac*) similar to levels in *armi* and *piwi* control knockdowns, but has no effect on transposons specific to somatic cells (e.g., *ZAM*). Germline-specific depletion of Piwi, Armi, or Shu caused severe sterility, where none of the eggs were viable (Fig. 2-9b). In addition, *shu* knockdown in germline cells results in an egg phenotype with fused dorsal appendages similar to other reported piRNA mutants (Fig. 2-9c). This phenotype is indicative of patterning defects during oogenesis as a consequence of transposition-induced DNA breaks and subsequent meiotic checkpoint activation (Theurkauf *et al.*, 2006; Chen *et al.*, 2007b; Klattenhoff *et al.*, 2007). Females with germline-specific depletion of Shu also produced significantly fewer eggs than other mutants (Fig. 2-9b), indicating additional defects associated with shu depletion.

Under normal conditions, Piwi shows nuclear localization in both somatic follicle and germline cells (Cox *et al.*, 1998; Cox *et al.*, 2000; Saito *et al.*, 2006; Brennecke *et al.*, 2007), but impaired primary biogenesis results in unloaded Piwi, which remains cytoplasmic (Malone *et al.*, 2009, Saito *et al.*, 2009; Klenov *et al.*, 2011). Aub and AGO3, in contrast, are exclusively expressed in nurse cells of the germline and localize to a perinuclear structure called nuage, with disruptions of secondary piRNA production causing the redistribution of Aub and AGO3 to sparse, cytoplasmic foci (Brennecke *et al.*, 2007; Gunawardane *et al.*, 2007; Lim & Kai, 2007; Nishida *et al.*, 2009). Thus, localization patterns of PIWI-clade proteins can serve as indicators for defects in primary biogenesis and ping-pong cycle (Malone *et al.*, 2009).

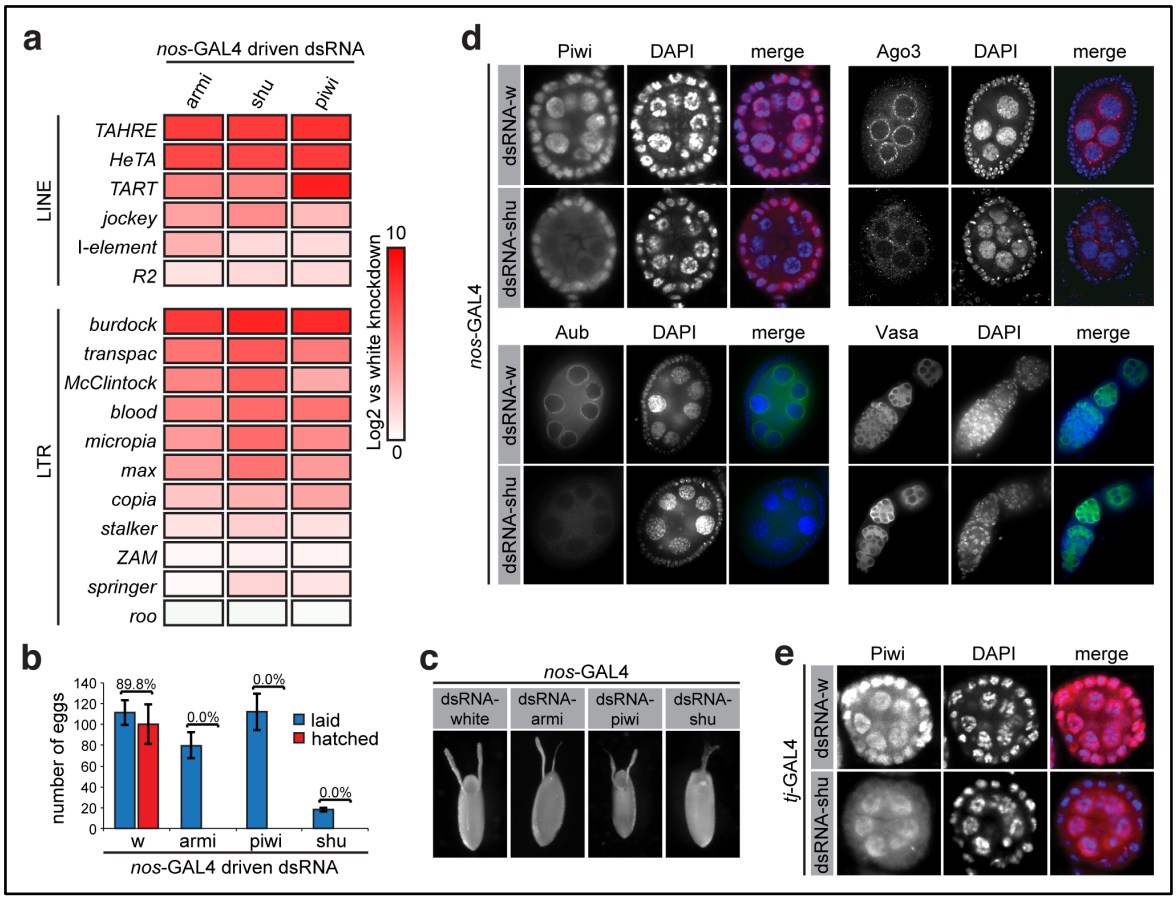


Fig. 2-9. Knockdown of *shu* in the germline results in typical piRNA pathway phenotypes. (a) Heat map showing global transposon de-repression upon germline knockdown of *shu* compared to depletion of white, armi and piwi. **(b)** Shu depletion and RNAi against *armi* and *piwi* in germline cells results in female sterility (hatching rates are zero). Shu depleted females lay fewer eggs than flies depleted of other piRNA pathway factors. **(c)** Fused dorsal appendages indicate patterning defects upon germline knockdown of *shu*. **(d)** Germline knockdown of *shu* causes delocalization of Piwi from nuclei. In addition, Aub and AGO3 delocalized from nuage, with nuage still intact as shown by Vas staining. RNAi against *white* served as control **(e)** Soma-specific depletion of Shu results in delocalization of Piwi from follicle cell nuclei. Knockdown of *white* is shown as control (Image and legend modified from: Preall, Czech *et al.*, 2012)

Ovaries with reduced Shu expression show delocalization of all PIWI-clade proteins, with Piwi redistributed to the cytoplasm, and Aub and Ago3 absent from nuage (Fig. 2-9d). However, the nuage structure in general was not disrupted, as Vasa, a major factor for nuage organization (Liang *et al.*, 1994; Malone *et al.*, 2009), showed no altered localization (Fig. 2-9d). Knockdown of *shu* in somatic cells using the follicle cell specific *tj*-GAL4 driver (Tanentzapf *et al.*, 2007; Olivieri *et al.*, 2010) also delocalizes Piwi from nuclei (Fig. 2-9e), while Yb-bodies, the suggested sites of Piwi loading (Szakmary *et al.*, 2009; Olivieri *et al.*, 2010; Saito *et al.*, 2010; Qi *et al.*, 2011), remained unaltered. Our results suggest that *shu* is a *bona fide* piRNA pathway component and plays a role in primary biogenesis and in the ping-pong amplification loop.

We further analysed the impact on small RNA profiles upon depletion of shu in the soma and germline. Small RNA libraries were generated from germline knockdowns (using the *nos*-GAL4 driver), and normalized using read counts of piRNAs uniquely mapped to the soma-specific *flam* cluster, which should be unaffected. Germline-specific depletion of shu caused a dramatic reduction in piRNAs derived from all germline clusters (Fig. 2-10a). Compared to the *white* control, shu depletion caused a 11.4-fold reduction in piRNA levels from the *42AB* cluster. This is more severe than the *piwi* knockdown (2.8-fold reduction), though Piwi-independent loading of Aub and AGO3 with *42AB* piRNAs could account for the less pronounced impact. *Shu* knockdowns in germ cells also dramatically reduce piRNA levels of other germline clusters (e.g., *20A*, *38C*, *80E*), but have no effects on somatic clusters (Fig. 2-10b).

To assess the effect of *Shu* on the piRNA amplification loop, we analysed ping-pong signatures in piRNAs, which are defined by the relative frequency of 10-nt overlaps between two reads of opposite orientations. Consistent with the significant depletion of *42AB*-derived piRNAs, *shu* knockdowns display a dramatic reduction in ping-pong signatures (Fig. 2-10a). This was in contrast to *piwi* knockdowns, which caused no change of ping-pong signatures as predicted due to unperturbed Aub and AGO3 loading. Furthermore, depletion of Shu did not change miRNA levels, indicating a specific role in the piRNA pathway.

The somatic follicle cells only express Piwi, and therefore only employ a primary biogenesis pathway without ping-pong amplification. In these cells, Piwi is mainly associated with piRNAs derived the somatic *flam* cluster (Brennecke *et al.*, 2007; Lau *et al.*, 2009; Malone *et al.*, 2009; Saito *et al.*, 2009). To address the role of *Shu* in ovarian somatic cells, we generated small RNA libraries from follicle cell knockdowns (using the *tj*-GAL4 driver), and normalized using read numbers of piRNAs uniquely mapping to the germline-specific *42AB* cluster, which was not affected by somatic knockdowns. Depletion of Piwi or Shu caused a marked reduction of piRNAs derived from the *flam* locus compared to the control (5.2-fold and 2.9-fold respectively), with similar results at another soma-specific locus (the 3' UTR of *tj*) (Fig. 2-10c) (Lau *et al.*, 2009; Saito *et al.*, 2009). Consistent with previous reports, germline clusters remained unchanged (Fig. 2-10d), and *flam*-derived piRNAs show no ping-pong signature upon either knockdown (Fig. 2-10c). Lastly, other small RNA classes, including miRNAs and endo-siRNAs, were not affected by *shu* knockdowns.

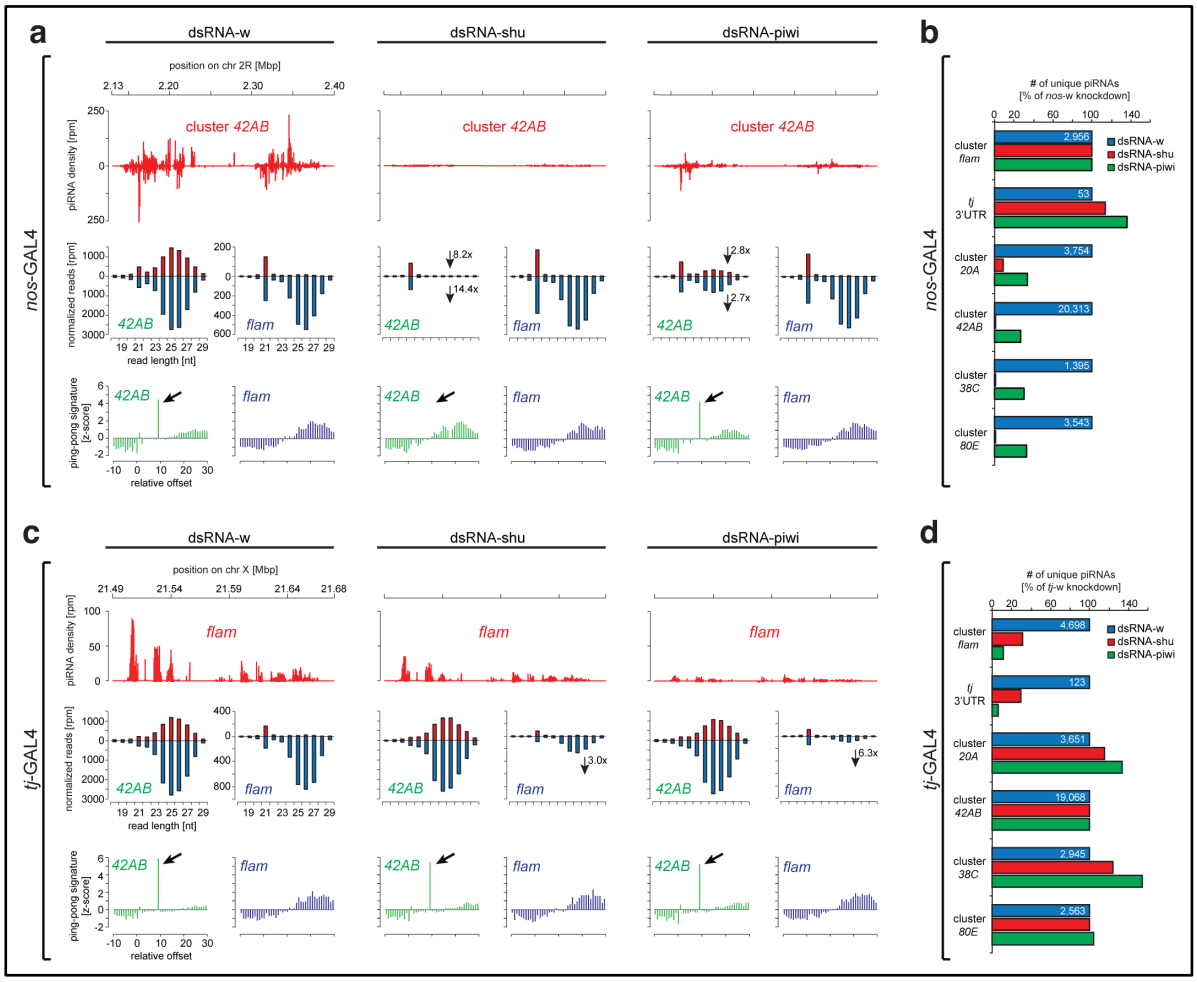


Fig. 2-10. Shu depletion causes loss of cluster-derived piRNAs in germline and somatic tissues. (a) Density plots of piRNAs mapping to the germline cluster *42AB* are shown in ovaries from flies with indicated germline knockdowns using *nos*-GAL4 (top). The size distribution of sense (red) and antisense (blue) reads matching to the germline *42AB* and the somatic *flam* clusters are shown as histograms (middle). Ping-pong signatures (peak at position 9 highlighted by arrow) of piRNAs mapping to *42AB* and *flamenco* clusters are displayed as histograms (bottom). **(b)** Histograms showing relative piRNA levels for indicated germline and somatic clusters. As somatic piRNA populations were unaffected by germline specific knockdowns, piRNAs mapping to *flamenco* were used for normalization across libraries. For better comparison, reads in *white* controls were set to 100%. **(c)** and **(d)** are similar to (a) and (b), but knockdowns were in somatic follicle cells (using the *tj*-GAL4 driver). Instead of *42AB*, reads mapping to *flam* are shown in **(c)**; and libraries in **(d)** were normalized to reads mapping to *42AB* (not affected by soma-specific knockdown). (Image and legend modified from: Preall, Czech *et al.*, 2012)

We next analysed the impact of germline-specific *shu* knockdowns on transposon-derived piRNAs by quantifying sense and antisense piRNAs matching to a set of 75 established *Drosophila* transposable elements. PiRNAs corresponding to transposons with prominent germline expression (e.g., *HeTA*, *TAHRE*, *TART*, *Rt1b*, *I-element*) were substantially reduced by *shu* knockdowns, with sense and antisense piRNAs affected to similar extents. This suggested loading defects of all three PIWI-clade proteins upon Shu depletion. Consistent with our results from cluster piRNAs, only *shu* knockdowns resulted in reduced ping-pong signatures of piRNAs matching to

germline-biased elements (e.g., *Rt1b*, *roo*), whereas knockdown of *piwi* had no effect. Lastly, piRNAs matching to transposons known to be active only in somatic cells (e.g., *gypsy*, *tabor*, *ZAM*) were not altered significantly upon germline depletion of Shu or Piwi.

Conclusions

Previous studies from fly genetics and proteomics of PIWI-clade proteins in mice have implicated proteins of the FKBP family as components of the piRNA pathway (Schüpbach & Wieschaus, 1991; Munn & Steward, 2000; Vagin *et al.*, 2009). Here, we depleted each of the eight *Drosophila* FKBP proteins in germ cells of the female ovary to assess their impact, and only *shutdown* was shown to be required for transposon silencing and female fertility. A detailed phenotypic analysis revealed additional defects upon germline-specific depletion of Shu that are reminiscent of known piRNA mutants, such as egg patterning issues and mislocalization of PIWI-clade proteins. Small RNA profiles from soma- or germline-specific *shu* knockdowns also display a dramatic reduction of primary and secondary piRNA levels, which consequently results in a loss of ping-pong signatures. These results implicate Shu as a bona fide piRNA pathway component. Interestingly, Shu is the first gene identified to affect the levels of both primary and secondary piRNAs.

Our results point to an evolutionarily conserved role for FKBP proteins in the piRNA pathway. Various reports have demonstrated an interaction of FKBP proteins with HSP90 chaperone complexes via their TPR domain (Pratt, 1998; Pratt *et al.*, 2004; Allan & Ratajczak, 2011), with HSP90 complexes shown to facilitate loading of RISC with small RNAs (Iki *et al.*, 2010; Iwasaki *et al.*, 2010; Miyoshi *et al.*, 2010b; Iki *et al.*, 2012). Therefore, it is possible that *shu* might function in loading of piRNAs into PIWI-clade effector complexes, and play important roles in assembling piRNA-primed RISC in the *Drosophila* germline. A more detailed characterization of Shu will require the combination of biochemical and genetic approaches.

2.2.2 A transcriptome-wide RNAi screen in *Drosophila* ovaries reveals novel factors of the germline piRNA pathway

Background

Eukaryotic organisms of all phyla are constantly challenged by genomic parasites known as transposons. Uncontrolled mobilization of transposable elements causes malignant alterations to the host genome such as disruption of regulatory or coding regions, DNA breaks or high-order chromosomal rearrangements (McClintock, 1951), and dramatically compromise reproductive fitness if occurring in germ cell lineages

(Slotkin & Martienssen, 2007; Malone & Hannon, 2009). Potent repression of mobile elements in animal germ cells relies on the piRNA pathway, which consists of 23- to 28-nt piRNAs and their PIWI-clade Argonaute partner proteins (Malone & Hannon, 2009; Khurana & Theurkauf, 2010; Senti & Brennecke, 2010; Siomi *et al.*, 2011). In *Drosophila*, two inter-related branches of the piRNA system exist, with primary biogenesis being the sole pathway in somatic follicle cells (but also active in germ cells) and the ping-pong amplification cycle operating exclusively in the germline (Malone *et al.*, 2009). Although we are beginning to understand basic concepts of piRNA-mediated silencing, many key aspects of the pathway are still enigmatic, due to gaps in our knowledge of central factors such as nucleases and silencing effectors.

Most of the factors that play essential roles the piRNA pathway were originally uncovered by classic genetic screens, which aimed to identify mutations responsible for female fertility or oogenesis (Schüpbach & Wieschaus, 1989; Schüpbach & Wieschaus, 1991). Due to technical limitations in the female germline, reverse genetic approaches have not been deployed for *de novo* identification of piRNA components. However, we and others have recently shown that efficient gene silencing in germ cells can be achieved by over-expression of Dcr2 (Handler *et al.*, 2011; Wang & Elgin, 2011; Preall *et al.*, 2012). In order to reveal the entire repertoire of genes required for proper transposon silencing, we designed an RNAi-based screen to systematically probe for missing components of the piRNA pathway *in vivo*. We specifically targeted the germline pathway, which comprises both Drosophila piRNA pathway branches – primary biogenesis and ping-pong amplification – and will therefore cover the broadest range of novel factors. Here, we report the results from our screen that encompassed the ovarian transcriptome and identified 74 genes, including already known piRNA pathway components, whose knockdown resulted in dramatic loss of transposon silencing of four distinct mobile elements that represent different transposon classes.

This work is under preparation for submission to *Cell* as Czech, Preall *et al.* The final manuscript version might vary in content from the one presented here.

Results

We first catalogued the expressed ovarian transcriptome by RNAseq, with the goal of identifying genes that showed expression above basal levels, which is suggestive of functional relevance in the ovary. Quantification of mRNAs from ovaries of our screening stock, expressing UAS-*Dcr2* and *nos*-GAL4, revealed expression of 8,396 protein-coding genes with an average fragments per kilo base per million reads (FPKM) greater than 1 in two biological replicates (Fig. 2-11a, top). This corresponds to 60.86%

of the 13,795 annotated genes in *Drosophila melanogaster* (Refseq release 5.48; non-coding transcripts were removed). RNAi lines were obtained from the Vienna *Drosophila* RNAi Center (VDRC; Dietzl *et al.*, 2007) for 8,171 of these genes, covering 97.32% of ovary-expressed coding transcripts that we identified (Fig. 2-11a, bottom). For the RNAi-based screen, we used our previously reported knockdown strategy for germline-specific gene depletion (Preall *et al.*, 2012), where males possessing the UAS-driven dsRNA were mated with virgin females carrying the germline-specific *nos*-GAL4 driver and a UAS-*Dcr2* transgene, which increases RNAi efficiency (Handler *et al.*, 2011; Wang & Elgin, 2011). As a proof of concept, depletion of Piwi or Armi resulted in strong and specific de-repression of germline-dominant mobile elements, as shown by increased transposon transcript levels detected in dissected ovaries (Fig. 2-11b).

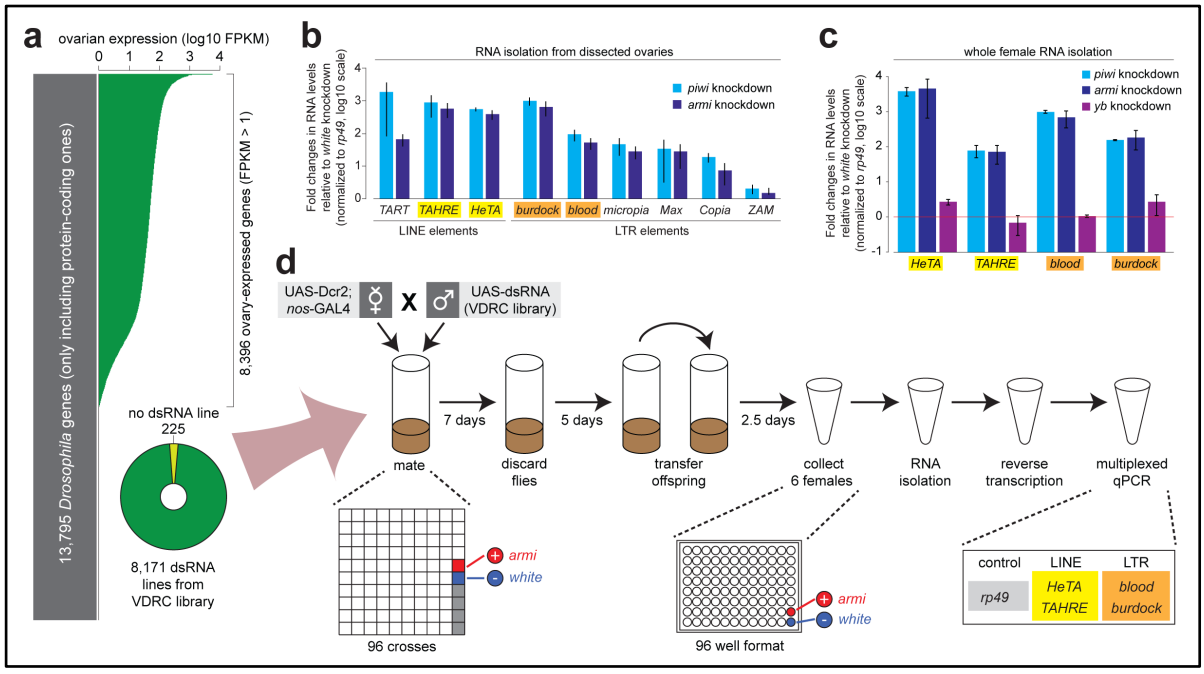


Fig. 2-11. Establishment of screening assays to monitor transposon de-repression. (a) Relative expression levels of protein-coding genes in *Drosophila melanogaster* are shown for ovarian RNAseq data as histogram. Green bars highlight ovary-expressed genes with FPKM > 1 (top). Doughnut diagram showing screened genes where dsRNA line was available from the VDRC (bottom). **(b)** Histograms show the relative expression levels of indicated transposons detected in ovaries from *Drosophila* that express dsRNA against *piwi* or *armi* in germline cells. Fold changes are relative to knockdown of *white*. Measurements were carried out on ovary-dissected total RNA. Error bars indicate standard deviation (n = 3). **(c)** Relative expression levels of the indicated mobile elements upon germline-specific knockdown of *piwi* and *armi* are shown. Depletion of Yb served as control. Fold changes relative to dsRNA against *white* (indicated by red line) were calculated. Measurements were carried out on RNA extracted from whole female flies. Error bars indicate standard deviation (n = 3). **(d)** Scheme of the screen setup. A germline-specific driver, *nos*-GAL4, was used to express UAS-dsRNA constructs in germ cells of the developing oocyte. UAS-*Dcr2* was co-expressed specifically in germ cells to enhance the RNAi response. Two and a half day old female offspring flies were collected and following RNA isolation and reverse transcription probed for de-repression of four transposons by multiplexed qPCR. Crosses were carried out in trays of 96 that contained a positive (*armi*) and negative (*white*) control knockdown. (Image and legend modified from: Czech, Preall *et al.*, in preparation)

To assess transposon de-repression upon knockdown of ovarian genes, we assay the expression of two LINE-like transposons, *HeTA* and *TAHRE*, as well as two LTR elements, *blood* and *burdock*, chosen for their consistent and robust fold changes when known piRNA components were depleted. RNA was isolated from whole adult females, and transposon transcripts were quantified simultaneously using custom TaqMan probes and established multiplexed qPCR assays. The results were normalized to transposon levels observed in knockdown of *white*, a gene not associated with the piRNA pathway. We were able to demonstrate high specificity and dynamic range of our assays, as shown by strong transposon de-repression phenotypes when Armi or Piwi were depleted (Fig. 2-11c). It also verified that whole adult females could substitute for dissected ovaries in our assays, which significantly simplified the workflow for large-scale screening. In addition, these results show the restriction of knockdown to the germline lineage, as depletion of the soma-specific factor Yb did not alter transposon levels.

The large-scale screening to identify components required for piRNA-mediated transposon silencing was carried out using the following workflow. Virgin females expressing UAS-*Dcr2* and *nos*-GAL4 were mated to males carrying a dsRNA hairpin against the gene of interest under the control of the UAS promoter. Crosses were setup in parallel in batches of 96 matings (94 experimental dsRNA lines with 2 controls targeting *white* and *armi*). Flies were mated for seven days, after which the adults were disposed, leaving the F1 larvae behind. After another five days, eclosing offspring flies were transferred to fresh food vials and left to mature for an additional 2.5 days. Once matured, six females from each cross were transferred to collection tubes, and total RNA was isolated, reverse transcribed and used as template in multiplexed qPCRs for the four transposons (Fig. 2-11d). To compare individual de-repression patterns across batches of matings, primary data was converted into *z*-scores, which indicates the deviation of the observed result (qPCR cycle number of *rp49* normalization subtracted from cycle number for transposon) from the mean value, expressed as a multiple of the standard deviation. *Z*-scores with negative values indicate transposon de-repression.

Using this strategy, we screened a collection of 8,171 dsRNA lines *in vivo* for genes involved in transposon silencing of *HeTA*, *TAHRE*, *blood*, and *burdock* elements. To increase the stringency for potential candidates, the average *z*-scores of all four transposons were calculated and heat maps were computed (Fig. 2-12a). We established a threshold of -1.5 or lower in the average *z*-score to identify potential candidates, and based on this cut-off, we found 216 dsRNA lines that result in strong loss of transposon silencing, corresponding to 2.64% of all lines tested. Strikingly, requiring these criteria for candidate selection, all of the positive *armi* controls present

within each screened batch were included, whereas all negative *white* controls were excluded (Fig. 2-12c), suggesting that our threshold was sufficiently stringent.

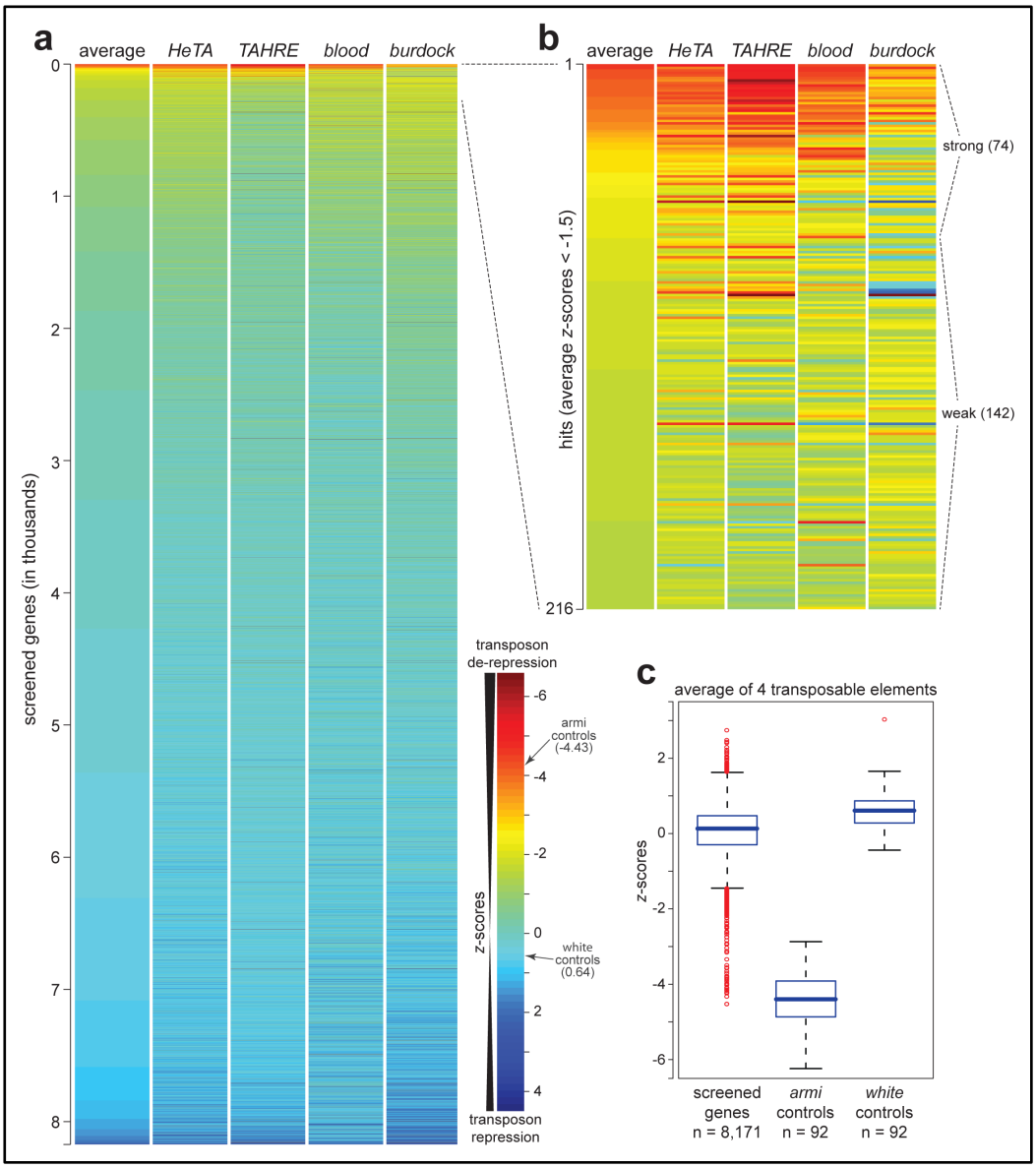


Fig. 2-12. Summary of primary screen results and determination of candidate hits. (a) Heat map displaying transposon de-repression (as *z*-scores) for all 8,171 investigated ovary-expressed genes in red-blue scale. The average of the four tested transposons is shown along the separate z-scores of *HeTA*, *TAHRE*, *blood*, and *burdock*. Negative z-scores indicate transposon de-repression (shown in red). **(b)** Close-up of the heat map for 216 candidate hits with average *z*-score < -1.5. **(c)** Box plots summarizing *z*-scores of all screened genes, positive *armi* controls, and negative *white* controls. Average data from four transposons was used for the analysis. (Image and legend modified from: Czech, Preall *et al.*, in preparation)

Depending on the magnitude of de-repression (as denoted by the *z*-score), we further categorized the 216 candidates into “weak” and “strong” groups, using an average *z*-scores of -2.0 or lower as cut-off for inclusion into the “strong” category (Fig. 2-12b). Using these very stringent criteria, we uncovered 74 genes with strong loss of

transposon silencing, which contain all known piRNA pathway components, except eggless, thus providing internal validation of our screen.

To confirm and extend our findings, the 216 candidates identified in our primary screen were re-examined by crossing our driver stock with dsRNA lines used in the original screening, and where available, a second independent RNAi line targeting a different region of the candidate gene (dsRNA or shRNA derived either from the VDRC or TRiP stock centres). Transposon levels were measured using multiplexed qPCRs, with two additional germline-expressed genes, *nos* and *γTub37C*, also investigated. To improve comparison between data sets, we normalized all expression changes to the average of two *armi* control knockdowns performed concurrently. Furthermore, we analysed potential fertility defects of F1 offspring flies by counting the number of larvae and pupae. Heat maps summarizing all assayed parameters are shown for the 75 strongest hits (Fig. 2-13). Overall, we found significant overlap between primary screen data and the validation experiments using identical dsRNA lines and independent RNAi lines targeting the same candidate gene. These results provide convincing evidence that the screen and associated assay is highly reproducible, and that our data is very robust. Notably, we were also able to demonstrate correlations between compromised fertility and either reduced expression of the germline markers *nos* and *γTub37C*, or general transposon de-repression (similar to *armi* knockdowns).

All known piRNA pathway components previously reported to affect the germline pathway – Shu, AGO3, Vas, Rhi, Squ, Piwi, Tej, Mael, Cuff, Vret, Zuc, Armi, Spn-E, BoYb, CG5508 (recently named Mino), Aub and Qin – were uncovered by our re-screening efforts, with the exception the aforementioned Egg and Krimp, which was not screened due to the unavailability of dsRNA lines. We also identified numerous novel candidates that, when depleted from germ cells, result in severe transposon de-repression and is often accompanied by fertility defects. An intriguing candidate exhibiting a dramatic transposon de-repression phenotype was *deadlock* (*del*), which was previously reported as important for germline maintenance and female fertility (Schüpbach & Wieschaus, 1991; Wehr *et al.*, 2006). Mutations in *del* result in defects during early oogenesis (Wehr *et al.*, 2006). Another dsRNA line that caused strong de-repression of all four measured transposable elements targeted the gene *CG2183*, which we named *GASZ* after its nearest vertebrate counterpart (GASZ, Germ cell specific protein with Ankyrin repeats, Sterile alpha motif, and putative basic leucine Zipper domain) (Yan *et al.*, 2002; Yan *et al.*, 2004). *Drosophila* and vertebrate GASZ share the ankyrin repeats and sterile alpha motif, but not the leucine zipper domain. Murine GASZ

is characterized as nuage component, with its loss resulting in sterility, accompanied by transposon de-repression and significantly reduced piRNA levels (Ma *et al.*, 2009).

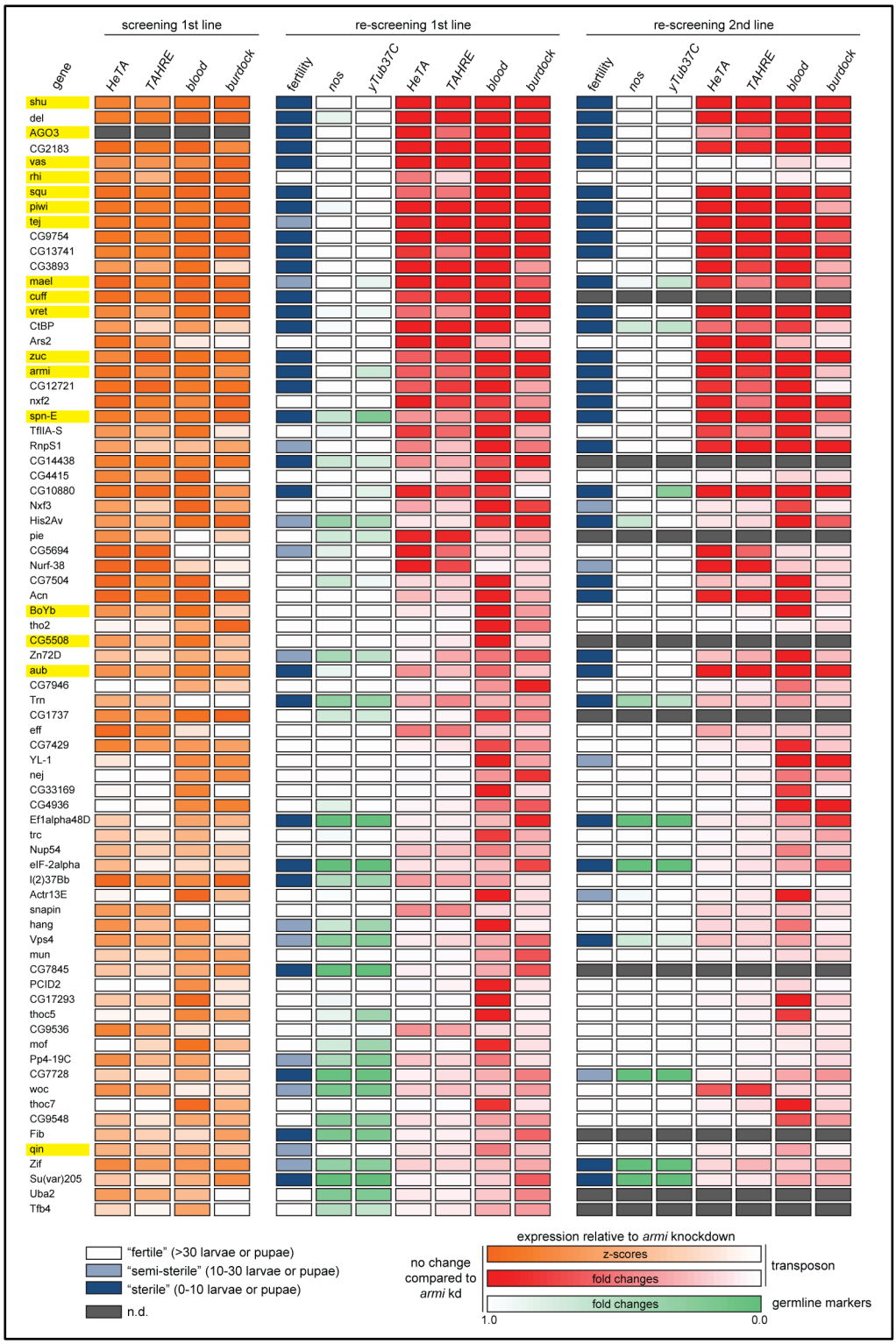


Fig. 2-13. Identification and validation of strong candidate genes. Heat maps summarizing transposon de-repression, germline marker gene expression, and sterility phenotypes upon germline-specific knockdown of indicated genes. Data is presented relative to depletion of Armi. Yellow boxes highlight known piRNA pathway components. (Image and legend modified from: Czech, Preall *et al.*, in preparation)

Yet another candidate hit, Ars2, was previously reported to impact miRNA and siRNA biogenesis (Gruber *et al.*, 2009; Sabin *et al.*, 2009), but has yet to be linked to the piRNA pathway. In all, we identified a total of 74 factors with pronounced effects on all transposons tested, with 58 genes previously not associated with transposon silencing.

To characterize the functions of GASZ and Del at the molecular level, we first studied the subcellular localization pattern of selected core components of the piRNA pathway upon germline-specific knockdown (Fig. 2-14). Knockdown of *white* (*w*), which does not affect the piRNA pathway, serves to illustrate the wild-type distribution for the investigated proteins: Piwi is localized to nurse cell nuclei, Aub and AGO3 are both enriched in nuage granules, and Armi is detected in diffuse perinuclear structures reminiscent of nuage (Cox *et al.*, 1998; Cox *et al.*, 2000; Saito *et al.*, 2006; Brennecke *et al.*, 2007; Gunawardane *et al.*, 2007; Lim & Kai, 2007; Nishida *et al.*, 2007; Saito *et al.*, 2010).

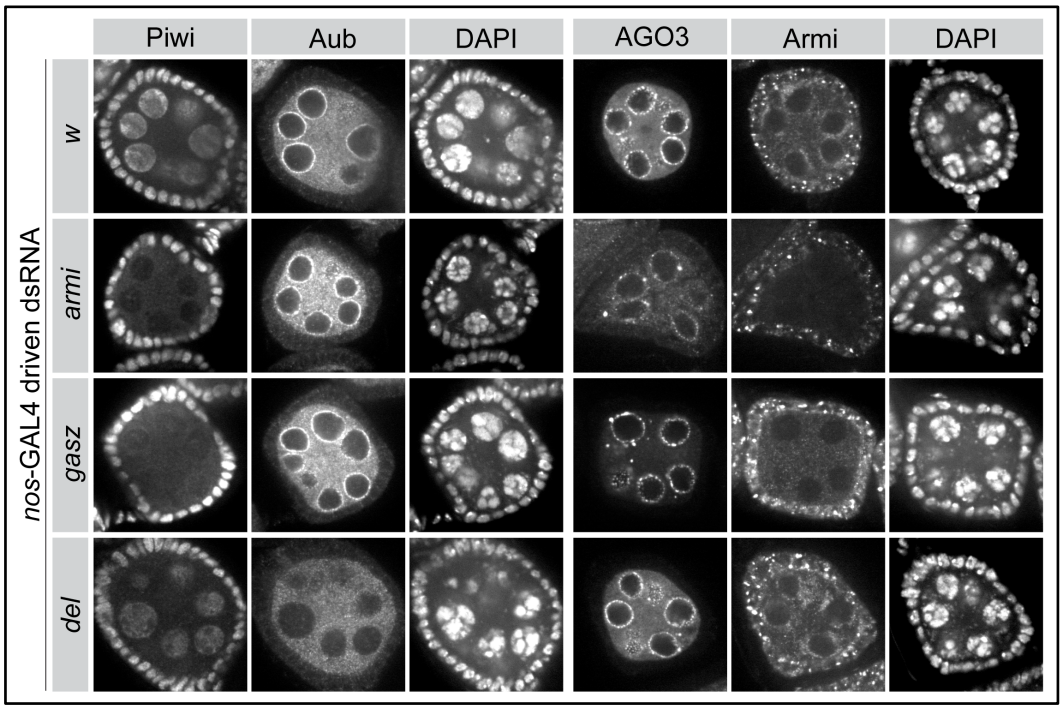


Fig. 2-14. Subcellular localization phenotypes of core piRNA pathway components upon depletion of the candidate factors GASZ and Del. Knockdown of *armi* and *gasz* in the germline using *nos*-GAL4 causes Piwi delocalization from nuclei and redistribution of Armi from nuage-like sites. The localization of Aub and AGO3 is not changed. *Nos*-GAL4-driven dsRNA against *del* results in redistribution of Aub, whereas the localization of Piwi, AGO3 and Armi is not affected. Knockdown of *white* is shown as control. (Image and legend modified from: Czech, Preall *et al.*, in preparation)

Germline knockdown of *armi* resulted in loss of protein signal below detectable levels, while Armi staining in adjacent follicle cells remained intact. Consistent with its function in primary piRNA biogenesis, depletion of Armi caused a clear redistribution of Piwi from nurse cell nuclei, but had no effect on the localization of Aub and AGO3. Upon

knockdown of *gasz*, nuclear localization of Piwi was severely compromised, whereas Aub and AGO3 appeared normal. Interestingly, we also observed a redistribution of Armi from nuage granules into cytoplasmic speckles, suggesting a function in primary piRNA biogenesis. Depletion of Del, in contrast, did not affect the localization of Piwi or Armi. However, we detected a pronounced redistribution of Aub away from nuage, while AGO3 remained normal. Thus, Del could play specific roles in the ping-pong cycle.

Transposon de-repression, reduced fertility, and disturbed protein localization of PIWI-clade members are hallmarks of bona fide piRNA pathway mutants. To directly assess likely effects on piRNA levels, we cloned and sequenced small RNA libraries from ovaries with germline-specific depletion of GASZ or Del. Small RNA reads were normalized to the number of piRNAs uniquely derived from the *flam* cluster, which are unchanged due to their somatic origin. These were then compared to samples generated from germline knockdowns of several known piRNA pathway factors (Armi, Spn-E, Aub), and genes not affected by germline knockdowns (*white*, *yb*). As expected, germline-specific depletion of Yb and White did not alter piRNA populations, whereas all other dsRNAs resulted in significantly reduced levels of germline-derived piRNAs (Fig. 2-15a). As a consequence of fewer piRNA reads, we observed slightly increased miRNA fractions in knockdowns of *armi*, *gasz*, *spn-E*, *aub* and *del*. PiRNAs (23- to 29-nt) uniquely matching the germline-exclusive *42AB* locus were dramatically reduced upon depletion of Armi (35.5x), GASZ (57.4x), Spn-E (26.9x), Aub (6.0x), or Del (13.3x) (Fig. 2-15a). The abundance of cluster-derived endo-siRNAs was similar in all knockdowns, with a marginal increase observed in cases where piRNAs were depleted.

Next, we analysed the effect of germline knockdown of *gasz* and *del* on the ping-pong amplification cycle. The frequencies of read pairs for *42AB*-derived piRNAs generated from opposing strands and overlapping by 10-nt, were calculated as a measure of ping-pong amplification (Fig. 2-15b). Knockdown of *armi* and *gasz* resulted in increased ping-pong signatures when compared to *white* or *yb*. In contrast, knockdown of *spn-E* and *aub*, factors essential to the ping-pong loop, as well as *del*, which affects proper localization of Aub, resulted in significantly reduced ping-pong pair frequencies.

We also inspected the impact of germline knockdowns on piRNAs corresponding to a set of 80 established *Drosophila* transposons (Fig. 2-15c). Based on previous data, these elements can be separated into those that dominate in somatic cells, intermediate transposons expressed in both lineages, and mobile elements predominantly active in germ cells (Malone *et al.*, 2009).

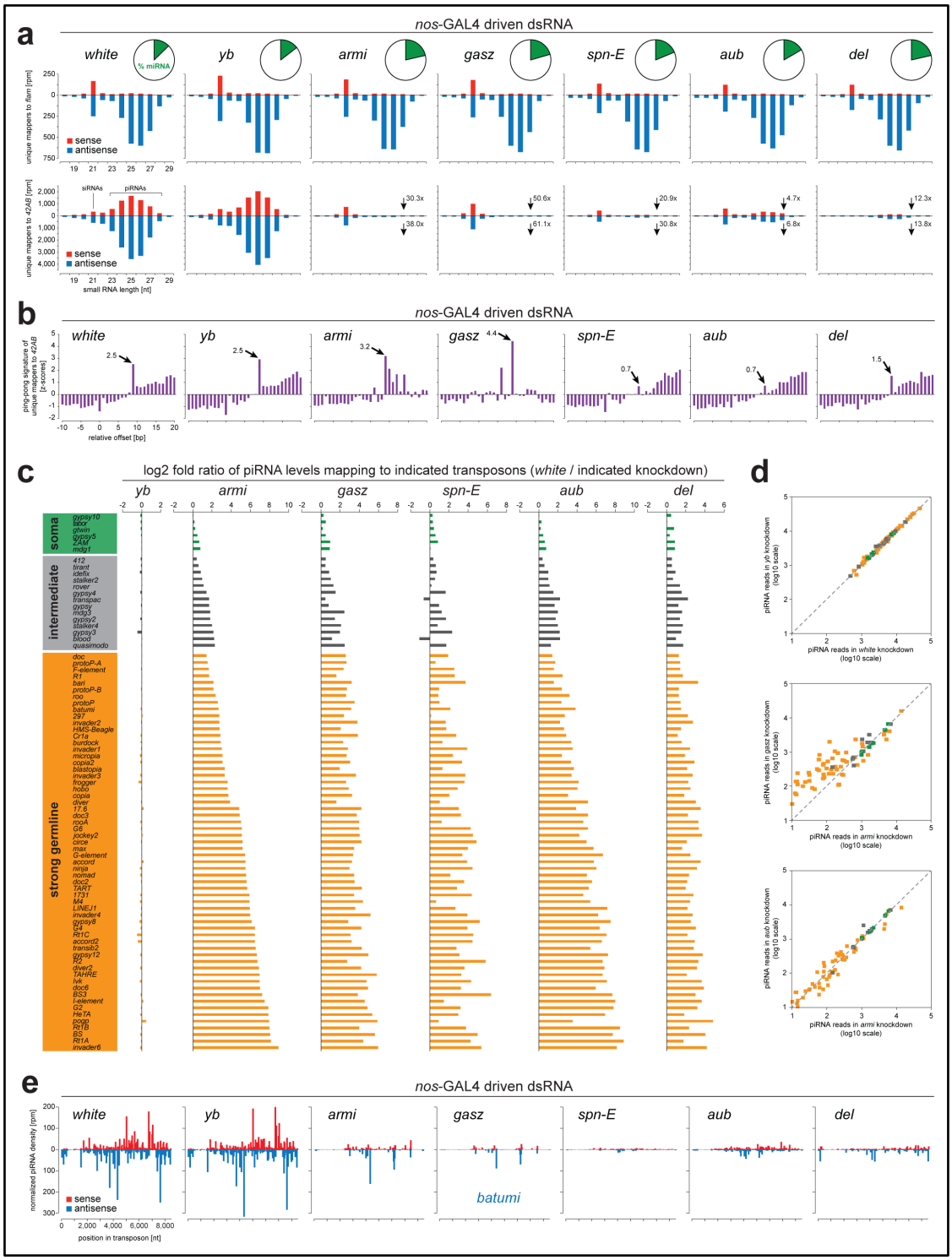


Fig. 2-15. Knockdown of *gasz* and *del* affects different steps of piRNA biogenesis. (a) The size distribution of small RNAs derived from each strand of the *flam* and *42AB* clusters are shown as histogram. The miRNA fraction (green) for the indicated libraries is highlighted in the cake diagrams. **(b)** Histograms showing the relative enrichment of piRNAs overlapping by the indicated number of nucleotides are plotted for *42AB*-derived sequences. The peak at position 9 (arrow) is suggestive of a ping-pong signature. The number represents the *z*-score for position 9. **(c)** Histograms showing the abundance of piRNAs mapping to soma dominant (green), intermediate (grey), or germline dominant (orange) transposons in *white* knockdowns compared to depletion of the indicated genes ($\log_2$ scale). **(d)** Scatter plots ($\log_{10}$ scale) comparing the piRNA abundance of the same transposons shown in (c). **(e)** Histograms of piRNAs mapping to the consensus sequence of the germline dominant *batumi* LTR transposon are shown for the indicated knockdowns. (Image and legend modified from: Czech, Preall *et al.*, in preparation)

Compared to *white* RNAi, germline depletion of Yb caused no significant changes in piRNAs derived from any transposon category (Fig. 2-15d, top), while all other knockdowns showed a variable but consistent reduction of piRNA levels from germline dominant elements (Fig. 2-15c). In contrast, piRNA levels of intermediate transposons were only mildly affected, and soma-enriched elements not changed significantly (less than 2-fold). The abundance of piRNAs matching to germline-enriched transposons was impaired to similar extent upon depletion of Armi and GASZ, suggesting related functions in primary piRNA biogenesis (Fig. 2-15d, middle). Furthermore, knockdown of *armi* and *aub* resulted in highly correlated reduction of piRNA levels (Fig. 2-15d, bottom), which is in agreement with Aub receiving inputs from primary biogenesis (Olivieri *et al.*, 2012) despite normal Aub localization upon *armi* knockdown (Fig. 2-14).

Density plot analysis for the LTR transposon *batumi*, which is predominantly active in the germline lineage, confirmed the interactions between the two candidates and the piRNA pathway (Fig. 2-15e). While knockdown of *yb* showed highly similar profiles to *white* controls, depletion of Armi or GASZ resulted in severely reduced piRNA levels, with some ping-pong-derived pairs persisting. Knockdown of *spn-E*, in contrast, caused dramatic loss of piRNA populations. Depletion of Aub resulted in significantly reduced piRNA abundances, with the remaining sequences likely associated with the other PIWI-clade proteins, Piwi and AGO3. Knockdown of *del* also caused a severe reduction of piRNAs matching the *batumi* transposon, but had distinct patterns to all other knockdowns.

Transposons are catalogued based on their sequence similarity, replication intermediates, and transposition strategy, and can be separated broadly into DNA elements and retrotransposons (Slotkin & Martienssen, 2007). The latter are further subdivided into LTR transposons (including *blood* and *burdock*), and non-LTR elements to which the LINE-like *HeTA* and *TAHRE* transposons belong. Given the diversity, it is likely that transposon-specific adaptions evolved for efficient silencing of each type of mobile element. To determine if the repression of certain transposons relies on particular genes, we compared de-silencing phenotypes (expressed as *z*-scores) between elements investigated in our primary, transcriptome-wide screen. As expected by their similar replication cycle, *z*-scores for *HeTA* and *TAHRE* are highly correlated amongst the top 500 genes scored as hits (R^2 = 0.62) (Fig. 2-16a). In contrast, correlations between *HeTA* and the LTR elements *burdock* (Fig. 2-16b) or *blood* (Fig. 2-16c) were much weaker (R^2 = 0.01 and R^2 = 0.08, respectively), probably reflecting important differences in the silencing determinants for these element classes. Strikingly, all 16 established piRNA pathway components showed strong transposon de-repression for all

four elements, with an average *z*-score of -3.66 (±0.57). Thus, knockdowns that robustly de-silence all four tested elements are highly enriched for core factors of the piRNA machinery. We can classify 17 new genes (including GASZ and del) that demonstrate consistent de-repression of all four transposons at levels similar to core components of the piRNA pathway, suggesting that they are likely to be central players in piRNA-mediated transposon silencing.

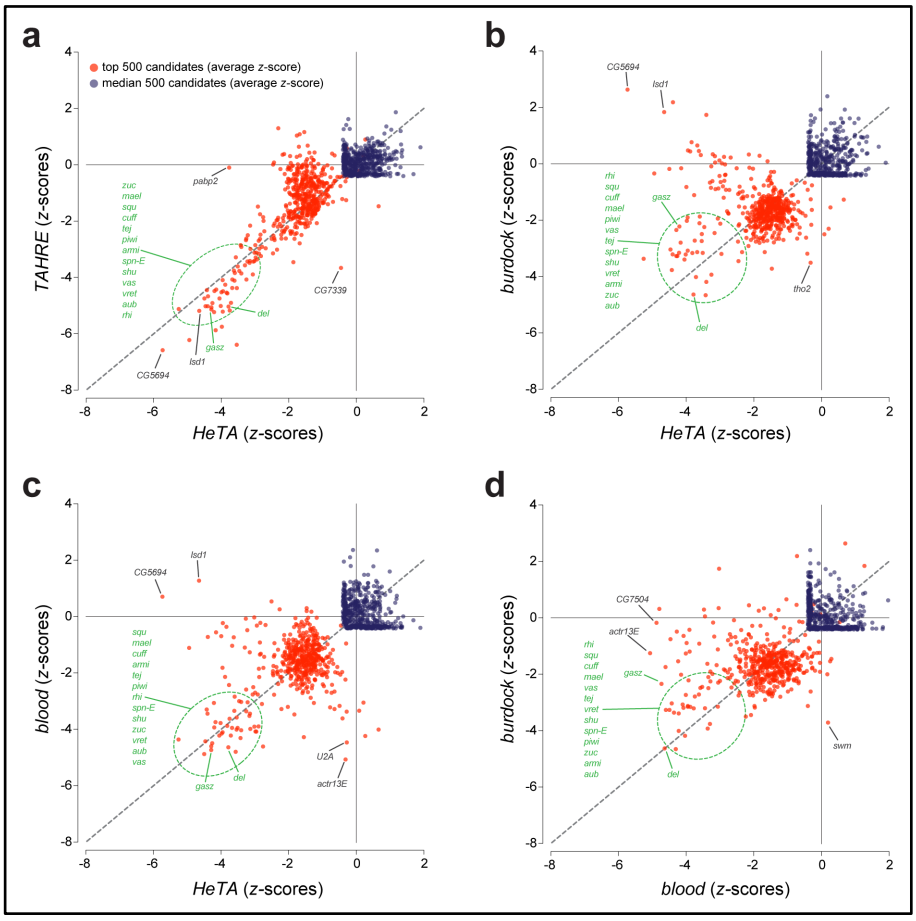


Fig. 2-16. Specific requirements for silencing of different transposon types. (a) Scatter plot comparing de-repression (as *z*-scores) for *HeTA* and *TAHRE* transposons (top 500 candidates from the primary screen are shown in red, median 500 candidates are indicated in blue). Known piRNA pathway components are highlighted in green. **(b)** Similar to (a), but *HeTA* de-repression is compared to the LTR element *burdock*. **(c)** Similar to (a), but comparing *HeTA* to *blood*. **(d)** Similar to (a) except levels of *blood* are compared to levels of *burdock*. (Image and legend modified from: Czech, Preall *et al.*, in preparation)

In addition to identifying novel components of the piRNA pathway, our screen also revealed an array of factors that participate in suppression of specific mobile elements or transposon families. For example, *CG5694* and *lsd1* (also known as *Su(var)3-3*), are critical for silencing both *HeTA* and *TAHRE* (Fig. 2-16a), but have no effect on *burdock* (Fig. 2-16b) or *blood* (Fig. 2-16c). *Lsd1* encodes for a histone demethylase, suggesting a specific requirement for certain chromatin modifiers in silencing LINE-like elements. In contrast, *Actr13E* was only necessary for the repression of *blood*, with little to no effect

on the other tested elements. Although showing little correlation (R^2 = 0.07), a comparison of *blood* and *burdock* repression uncovered only very few factors that dramatically compromised silencing in only of the two LTR transposons (Fig. 2-16d), suggesting that both elements share a common silencing mechanism beyond the core piRNA machinery.

Conclusions

Until now, no screen has been published that aims to systematically uncover novel components of the highly complex germline piRNA pathway. Here, we describe the results from RNAi screen encompassing the ovarian transcriptome, looking for novel piRNA pathway components and factors required for general transposon silencing. We identified 74 genes whose depletion severely compromised suppression of four transposons from different families. Importantly, all known piRNA machinery factors included in the screen (with the exception of Egg) were detected within the candidate list of 74 genes. We characterized two newly emerged candidates, GASZ and Del, which were both previously linked to germline development (Schüpbach & Wieschaus, 1991; Wehr *et al.*, 2006; Ma *et al.*, 2009). They appear to be involved at different steps in piRNA biogenesis, with GASZ likely acting in primary processing, whereas Del appears to play a role in the ping-pong amplification cycle.

In addition, we uncovered numerous genes that were required only for the suppression of specific transposons or element classes. Depletion of *CG5694* and *lsd1* ranked amongst the strongest candidates in de-repressing the LINE-like elements *HeTA* and *TAHRE*, but merely affected the LTR transposons *blood* and *burdock*. The identification of *lsd1*, which encodes a putative H3K4 histone demethylase, provides a tempting link between the nuclear piRNA pathway (represented by Piwi) and transcriptional regulation through histone modifications. It also raises the possibility that distinct chromatin modifiers (and their associated histone modifications) can contribute to the suppression of some, but not all, transposons.

Taken together, our results have introduced a significant number of novel factors to piRNA-mediated transposon regulation. These candidates now await further characterization at the molecular level, and will no doubt provide invaluable insights into the biological mechanisms underlying the piRNA pathway.

3 Discussion

3.1 The role of dsRBD proteins in dsRNA-derived small RNA biogenesis

An early model suggested two parallel, strictly separated biogenesis pathways for *Drosophila* miRNAs and siRNAs of exogenous origin. In this model, miRNAs are processed by Dcr1 and its co-factor Loqs, and associate with AGO1, whereas siRNAs depend on Dcr2 and the dsRBD protein R2D2 and load into AGO2. The discovery of endogenous siRNAs (Czech *et al.*, 2008; Ghildiyal *et al.*, 2008; Kawamura *et al.*, 2008; Okamura *et al.*, 2008a) and the uncovering of their biogenesis requirements (Czech *et al.*, 2008; Okamura *et al.*, 2008a) raised initial doubts on the model. While processing of endo-siRNAs from structured loci and viruses depended on Dcr2, they merely required R2D2 (Czech *et al.*, 2008; Okamura *et al.*, 2008a). Instead, both siRNA types surprisingly showed a strong dependence on the dsRBD protein Loqs. Data presented in this thesis extend these findings to other endo-siRNA classes, such as those derived from repeats and the *klarsicht* locus, and those produced from convergent transcription units (Zhou *et al.*, 2009). All endo-siRNA types were significantly reduced upon Loqs knockdown in S2 cells or in *loqs* mutant flies, whereas loss of R2D2 had little to no effect on their production (Zhou *et al.*, 2009).

Consistent with the requirement of Loqs and Dcr2 for endo-siRNA production, protein-protein interactions between these two proteins were identified (Czech *et al.*, 2008). Mining of quantitative proteomics uncovered a previously unknown Loqs isoform, Loqs-PD, in addition to the established variants Loqs-PA, Loqs-PB, and Loqs-PC (Förstemann *et al.*, 2005; Jiang *et al.*, 2005; Saito *et al.*, 2005; Park *et al.*, 2007; Hartig *et al.*, 2009; Zhou *et al.*, 2009). Under our conditions, isoforms PA and PC showed no impact on the processing of miRNAs or siRNAs, probably due to their more restrictive expression patterns. Indeed, a later report implicated Loqs-PA in miRNA biogenesis specifically in testes (Miyoshi *et al.*, 2010a), which parallels its expression. We found that the biogenesis of all endo-siRNA types solely depended on isoform Loqs-PD, and confirmed previous reports implying variant Loqs-PB in the production of miRNAs (Park *et al.*, 2007). Protein interactions between Dicer proteins and specific Loqs isoforms support these findings, with Loqs-PB showing the highest affinity for Dcr1, and Loqs-PD (as well as R2D2) strongly biased to interact with Dcr2 (Hartig *et al.*, 2009; Zhou *et al.*, 2009; Miyoshi *et al.*, 2010a). A subsequent report identified heterotrimeric complexes of Dcr2, Loqs-PD and R2D2, suggesting the simultaneous interaction of Dcr2 with both dsRBD proteins (Miyoshi *et al.*, 2010a).

Further studies confirmed our observation that R2D2 was dispensable for the production of endo- and exo-siRNA (Marques *et al.*, 2010; Miyoshi *et al.*, 2010a). R2D2 appears to act downstream of processing and instead assists Dcr2 in loading small RNA duplexes into AGO2. Loqs-PD, in contrast, is essential for efficient processing of dsRNAs into siRNA duplexes by Dcr2 *in vitro*. These results suggest sequential, non-redundant functions of Loqs-PD and R2D2 (Marques *et al.*, 2010; Miyoshi *et al.*, 2010a). Thus, both dsRBD co-factors are required for siRNA-mediated silencing, but they likely act at different steps within the siRNA pathway.

Although operating in a sequence-independent fashion, fly Dicer proteins show remarkable substrate specificity (Lee *et al.*, 2004b). Dcr1 prefers short hairpins that typically contain several mismatches. In contrast, Dcr2 usually operates on perfect or near-perfect dsRNA substrates. What parameters contribute to substrate specificity? Loqs isoforms and R2D2 differ in the number of dsRNA-binding domains. While Loqs-PB possesses three domains, Loqs-PD and R2D2 only feature two (Hartig *et al.*, 2009; Zhou *et al.*, 2009; Miyoshi *et al.*, 2010a), thus it is tempting to speculate that the different domain structures of these co-factors contribute to substrate specificity. This is unlikely to be true for Dcr1, as a recent report suggests that it recognizes the single-stranded loop region of pre-miRNAs via its amino-terminal helicase domain (Tsutsumi *et al.*, 2011), thereby excluding long dsRNAs, which lack single-stranded regions, as substrates. Biochemical studies also showed that Loqs-PB enhances the catalytic activity of Dcr1 (Jiang *et al.*, 2005; Saito *et al.*, 2005), without influencing substrate specificity (Tsutsumi *et al.*, 2011). In contrast, recombinant Dcr2 is capable to process pre-miRNAs, with binding of R2D2 and physiological concentrations of inorganic phosphate abolishing this activity (Cenik *et al.*, 2011). Loqs-PD and (to a lesser extent) R2D2 increase the affinity of Dcr2 for long dsRNA substrates *in vitro*, thus turning Dcr2 into a dsRNA-specific enzyme without modulating its catalytic activity (Cenik *et al.*, 2011).

Interestingly, the association of Dicer proteins with dsRBD partners is conserved in other organisms. The sole mammalian Dicer protein was shown to interact with two dsRNA-binding domain proteins, TAR RNA-binding protein (TRBP) and PKR activating protein (PACT) (Chendrimada *et al.*, 2005; Haase *et al.*, 2005; Lee *et al.*, 2006; Kok *et al.*, 2007), and is responsible for the production of both miRNAs and endo-siRNAs (Zhang *et al.*, 2002; Provost *et al.*, 2002; Watanabe *et al.*, 2008). To date, it is unclear whether TRBP and PACT carry out specialized functions in either pathway, with conflicting reports suggesting general roles such as substrate selection, enhancement of dicing or improved loading into one of the four Argonaute effector complexes. Similarly, *Caenorhabditis elegans* also encodes a single Dicer protein, DCR-1, involved in the

biogenesis of miRNAs and siRNAs (Ketting *et al.*, 2001). The dsRBD protein RDE-4, an ortholog of fly R2D2, associates with worm DCR-1 and specifically functions in the siRNA pathway, while being dispensable for miRNA biogenesis (Tabara *et al.*, 2002). RDE-4 shows a high affinity to bind long dsRNA, but not small RNA duplexes (Parker *et al.*, 2006). Therefore, RDE-4 is thought to aid the conversion of long dsRNA into siRNAs by recruiting DCR-1 to long dsRNA substrates. In *Arabidopsis*, which possesses several dsRBD proteins, HYPONASTIC LEAVES 1 (HYL1, an ortholog of fly Loqs) was reported to assist DCL1 in the production of miRNAs (Han *et al.*, 2004b; Vazquez *et al.*, 2004; Kurihara *et al.*, 2006). In addition, HYL1 is required for the biogenesis of nat-siRNAs in response to biotic (22-nt, in collaboration with DCL1) and abiotic stress (24-nt, as partner of DCL2) (Borsani *et al.*, 2005; Katiyar-Agarwal *et al.*, 2006), pointing to general roles in dsRNA substrate detection. Another dsRBD protein, DRB4, interacts with DCL4 to produce phased 21-nt duplexes in ta-siRNA biogenesis (Adenot *et al.*, 2006; Nakazawa *et al.*, 2007). The detailed molecular functions of HYL1 and DRB4, and the roles of other plant dsRBD proteins in small RNA pathways, remain to be investigated.

My thesis research revealed discrete functions for specific *Drosophila* dsRBD proteins within small RNA pathways. In summary, our data and reports from other model organisms delineate different modes of action for distinct dsRBD proteins, which can contribute to substrate specificity or enhancement of catalysis potentially through altered structural conformations. The details of their molecular mechanisms will need to be resolved with future experiments to better characterize their influence on small RNA biogenesis.

3.2 A hierarchy of determinants ensures proper small RNA loading

The biogenesis of *Drosophila* miRNAs and siRNAs was proposed to occur through separate, unconnected pathways, though this model relied on the analysis of a limited number of sequences that were available at that time. The strict association of miRNAs with AGO1 and siRNAs with AGO2, respectively, suggested coupled processing and loading. Evidence against such mechanism was uncovered with the report of a microRNA, *miR-277*, which was preferentially bound to AGO2 (Förstemann *et al.*, 2007). Instead of coupled biogenesis and loading, the data suggested that small RNA duplexes dissociate from Dicer proteins following their production, and are subsequently distributed into Argonaute proteins via a step termed “sorting”, with perfectly matched duplexes entering AGO2, whereas mismatched duplexes associating with AGO1 (Förstemann *et al.*, 2007; Tomari *et al.*, 2007).

The availability of high throughput sequencing technologies has enabled us to profile the full repertoire of endogenous small RNA populations of AGO1 and AGO2, and allowed us to test existing hypotheses on small RNA sorting (Czech *et al.*, 2009). Profiling of AGO2-associated RNA species revealed a strong association with various endo-siRNA types. In addition, we detected multiple AGO2-loaded miRNAs (including *miR-277*), of which the majority corresponded to the miR* strand. AGO1, in contrast, was almost exclusively occupied by miR strands, with only very few sequences detected that match to endo-siRNAs (e.g., *esi-2.3*). Our results have two major implications: first, coupled processing and loading of miRNAs into AGO1 and siRNAs into AGO2 complexes, respectively, does not occur in flies. Instead, miR:miR* and siRNA duplexes (consisting of guide and passenger strands) appear to be freely sorted independent of their upstream biogenesis events. Second, loading of miR strands into AGO1 and miR* strands into AGO2 suggests the assessment of each individual strand within a duplex, therefore extending the cells repertoire to pair small RNAs with Argonaute effectors. As the association of small RNAs with Argonaute complexes is non-random, intricate mechanisms must underlie proper sorting.

Understanding the small RNA sorting process would be advantageous in generating potent silencing triggers for therapeutics and research. By mining our high throughput data, we extracted determinants that mediate the sorting of small RNA duplexes into specific Argonaute proteins (Czech *et al.*, 2009). Experimental testing of these determinants indicated that they follow a hierarchy, and enabled us to derive sorting rules to predict the fate of a small RNA. The dominant determinant is the duplex structure, especially at the central region (nucleotides 9 and 10 counted from the 5' end), with unpaired bases resulting in association with AGO1 and duplexes without mismatches favouring AGO2 (Tomari *et al.*, 2007; Czech *et al.*, 2009; Okamura *et al.*, 2009; Ghildiyal *et al.*, 2010). In addition, AGO2-associated duplexes showed stronger pairing at their 5' end (of the incorporated strand), compared to those binding to AGO1 (Czech *et al.*, 2009; Okamura *et al.*, 2009; Ghildiyal *et al.*, 2010). For imperfect duplexes, central mismatches impact strand choice, with unpaired bases at position 9 and 10 marking a strand to be retained. In contrast, thermodynamic asymmetry is the major determinant for strand selection of perfect dsRNA duplexes, as originally proposed (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003). Although sequences bound to *Drosophila* Argonaute proteins feature pronounced biases for terminal nucleotides (1U for AGO1 and 1C for AGO2) (Czech *et al.*, 2008; Ghildiyal *et al.*, 2008; Czech *et al.*, 2009; Okamura *et al.*, 2009; Ghildiyal *et al.*, 2010), their respective contribution to sorting and strand selection appears to be minimal in our experiments (Czech *et al.*, 2009). This is in

contrast to other reports that suggest a stronger impact of the 5' terminal nucleotide (Okamura *et al.*, 2009; Ghildiyal *et al.*, 2010), though this discrepancy is perhaps due to different sets of duplexes that were investigated. Structurally, the 5' nucleotide of small RNAs is anchored in a binding pocket, which is formed by the mid domain of Argonaute proteins (Ma *et al.*, 2005; Parker *et al.*, 2005, Wang *et al.*, 2009), with recent work providing evidence that mid domains discriminate between terminal nucleotides (Frank *et al.*, 2010). Human AGO2, the homolog of fly AGO1, shows a preference for terminal U (or A) nucleotides, while excluding C or G nucleotides through a nucleotide specificity loop (Frank *et al.*, 2010). Interestingly, this nucleotide specificity loop is absent from *Drosophila* AGO2, potentially explaining differential nucleotide biases, with details of its contribution to sorting in flies awaiting further investigation.

A number of endo-siRNAs from structured loci, containing several mismatches (similar to *esi-2.3*) and prominent 1U bias, was found to accumulate in AGO2, rather than AGO1 as predicted (Ameres *et al.*, 2011). Interestingly, these endo-siRNAs were loaded into AGO1 *in vitro*, but are unstable in AGO1 *in vivo* due to target-directed small RNA destruction, via the tailing and trimming pathway (Ameres *et al.*, 2010; Ameres *et al.*, 2011). In contrast, AGO2-loaded endo-siRNAs were protected from degradation by their 2'-*O*-methylated 3' termini (Horwich *et al.*, 2007; Saito *et al.*, 2007), and therefore stabilized *in vivo*. Thus, mechanisms that operate after the sorting and loading steps are actively influencing the steady-state distribution of small RNAs, thereby adding further intricacy to (studies of) the sorting pathways.

Another layer of complexity in small RNA sorting emerged from studies of siRNA responses to latent virus infections (Aliyari *et al.*, 2008; Flynt *et al.*, 2009). Although produced by Dcr2/Loqs-PD complexes, a significant proportion of virus-derived siRNAs was not loaded into AGO2, as shown by the partial absence of modified 3' termini (Aliyari *et al.*, 2008; Czech *et al.*, 2009; Flynt *et al.*, 2009), and consequently lack of dependency on R2D2 (Czech *et al.*, 2008). Furthermore, deep sequencing detected no interaction of unmodified virus-siRNAs with AGO1, suggesting that these siRNAs were unloaded and instead free-floating (Czech *et al.*, 2008; Flynt *et al.*, 2009). It is unclear why these siRNA duplexes, although being of perfect dsRNA nature, escape the loading machinery. A possible explanation is that the host cell aims to destroy replication intermediates of the viral genome, without disrupting endogenous small RNA pools and thereby maintaining optimized Argonaute protein occupancy.

Profiling of small RNAs from distinct *Drosophila* tissues or developmental stages often detected extreme differences in ratios between miR and miR* strands (Ro *et al.*, 2007; Okamura *et al.*, 2008b; Ghildiyal *et al.*, 2010; Benjamin Czech & Greg Hannon,

unpublished). While some of these variations can be attributed to varying expression levels of the loading machineries or Argonaute proteins themselves, tissue-specific expression of additional factors affecting strand choice is also possible. Strikingly, recent work uncovered the existence of sequence signals located outside the duplex that determine strand choice (Griffiths-Jones *et al.*, 2011). When the primary transcripts of *miR-10* from *Tribolium castaneum* and *Drosophila melanogaster*, respectively, were expressed in S2 cells (lacking detectable endogenous levels of *miR-10*), both pri-miRNAs were processed into identical miR:miR* duplexes, but resulted in the accumulation of different strands. Duplexes from the *Tribolium* pri-miRNA selectively stabilized the strand from the 5' arm, whereas duplexes that originated from *Drosophila* precursors predominantly retained the strand from the 3' arm (Griffiths-Jones *et al.*, 2011). Thus, there are additional parameters that lie outside the small RNA duplex that are capable in affecting strand selection. It will be interesting to identify the molecular nature of these signals and determine if these are unique adaptions or part of a broader phenomenon.

What do we know about small RNA sorting in other organisms? In contrast to flies, all four mammalian AGO-clade proteins (AGO1 through AGO4) appear to be loaded with highly similar miRNA populations without biases for structure or 5' nucleotide identities (Liu *et al.*, 2004; Meister *et al.*, 2004; Yoda *et al.*, 2010), arguing for the absence of stringent sorting pathways. It has not been investigated yet if endo-siRNAs show specific loading patterns, for instance, into the only slicing-competent AGO-clade member, AGO2. In addition, *miR-451* was shown to require the catalytic activity of AGO2 for its maturation (Cheloufi *et al.*, 2010; Cifuentes *et al.*, 2010; Yang *et al.*, 2010), but whether its precursor is sorted specifically into AGO2 remains elusive.

C. elegans seems to possess sophisticated sorting machineries, as miRNAs with central mismatches are loaded into ALG1 and ALG2, whereas siRNAs from perfect dsRNA associate with RDE-1 (Steiner *et al.*, 2007; Jannot *et al.*, 2008). The distribution of miRNAs and siRNAs is influenced by protein factors, with the details that determine distinct sorting, like the contribution of nucleotide biases, not extensively studied. In contrast to miRNAs and primary siRNAs, RdRPs directly synthesize secondary siRNAs (Pak & Fire, 2007; Sijen *et al.*, 2007). As they are of single-stranded origin, no discrimination between guide and passenger strands is required. However, loading of single-stranded RNAs into WAGO complexes takes place through unknown mechanisms, with their 5' triphosphate termini potentially contributing via interactions with WAGO mid domain.

Small RNA sorting is particularly important in plants, with different dsRNA-derived small RNAs (miRNAs and several siRNA classes) sorted into multiple Argonaute proteins. *Arabidopsis* encodes ten AGO-clade members, which show different nucleotide biases and preferences for different small RNA lengths. MiRNAs, which are ~21-nt in length and processed by DCL1, are typically associated with AGO1, whereas DCL3-produced hc-siRNAs are ~23-nt long and preferentially load into AGO4. Profiling of small RNAs associated with different Argonautes identified strong biases for 5' terminal nucleotides (Mi *et al.*, 2008; Montgomery *et al.*, 2008; Takeda *et al.*, 2008): AGO1 was mainly loaded with sequences with terminal U, AGO2 and AGO4 preferred sequences with 5' A, and AGO5 showed a bias towards a terminal C. Moreover, altering the terminal nucleotides redirected small RNAs into different Argonaute complexes, indicating that the 5' nucleotide is a dominant determinant for sorting in plants (Mi *et al.*, 2008; Montgomery *et al.*, 2008). A subset of these biases was recently attributed to diverse nucleotide specificity loops residing in mid domains of the different plant Argonaute proteins AGO1, AGO2, and AGO5 (Frank *et al.*, 2012). However, there are numerous exceptions to these simple rules. For instance, mature *miR-390* possesses a terminal A, which would predict association with AGO2, but instead exclusively occupies AGO7 (Montgomery *et al.*, 2008), with changes of the 5' nucleotide incapable to redirect *miR-390* to other Argonautes. Thus, additional determinants that may include the duplex structure, degree of base pairing, and thermodynamic properties likely contribute to sorting and strand selection in plants (Eamens *et al.*, 2009), with details yet to be examined.

The work in this thesis contributed to uncover the complex sorting mechanisms that have evolved in *Drosophila melanogaster* to ensure proper pairing of duplexes with their Argonaute partners. Furthermore, the knowledge gained from this study has enabled us to co-opt the miRNA pathway for transgenic RNAi, and led to the development of shRNAs as potent silencing triggers in *Drosophila* (Ni *et al.*, 2011). In summary, much progress has been made in the past years, largely aided by high throughput sequencing, but additional biochemical and structural studies will be necessary to disentangle the remaining secrets this pathway holds.

3.3 Towards an understanding of piRNA biogenesis

Since the discovery and recognition of piRNAs as a major small RNA class, numerous genes have been linked to the piRNA pathway. The vast majority of these factors were originally uncovered in classic forward genetic screens looking for mutations that affect oogenesis, female fertility or spindle formation (Schüpbach & Wieschaus, 1989; Schüpbach & Wieschaus, 1991; Gonzalez-Reyes *et al.*, 1997). Reverse genetic

approaches to identify novel piRNA pathway components became possible with the recent availability of tools like the ovarian somatic sheet (OSS) cell line (Niki *et al.*, 2006; Lau *et al.*, 2009; Saito *et al.*, 2009), or RNAi systems that allow gene knockdowns in somatic and germline cells of the fly ovary (e.g., shRNAs, dsRNAs facilitated by Dcr2 overexpression) (Olivieri *et al.*, 2010; Handler *et al.*, 2011; Ni *et al.*, 2011; Wang & Elgin, 2011; Preall *et al.*, 2012). However, systematic RNAi screens have yet to be utilized to uncover missing piRNA biogenesis factors on a genome-wide scale. While we are beginning to understand key concepts of piRNA silencing, many aspects of the pathway still remain enigmatic. The identification of as-yet unidentified key factors (like nucleases) will be critical to elucidate the molecular mechanisms behind piRNA biogenesis.

To uncover the entire repertoire of piRNA pathway components, we carried out an unbiased, transcriptome-wide RNAi screen in germline cells of the female *Drosophila* ovary, and probed for transposon de-repression. Data presented in this thesis identified 74 genes that caused significant loss of silencing of four distinct transposons. Only 16 of these represent known piRNA pathway components, with the remaining candidates comprising new putative piRNA factors (see Results 2.2.2). Downstream assays enabled us to distinguish factors required for primary or secondary piRNA biogenesis from those affecting piRNA-mediated silencing. We characterized two new factors that act in different biogenesis pathways. *GASZ* compromises primary biogenesis in somatic and germline cells, as evidenced by reduced piRNA levels and Piwi redistribution to the cytoplasm upon GASZ depletion. In contrast, *deadlock* (*del*) specifically functions in germline cells and is essential for a proper ping-pong cycle, with the precise molecular roles of these two biogenesis factors yet to be elucidated. Additional genome-wide RNAi screens have been carried out in OSS cells (Muerdter, Guzzardo *et al.*, in preparation) and follicle cells of the female ovary (Julius Brennecke, personal communication). These will complement our efforts in identifying the majority of missing pathway components, particularly the nucleases involved in biogenesis. Further efforts to characterize emerging candidates will help to shed light on piRNA pathway functions.

I believe that the next key breakthrough will be the identification of the enzymatic activities responsible for 5' and 3' end formation. Current models suggest the 5' end generation to be carried out by an endonuclease specific to single-stranded RNA, with Zuc being a promising candidate (Malone *et al.*, 2009; Haase *et al.*, 2010; Olivieri *et al.*, 2010; Saito *et al.*, 2010; Ipsaro *et al.*, *in press*). Recent studies using a silkworm (*Bombyx mori*) cell line to probe the 3' end formation of piRNA intermediates bound by PIWI-clade proteins in an cell-free system identified an exonucleolytic processing activity, termed "trimmer" (Kawaoka *et al.*, 2011), with the molecular identity of this enzyme yet to

be uncovered. The use of *in vitro* systems has significantly accelerated the progress made in understanding the molecular mechanisms of miRNA and siRNA processing (Tuschl *et al.*, 1999; Hammond *et al.*, 2000; Zamore *et al.*, 2000). The availability of cell lines that recapitulate primary piRNA biogenesis (*Drosophila* OSS cells) (Niki *et al.*, 2006; Lau *et al.*, 2009; Saito *et al.*, 2009) or active ping-pong amplification (*Bombyx mori* cells) (Kawaoka *et al.*, 2009) should facilitate the establishment of *in vitro* systems for specific aspects of piRNA biology. However, the pronounced dependence of piRNA biogenesis on scaffolding Tudor family proteins implies that complex cellular structures, like nuage or Yb-bodies, might be essential to process and load piRNAs efficiently, potentially by bringing together all necessary protein components as well as RNA substrates. When coupled with the large number of critical piRNA factors (~20 in soma and germline combined), an *in vitro* system that recapitulates all aspects of piRNA biogenesis might be difficult to develop.

We and others have recently identified *shutdown* (*shu*), which was also recovered in our screen, as an important component of the piRNA pathway (Olivieri *et al.*, 2012; Preall *et al.*, 2012). Shu is unique, as it is the only known factor that appears to affect primary biogenesis and the ping-pong amplification loop, with loss of Shu dramatically impairing small RNA populations of all three fly PIWI-clade proteins (Olivieri *et al.*, 2012; Preall *et al.*, 2012). Given these phenotypes and the high similarity of Shu to conserved co-chaperones, we speculate that Shu plays a role in loading of piRNA precursors into PIWI-clade proteins, potentially aided through interaction with Hsp90. Interestingly, primary piRNAs are only loaded into Piwi and Aub, but not AGO3 (Olivieri *et al.*, 2012). The detailed molecular mechanisms that pair single-stranded precursor RNAs with Piwi-clade proteins remain enigmatic. Since loading of single-stranded RNAs does not require the discrimination between two strands (as described for miRNA or siRNA duplexes), other parameters must determine the sorting and incorporation into PIWI family members. A related question is how asymmetry is established within specific PIWI-clade protein populations. Although the exclusive association of primary piRNAs with Piwi and Aub might account for the pronounced antisense bias of these two proteins (Olivieri *et al.*, 2012), mechanisms excluding sense transcripts from associating with these factors and instead funnels them into AGO3 remain to be resolved.

The determinants that flag transcripts for processing into mature piRNAs are entirely unknown. One hypothesis proposed that the genomic location of piRNA clusters in heterochromatin plays a role in tagging transcripts for piRNA biogenesis, which is supported indirectly by the requirement of heterochromatin marks for piRNA cluster transcription (Rangan *et al.*, 2011). However, artificially introduced piRNA clusters

inserted in euchromatic regions still produced piRNAs across various species (Kawaoka *et al.*, 2012; Muerdter *et al.*, 2012), arguing against this model. Moreover, the discovery of piRNAs derived from protein-coding transcripts in OSS cells suggests a more complex mechanism (Robine *et al.*, 2009; Saito *et al.*, 2009). Genic piRNAs are mainly derived from exons and 3' UTRs (therefore predominantly in sense orientation to mRNAs), but only a subset of transcripts acts as substrates for piRNA generation, with little correlation to their cellular abundance (Robine *et al.*, 2009). The signals that funnel cellular transcripts, especially protein-coding ones, into the piRNA biogenesis machinery are yet to be elucidated.

There is evidence that transcripts can be targeted preferentially for piRNA biogenesis. Most germline clusters are transcribed from both strands, which can lead to the formation of dsRNA that is processed via Dcr2 into siRNAs. However, these siRNAs generated from clusters accumulate at significantly lower levels than piRNAs (Czech *et al.*, 2008). Interestingly, knockdown of several piRNA pathway components in the germline leads to increased levels of siRNAs corresponding to piRNA cluster transcripts, while control knockdowns lack elevated siRNA levels (Benjamin Czech & Greg Hannon, unpublished). These results cannot be explained simply as a consequence of Dcr2 overexpression and rather suggest an active mechanism that prevents the siRNA pathway from processing cluster transcripts and instead makes them available for piRNA biogenesis, likely by stabilizing single-stranded cluster transcripts. Thus, the flagging of cluster transcripts and protein-coding mRNAs for the production of piRNAs remains an open question. Binding of specific proteins to nascent cluster transcripts could likely stabilize single-stranded conformations, and subsequently guide them to the piRNA processing sites, Yb-bodies and nuage. Alternatively, cluster transcripts could selectively be tagged by one of myriad modifications found in ncRNAs and mRNAs (e.g., 5-methylcytosine, m^5C; or N(6)-methyladenosine, m^6A) (Dominissini *et al.*, 2012; Squires *et al.*, 2012). Further studies are required to elucidate the molecular mechanisms that guide transcripts to the piRNA biogenesis machinery.

Taken together, my thesis research helped to identify numerous additional factors required for piRNA processing. Further efforts to characterize the molecular functions of emerging candidates will help to place them into a model and shed light on the production of piRNAs. I expect significant knowledge gains from the combination of structural and biochemical approaches, along with existing genetic data, which will ultimately lead to a detailed understanding of the piRNA biogenesis.

3.4 PiRNA-mediated silencing: transposons and beyond

The piRNA system comprises genetically inheritable memory of transposon activity in form of piRNA clusters, and specific amplification against acute mobilization of repetitive elements via the ping-pong cycle. Thus, it can be seen as an RNA-based adaptive immune system targeting deleterious genetic elements (Brennecke *et al.*, 2007). In *Drosophila*, piRNA clusters contain diverse transposon fragments resulting from past exposure to mobile elements (Brennecke *et al.*, 2007). Transposon activity is therefore evolutionary recorded by transposition events into piRNA clusters, but also contributes to primary piRNA pools, which enable cells to detect and silence parasitic nucleic acids. However, the hypothesis that piRNA clusters specifically attract mobile elements is still a matter of controversial debate (Brennecke *et al.*, 2007), with one report claiming the *X-TAS* cluster as transposition hotspot of *P*-elements (Karpen & Spradling, 1992), whereas another work suggesting random or only moderately biased transposon integration into clusters (Khurana *et al.*, 2011). Either way, both hypotheses imply that once a novel transposon is integrated into a piRNA cluster, trans-generational memory is established and cells are enabled to silence other mobile element copies *in trans*. Strikingly, piRNA clusters can easily adapt to novel foreign sequences and produce mature silencing triggers, as evidenced by artificial piRNAs against *EGFP* when inserted into piRNA clusters in fly, mouse, and silkworm (Kawaoka *et al.*, 2012; Muerdter *et al.*, 2012). Whether artificial piRNAs can also trigger a ping-pong response remains to be tested, as current studies did not provide the required second strand required as substrate for amplification. Studies of the hybrid dysgenesis phenomena point towards the requirement of piRNA amplification for efficient transposon silencing (Brennecke *et al.*, 2008; Khurana *et al.*, 2011). In this context, it should also be noted that piRNA clusters not only serve as source of primary piRNAs by providing memory of transposon history, but also function as relays stations in the ping-pong cycle through supply of RNA substrates (Brennecke *et al.*, 2007).

Intriguingly, a similar but non-homologous concept is used by prokaryotes, which use small ncRNAs to defend their genomes from foreign nucleic acids like phages or plasmids (Horvath & Barrangou, 2010; Wiedenheft *et al.*, 2012). Here, fragments of invading nucleic acids are actively incorporated into specific genomic loci, called “clustered regularly interspaced short palindromic repeat” (CRISPR) arrays, with the spacers corresponding to DNA fragments of invading phages or plasmids (Barrangou *et al.*, 2007; Marraffini & Sontheimer, 2008). Following transcription, precursors are processed into small RNAs, called CRISPR RNAs (crRNAs), which associate with

CRISPR-associated (*Cas*) proteins (Brouns *et al.*, 2008; Carte *et al.*, 2008). The crRNAs guide *Cas* proteins to target nucleic acids (mainly DNA, with some reports also suggesting RNA), which results in their degradation through cleavage (Marraffini & Sontheimer, 2008; Hale *et al.*, 2009; Garneau *et al.*, 2010). Thus, although the CRISPR system is not an RNAi-related mechanism (no involvement of Argonaute proteins), it is mechanistically very similar to the piRNA system in protecting host genomes from parasitic nucleic acids through small RNAs, thereby showing conceptual similarities to adaptive immune systems.

While the repressive functions of repeat-derived piRNAs are well recognized, the significance of non-repeat matching piRNAs remains obscure. The selective production of genic piRNAs from certain precursor mRNAs in *Drosophila* ovaries, mouse testes and *Xenopus* eggs, and their enrichment for specific gene ontology categories suggest potential regulatory purposes (Robine *et al.*, 2009). Only a subset of murine piRNAs is derived from transposable elements, dependent on the developmental stage and involved PIWI-clade protein, with MIWI-associated piRNAs expressed at the pachytene stage of meiosis depleted of repetitive sequences and their targets still unknown (Aravin *et al.*, 2006; Girard *et al.*, 2006). Similarly, the vast majority of nematode piRNAs (also known as 21U RNAs) are not enriched for transposon sequences and have no known function (Ruby *et al.*, 2006; Batista *et al.*, 2008).

Recent studies in *C. elegans* propose a major role for piRNAs in the determination of “self” and “non-self” that persists over generations, with some piRNAs protecting endogenously expressed genes from undesired silencing, and other piRNAs detecting and repressing foreign nucleic acids (Ashe *et al.*, 2012; Bagijn *et al.*, 2012; Lee *et al.*, 2012; Shirayama *et al.*, 2012). Furthermore, piRNAs trigger the production of secondary siRNAs, which bind to dedicated nuclear WAGO proteins and mediate the transcriptional repression of mobile elements, aided by associated chromatin factors (Ashe *et al.*, 2012; Shirayama *et al.*, 2012). Thus, the *C. elegans* piRNA pathway cooperates with the endo-siRNA pathway to achieve efficient trans-generational repression of transposons. Interestingly, with the lack of ping-pong signatures of nematode piRNAs, engagement of the siRNA pathway might provide an elegant amplification mechanism for silencing triggers, likely required for efficient repression of active transposon threats. Whether a similar mechanistic coupling between endo-siRNA and piRNA pathways takes place in other species has not been studied in detail. However, numerous fly transposons are controlled and silenced to varying degrees by piRNA and endo-siRNAs (Czech *et al.*, 2008), suggesting an indirect cooperation between both pathways. Alternatively, both pathways could co-exist and serve as backup mechanisms in the event where one

system fails. Remarkably, studies of the small RNA populations of the parasitic nematode *Ascaris* (and the related parasite *Brugia*) uncovered the surprising loss of piRNAs and corresponding pathway components, while a myriad of miRNAs and siRNAs and their corresponding biogenesis machineries were detected (Wang *et al.*, 2011). Transposon silencing in these organisms appears to solely rely on secondary siRNAs, demonstrating an extreme example of the flexibility and adaption of small RNA pathways.

How do piRNAs perform their downstream function mechanistically and what factors are involved in piRNA-mediated silencing? Our current knowledge suggests several machineries depending on the PIWI-clade protein involved, with fly Aub/AGO3 and Piwi engaging different silencing mechanisms. Aub and AGO3 localize to nuage, where they actively participate in the ping-pong amplification loop, and are thought to cleave targeted transposons utilizing their slicer activity. Notably, Aub-associated piRNAs degrade targets without full sequence complementarity, as shown for interaction between *AT-chX*-derived piRNAs and *vasa* in fly testes (Nishida *et al.*, 2007). Most factors that affect silencing through Aub and AGO3 are nuage components, though proteins regulating their function awaits identification. Squash (Squ), which codes for a putative nuclease, likely acts at the effector step, as loss of Squ results in dramatic de-repression of transposons, while piRNA biogenesis appears unaffected (Pane *et al.*, 2007; Malone *et al.*, 2009; Haase *et al.*, 2010). Squ could play a role in the degradation of Aub- and AGO3-cleaved transposon transcripts, but these ideas remain to be tested. Piwi is localized to the nucleus, and its molecular silencing mechanism is currently unknown. Recent reports found the nuclear localization of Piwi essential for proper transposon suppression (Saito *et al.*, 2009; Saito *et al.*, 2010; Klenov *et al.*, 2011). Piwi's slicer activity is not required, however, as cells with catalytically inactive Piwi show normal transposon levels (Saito *et al.*, 2010), suggesting that Piwi functions through transcriptional silencing potentially by inducing chromatin modifications.

Similar to flies, the catalytic activities of mouse MILI and MIWI are essential for proper transposon silencing and fertility (De Fazio *et al.*, 2011; Reuter *et al.*, 2011). However, functions of murine piRNAs extend beyond post-transcriptional silencing, as pre-pachytene piRNAs participate in the establishment of epigenetic modifications that manifest in DNA methylation (Aravin *et al.*, 2007; Kuramochi-Miyagawa *et al.*, 2008). Reduced DNA methylation in *mili* and *miwi2* mutants correlates with increased mobility of several LINE and LTR elements, suggesting a failure to establish *de novo* methylation at transposon integrations during spermatogenesis (Aravin *et al.*, 2007; Kuramochi-Miyagawa *et al.*, 2008). This hypothesis is further supported by the partial overlap of the

expression and subcellular localization of MIWI2 at the developmental stage when *de novo* methylation takes place (Aravin *et al.*, 2007; Kuramochi-Miyagawa *et al.*, 2008), but the molecular details remain to be investigated.

Work presented within this thesis uncovered several new factors involved in piRNA-mediated silencing. Future work aimed to characterize the mode of action of these candidates in detail will be key to a better understanding of piRNA effector mechanisms, in particular at the molecular level.

In summary, my thesis work helped to shed light on the biogenesis, sorting and effector mechanisms of small RNAs in *Drosophila*. Taken together, tremendous progress of all aspects of small RNA biology has been made in recent years, with many intriguing avenues left for future research. Given the current pace, I envision a detailed mechanistic understanding of small RNA pathways to be within reach.

4 References

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouche, N., Gasciolli, V., and Vaucheret, H. 2006. DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. *Curr Biol* **16**(9): 927-932.

Aliyari, R., Wu, Q., Li, H.W., Wang, X.H., Li, F., Green, L.D., Han, C.S., Li, W.X., and Ding, S.W. 2008. Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in Drosophila. *Cell Host Microbe* **4**(4): 387-397.

Allan, R.K. and Ratajczak, T. 2011. Versatile TPR domains accommodate different modes of target protein recognition and function. *Cell Stress Chaperones* **16**(4): 353-367.

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**(5985): 1534-1539.

Ameres, S.L., Hung, J.H., Xu, J., Weng, Z., and Zamore, P.D. 2011. Target RNA-directed tailing and trimming purifies the sorting of endo-siRNAs between the two Drosophila Argonaute proteins. *Rna* **17**(1): 54-63.

Anand, A. and Kai, T. 2012. The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in Drosophila. *Embo J* **31**(4): 870-882.

Anne, J. and Mechler, B.M. 2005. Valois, a component of the nuage and pole plasm, is involved in assembly of these structures, and binds to Tudor and the methyltransferase Capsuleen. *Development* **132**(9): 2167-2177.

Anne, J., Ollo, R., Ephrussi, A., and Mechler, B.M. 2007. Arginine methyltransferase Capsuleen is essential for methylation of spliceosomal Sm proteins and germ cell formation in Drosophila. *Development* **134**(1): 137-146.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J.J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., and Tuschl, T. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**(7099): 203-207.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**(6): 785-799.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**(5825): 744-747.

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., Pintacuda, G., Sakaguchi, A., Sarkies, P., Ahmed, S., and Miska, E.A. 2012. piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. *Cell* **150**(1): 88-99.

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**(6094): 574-578.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**(5819): 1709-1712.

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**(2): 281-297.

Bartel, D.P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**(2): 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., Conte, D., Jr., Luo, S., Schroth, G.P., Carrington, J.C., Bartel, D.P., and Mello, C.C. 2008. PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol Cell* **31**(1): 67-78.

Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., and Lai, E.C. 2007. Mammalian mirtron genes. *Mol Cell* **28**(2): 328-336.

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**(6818): 363-366.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. 1998. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *Embo J* **17**(1): 170-180.

Bohnsack, M.T., Czaplinski, K., and Gorlich, D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* **10**(2): 185-191.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* **123**(7): 1279-1291.

Boswell, R.E. and Mahowald, A.P. 1985. tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. *Cell* **43**(1): 97-104.

Brand, A.H. and Perrimon, N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**(2): 401-415.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**(6): 1089-1103.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**(1): 25-36.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**(5906): 1387-1392.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L., and Voinnet, O. 2008. Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**(5880): 1185-1190.

Brodersen, P. and Voinnet, O. 2009. Revisiting the principles of microRNA target recognition and mode of action. *Nat Rev Mol Cell Biol* **10**(2): 141-148.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**(5891): 960-964.

Bushati, N. and Cohen, S.M. 2007. microRNA functions. *Annu Rev Cell Dev Biol* **23**: 175-205.

Bussing, I., Yang, J.S., Lai, E.C., and Grosshans, H. 2010. The nuclear export receptor XPO-1 supports primary miRNA processing in C. elegans and Drosophila. *Embo J* **29**(11): 1830-1839.

Cai, X., Hagedorn, C.H., and Cullen, B.R. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* **10**(12): 1957-1966.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12**(4): 503-514.

Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* **22**(24): 3489-3496.

Cenik, E.S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka Hall, T.M., and Zamore, P.D. 2011. Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol Cell* **42**(2): 172-184.

Chapman, E.J. and Carrington, J.C. 2007. Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet* **8**(11): 884-896.

Cheloufi, S., Dos Santos, C.O., Chong, M.M., and Hannon, G.J. 2010. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**(7298): 584-589.

Chen, C.H., Huang, H., Ward, C.M., Su, J.T., Schaeffer, L.V., Guo, M., and Hay, B.A. 2007a. A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. *Science* **316**(5824): 597-600.

Chen, X. 2004. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science* **303**(5666): 2022-2025.

Chen, Y., Pane, A., and Schupbach, T. 2007b. Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. *Curr Biol* **17**(7): 637-642.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**(7051): 740-744.

Chintapalli, V.R., Wang, J., and Dow, J.A. 2007. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat Genet* **39**(6): 715-720.

Chung, W.J., Okamura, K., Martin, R., and Lai, E.C. 2008. Endogenous RNA interference provides a somatic defense against Drosophila transposons. *Curr Biol* **18**(11): 795-802.

Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D., Wolfe, S.A., and Giraldez, A.J. 2010. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328**(5986): 1694-1698.

Cleary, M.A., Kilian, K., Wang, Y., Bradshaw, J., Cavet, G., Ge, W., Kulkarni, A., Paddison, P.J., Chang, K., Sheth, N., Leproust, E., Coffey, E.M., Burchard, J., McCombie, W.R., Linsley, P., and Hannon, G.J. 2004. Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis. *Nat Methods* **1**(3): 241-248.

Clegg, N.J., Frost, D.M., Larkin, M.K., Subrahmanyan, L., Bryant, Z., and Ruohola-Baker, H. 1997. maelstrom is required for an early step in the establishment of Drosophila oocyte polarity: posterior localization of grk mRNA. *Development* **124**(22): 4661-4671.

Cook, H.A., Koppetsch, B.S., Wu, J., and Theurkauf, W.E. 2004. The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell* **116**(6): 817-829.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. 1998. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev* **12**(23): 3715-3727.

Cox, D.N., Chao, A., and Lin, H. 2000. piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**(3): 503-514.

Cuperus, J.T., Fahlgren, N., and Carrington, J.C. 2011. Evolution and functional diversification of MIRNA genes. *Plant Cell* **23**(2): 431-442.

Czech, B. and Hannon, G.J. 2011. Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet* **12**(1): 19-31.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., Hannon, G.J., and Brennecke, J. 2008. An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**(7196): 798-802.

Czech, B., Zhou, R., Erlich, Y., Brennecke, J., Binari, R., Villalta, C., Gordon, A., Perrimon, N., and Hannon, G.J. 2009. Hierarchical rules for Argonaute loading in Drosophila. *Mol Cell* **36**(3): 445-456.

Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. 2005. RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Curr Biol* **15**(8): 743-749.

De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., Antony, C., Moreira, P.N., Enright, A.J., and O'Carroll, D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480**(7376): 259-263.

Deng, W. and Lin, H. 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell* **2**(6): 819-830.

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**(7014): 231-235.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Keleman, K., and Dickson, B.J. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature* **448**(7150): 151-156.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., Sorek, R., and Rechavi, G. 2012. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**(7397): 201-206.

Eamens, A.L., Smith, N.A., Curtin, S.J., Wang, M.B., and Waterhouse, P.M. 2009. The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *Rna* **15**(12): 2219-2235.

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

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 Dev* **21**(20): 2558-2570.

Fabian, M.R., Sonenberg, N., and Filipowicz, W. 2010. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **79**: 351-379.

Findley, S.D., Tamanaha, M., Clegg, N.J., and Ruohola-Baker, H. 2003. Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. *Development* **130**(5): 859-871.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**(6669): 806-811.

Flynt, A., Liu, N., Martin, R., and Lai, E.C. 2009. Dicing of viral replication intermediates during silencing of latent Drosophila viruses. *Proc Natl Acad Sci U S A* **106**(13): 5270-5275.

Flynt, A.S., Greimann, J.C., Chung, W.J., Lima, C.D., and Lai, E.C. 2010. MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. *Mol Cell* **38**(6): 900-907.

Förstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. 2007. Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**(2): 287-297.

Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. 2005. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* **3**(7): e236.

Frank, F., Hauver, J., Sonenberg, N., and Nagar, B. 2012. Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *Embo J* **31**(17): 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**(7299): 818-822.

Freeman, M. 1996. Reiterative use of the EGF receptor triggers differentiation of all cell types in the Drosophila eye. *Cell* **87**(4): 651-660.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**(7320): 67-71.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., and Zamore, P.D. 2008. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320**(5879): 1077-1081.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P.D. 2010. Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *Rna* **16**(1): 43-56.

Ghildiyal, M. and Zamore, P.D. 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**(2): 94-108.

Gillespie, D.E. and Berg, C.A. 1995. Homeless is required for RNA localization in Drosophila oogenesis and encodes a new member of the DE-H family of RNA-dependent ATPases. *Genes Dev* **9**(20): 2495-2508.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**(7099): 199-202.

Gonzalez-Reyes, A., Elliott, H., and St Johnston, D. 1997. Oocyte determination and the origin of polarity in Drosophila: the role of the spindle genes. *Development* **124**(24): 4927-4937.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**(7014): 235-240.

Griffiths-Jones, S., Hui, J.H., Marco, A., and Ronshaugen, M. 2011. MicroRNA evolution by arm switching. *EMBO Rep* **12**(2): 172-177.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**(1): 23-34.

Grivna, S.T., Beyret, E., Wang, Z., and Lin, H. 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* **20**(13): 1709-1714.

Gruber, J.J., Zatechka, D.S., Sabin, L.R., Yong, J., Lum, J.J., Kong, M., Zong, W.X., Zhang, Z., Lau, C.K., Rawlings, J., Cherry, S., Ihle, J.N., Dreyfuss, G., and Thompson, C.B. 2009. Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell* **138**(2): 328-339.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* **315**(5818): 1587-1590.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**(7308): 835-840.

Haase, A.D., Fenoglio, S., Muerdter, F., Guzzardo, P.M., Czech, B., Pappin, D.J., Chen, C., Gordon, A., and Hannon, G.J. 2010. Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. *Genes Dev* **24**(22): 2499-2504.

Haase, A.D., Jaskiewicz, L., Zhang, H., Laine, S., Sack, R., Gatignol, A., and Filipowicz, W. 2005. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep* **6**(10): 961-967.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**(5): 945-956.

Haley, B., Hendrix, D., Trang, V., and Levine, M. 2008. A simplified miRNA-based gene silencing method for Drosophila melanogaster. *Dev Biol* **321**(2): 482-490.

Hamilton, A.J. and Baulcombe, D.C. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**(5441): 950-952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**(6775): 293-296.

Han, B.W., Hung, J.H., Weng, Z., Zamore, P.D., and Ameres, S.L. 2011. The 3'-to-5' exoribonuclease Nibbler shapes the 3' ends of microRNAs bound to Drosophila Argonaute1. *Curr Biol* **21**(22): 1878-1887.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. 2004a. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**(24): 3016-3027.

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**(5): 887-901.

Han, M.H., Goud, S., Song, L., and Fedoroff, N. 2004b. The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc Natl Acad Sci U S A* **101**(4): 1093-1098.

Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. 2011. A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *Embo J* **30**(19): 3977-3993.

Harris, A.N. and Macdonald, P.M. 2001. Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. *Development* **128**(14): 2823-2832.

Hartig, J.V., Esslinger, S., Bottcher, R., Saito, K., and Forstemann, K. 2009. Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *Embo J* **28**(19): 2932-2944.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. 2006. Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* **38**(6): 721-725.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. 2008. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* **32**(2): 276-284.

Horvath, P. and Barrangou, R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**(5962): 167-170.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. 2007. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* **17**(14): 1265-1272.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., Plasterk, R.H., Hannon, G.J., Draper, B.W., and Ketting, R.F. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**(1): 69-82.

Huang, H.Y., Houwing, S., Kaaij, L.J., Meppelink, A., Redl, S., Gauci, S., Vos, H., Draper, B.W., Moens, C.B., Burgering, B.M., Ladurner, P., Krijgsveld, J., Berezikov, E., and Ketting, R.F. 2011. Tdrd1 acts as a molecular scaffold for Piwi proteins and piRNA targets in zebrafish. *Embo J* **30**(16): 3298-3308.

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**(5531): 834-838.

Hutvagner, G. and Simard, M.J. 2008. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**(1): 22-32.

Iki, T., Yoshikawa, M., Meshi, T., and Ishikawa, M. 2012. Cyclophilin 40 facilitates HSP90-mediated RISC assembly in plants. *Embo J* **31**(2): 267-278.

Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M.C., Matsumoto-Yokoyama, E., Mitsuhara, I., Meshi, T., and Ishikawa, M. 2010. In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol Cell* **39**(2): 282-291.

Iwasaki, S., Kawamata, T., and Tomari, Y. 2009. Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression. *Mol Cell* **34**(1): 58-67.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., and Tomari, Y. 2010. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* **39**(2): 292-299.

Jannot, G., Boisvert, M.E., Banville, I.H., and Simard, M.J. 2008. Two molecular features contribute to the Argonaute specificity for the microRNA and RNAi pathways in C. elegans. *Rna* **14**(5): 829-835.

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 Dev* **19**(14): 1674-1679.

Johnston, M., Geoffroy, M.C., Sobala, A., Hay, R., and Hutvagner, G. 2010. HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. *Mol Biol Cell* **21**(9): 1462-1469.

Johnston, M. and Hutvagner, G. 2011. Posttranslational modification of Argonautes and their role in small RNA-mediated gene regulation. *Silence* **2**: 5.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. 2006. MicroRNAS and their regulatory roles in plants. *Annu Rev Plant Biol* **57**: 19-53.

Karpen, G.H. and Spradling, A.C. 1992. Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics* **132**(3): 737-753.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Jr., Zhu, J.K., Staskawicz, B.J., and Jin, H. 2006. A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci U S A* **103**(47): 18002-18007.

Kawamata, T., Seitz, H., and Tomari, Y. 2009. Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol* **16**(9): 953-960.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. 2008. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**(7196): 793-797.

Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., and Katsuma, S. 2009. The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *Rna* **15**(7): 1258-1264.

Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. 2011. 3' end formation of PIWI-interacting RNAs in vitro. *Mol Cell* **43**(6): 1015-1022.

Kawaoka, S., Mitsutake, H., Kiuchi, T., Kobayashi, M., Yoshikawa, M., Suzuki, Y., Sugano, S., Shimada, T., Kobayashi, J., Tomari, Y., and Katsuma, S. 2012. A role for transcription from a piRNA cluster in de novo piRNA production. *Rna* **18**(2): 265-273.

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 Dev* **15**(20): 2654-2659.

Khurana, J.S. and Theurkauf, W. 2010. piRNAs, transposon silencing, and Drosophila germline development. *J Cell Biol* **191**(5): 905-913.

Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., Nowosielska, A., Li, C., Zamore, P.D., Weng, Z., and Theurkauf, W.E. 2011. Adaptation to P element transposon invasion in Drosophila melanogaster. *Cell* **147**(7): 1551-1563.

Khvorova, A., Reynolds, A., and Jayasena, S.D. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**(2): 209-216.

Kidwell, M.G., Kidwell, J.F., and Sved, J.A. 1977. Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics* **86**(4): 813-833.

Kirino, Y., Kim, N., de Planell-Saguer, M., Khandros, E., Chiorean, S., Klein, P.S., Rigoutsos, I., Jongens, T.A., and Mourelatos, Z. 2009. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nat Cell Biol* **11**(5): 652-658.

Kirino, Y., Vourekas, A., Sayed, N., de Lima Alves, F., Thomson, T., Lasko, P., Rappsilber, J., Jongens, T.A., and Mourelatos, Z. 2010. Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *Rna* **16**(1): 70-78.

Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A., and Theurkauf, W.E. 2007. Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell* **12**(1): 45-55.

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., Seitz, H., Zamore, P.D., Weng, Z., and Theurkauf, W.E. 2009. The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138**(6): 1137-1149.

Klenov, M.S., Sokolova, O.A., Yakushev, E.Y., Stolyarenko, A.D., Mikhaleva, E.A., Lavrov, S.A., and Gvozdev, V.A. 2011. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc Natl Acad Sci U S A* **108**(46): 18760-18765.

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**(5538): 2269-2271.

Kok, K.H., Ng, M.H., Ching, Y.P., and Jin, D.Y. 2007. Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J Biol Chem* **282**(24): 17649-17657.

Kozomara, A. and Griffiths-Jones, S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**(Database issue): D152-157.

Kulkarni, M.M., Booker, M., Silver, S.J., Friedman, A., Hong, P., Perrimon, N., and Mathey-Prevot, B. 2006. Evidence of off-target effects associated with long dsRNAs in Drosophila melanogaster cell-based assays. *Nat Methods* **3**(10): 833-838.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H., and Nakano, T. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**(7): 908-917.

Kurihara, Y., Takashi, Y., and Watanabe, Y. 2006. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *Rna* **12**(2): 206-212.

Kurihara, Y. and Watanabe, Y. 2004. Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci U S A* **101**(34): 12753-12758.

Landthaler, M., Yalcin, A., and Tuschl, T. 2004. The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol* **14**(23): 2162-2167.

Lau, N.C., Robine, N., Martin, R., Chung, W.J., Niki, Y., Berezikov, E., and Lai, E.C. 2009. Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome Res* **19**(10): 1776-1785.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. 2006. Characterization of the piRNA complex from rat testes. *Science* **313**(5785): 363-367.

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**(1): 78-87.

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**(5): 843-854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V.N. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**(6956): 415-419.

Lee, Y., Hur, I., Park, S.Y., Kim, Y.K., Suh, M.R., and Kim, V.N. 2006. The role of PACT in the RNA silencing pathway. *Embo J* **25**(3): 522-532.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. 2004a. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**(20): 4051-4060.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. 2004b. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**(1): 69-81.

Lehrbach, N.J., Armisen, J., Lightfoot, H.L., Murfitt, K.J., Bugaut, A., Balasubramanian, S., and Miska, E.A. 2009. LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans. *Nat Struct Mol Biol* **16**(10): 1016-1020.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., Kittler, E.L., Zapp, M.L., Klattenhoff, C., Schulz, N., Theurkauf, W.E., Weng, Z., and Zamore, P.D. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**(3): 509-521.

Liang, L., Diehl-Jones, W., and Lasko, P. 1994. Localization of vasa protein to the Drosophila pole plasm is independent of its RNA-binding and helicase activities. *Development* **120**(5): 1201-1211.

Lim, A.K. and Kai, T. 2007. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **104**(16): 6714-6719.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**(5689): 1437-1441.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. 2005. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol* **7**(7): 719-723.

Liu, N., Abe, M., Sabin, L.R., Hendriks, G.J., Naqvi, A.S., Yu, Z., Cherry, S., and Bonini, N.M. 2011. The exoribonuclease Nibbler controls 3' end processing of microRNAs in Drosophila. *Curr Biol* **21**(22): 1888-1893.

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. 2003. R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science* **301**(5641): 1921-1925.

Liu, X., Jiang, F., Kalidas, S., Smith, D., and Liu, Q. 2006. Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. *Rna* **12**(8): 1514-1520.

Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L.N., Grishin, N.V., and Liu, Q. 2009. C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science* **325**(5941): 750-753.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297**(5589): 2053-2056.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. 2004. Nuclear export of microRNA precursors. *Science* **303**(5654): 95-98.

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**(7033): 666-670.

Ma, L., Buchold, G.M., Greenbaum, M.P., Roy, A., Burns, K.H., Zhu, H., Han, D.Y., Harris, R.A., Coarfa, C., Gunaratne, P.H., Yan, W., and Matzuk, M.M. 2009. GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. *PLoS Genet* **5**(9): e1000635.

MacRae, I.J. and Doudna, J.A. 2007. Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Curr Opin Struct Biol* **17**(1): 138-145.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. 2009. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**(3): 522-535.

Malone, C.D. and Hannon, G.J. 2009. Small RNAs as guardians of the genome. *Cell* **136**(4): 656-668.

Markham, N.R. and Zuker, M. 2008. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* **453**: 3-31.

Marques, J.T., Kim, K., Wu, P.H., Alleyne, T.M., Jafari, N., and Carthew, R.W. 2010. Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. *Nat Struct Mol Biol* **17**(1): 24-30.

Marraffini, L.A. and Sontheimer, E.J. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**(5909): 1843-1845.

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**(4): 607-620.

McClintock, B. 1951. Chromosome organization and genic expression. *Cold Spring Harb Symp Quant Biol* **16**: 13-47.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**(2): 185-197.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G.J., and Qi, Y. 2008. Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133**(1): 116-127.

Miyoshi, K., Miyoshi, T., Hartig, J.V., Siomi, H., and Siomi, M.C. 2010a. Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in Drosophila. *Rna* **16**(3): 506-515.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. 2005. Slicer function of Drosophila Argonautes and its involvement in RISC formation. *Genes Dev* **19**(23): 2837-2848.

Miyoshi, T., Takeuchi, A., Siomi, H., and Siomi, M.C. 2010b. A direct role for Hsp90 in pre-RISC formation in Drosophila. *Nat Struct Mol Biol* **17**(8): 1024-1026.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. 2008. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell* **133**(1): 128-141.

Moussian, B., Schoof, H., Haecker, A., Jurgens, G., and Laux, T. 1998. Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis. *Embo J* **17**(6): 1799-1809.

Muerdter, F., Olovnikov, I., Molaro, A., Rozhkov, N.V., Czech, B., Gordon, A., Hannon, G.J., and Aravin, A.A. 2012. Production of artificial piRNAs in flies and mice. *Rna* **18**(1): 42-52.

Munn, K. and Steward, R. 2000. The shut-down gene of Drosophila melanogaster encodes a novel FK506-binding protein essential for the formation of germline cysts during oogenesis. *Genetics* **156**(1): 245-256.

Nakazawa, Y., Hiraguri, A., Moriyama, H., and Fukuhara, T. 2007. The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the trans-acting siRNA pathway. *Plant Mol Biol* **63**(6): 777-785.

Ni, J.Q., Zhou, R., Czech, B., Liu, L.P., Holderbaum, L., Yang-Zhou, D., Shim, H.S., Tao, R., Handler, D., Karpowicz, P., Binari, R., Booker, M., Brennecke, J., Perkins, L.A., Hannon, G.J., and Perrimon, N. 2011. A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat Methods* **8**(5): 405-407.

Niki, Y., Yamaguchi, T., and Mahowald, A.P. 2006. Establishment of stable cell lines of Drosophila germ-line stem cells. *Proc Natl Acad Sci U S A* **103**(44): 16325-16330.

Nishida, K.M., Okada, T.N., Kawamura, T., Mituyama, T., Kawamura, Y., Inagaki, S., Huang, H., Chen, D., Kodama, T., Siomi, H., and Siomi, M.C. 2009. Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. *Embo J* **28**(24): 3820-3831.

Nishida, K.M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., Siomi, H., and Siomi, M.C. 2007. Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. *Rna* **13**(11): 1911-1922.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. 2008a. The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453**(7196): 803-806.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* **130**(1): 89-100.

Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18**(14): 1655-1666.

Okamura, K., Liu, N., and Lai, E.C. 2009. Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. *Mol Cell* **36**(3): 431-444.

Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T., and Lai, E.C. 2008b. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol* **15**(4): 354-363.

Olivieri, D., Senti, K.A., Subramanian, S., Sachidanandam, R., and Brennecke, J. 2012. The Cochaperone Shutdown Defines a Group of Biogenesis Factors Essential for All piRNA Populations in Drosophila. *Mol Cell*.

Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K., and Brennecke, J. 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *Embo J* **29**(19): 3301-3317.

Paddison, P.J., Silva, J.M., Conklin, D.S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scobie, K., Chang, K., Westbrook, T., Cleary, M., Sachidanandam, R., McCombie, W.R., Elledge, S.J., and Hannon, G.J. 2004. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**(6981): 427-431.

Pak, J. and Fire, A. 2007. Distinct populations of primary and secondary effectors during RNAi in C. elegans. *Science* **315**(5809): 241-244.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. 2003. Control of leaf morphogenesis by microRNAs. *Nature* **425**(6955): 257-263.

Pane, A., Jiang, P., Zhao, D.Y., Singh, M., and Schupbach, T. 2011. The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. *Embo J* **30**(22): 4601-4615.

Pane, A., Wehr, K., and Schupbach, T. 2007. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev Cell* **12**(6): 851-862.

Park, J.K., Liu, X., Strauss, T.J., McKearin, D.M., and Liu, Q. 2007. The miRNA pathway intrinsically controls self-renewal of Drosophila germline stem cells. *Curr Biol* **17**(6): 533-538.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol* **12**(17): 1484-1495.

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**(5): 807-818.

Parker, J.S., Roe, S.M., and Barford, D. 2005. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**(7033): 663-666.

Patil, V.S. and Kai, T. 2010. Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas. *Curr Biol* **20**(8): 724-730.

Perrimon, N., Ni, J.Q., and Perkins, L. 2010. In vivo RNAi: today and tomorrow. *Cold Spring Harb Perspect Biol* **2**(8): a003640.

Petrella, L.N., Smith-Leiker, T., and Cooley, L. 2007. The Ovhts polyprotein is cleaved to produce fusome and ring canal proteins required for Drosophila oogenesis. *Development* **134**(4): 703-712.

Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., and Sontheimer, E.J. 2004. A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. *Cell* **117**(1): 83-94.

Pratt, W.B. 1998. The hsp90-based chaperone system: involvement in signal transduction from a variety of hormone and growth factor receptors. *Proc Soc Exp Biol Med* **217**(4): 420-434.

Pratt, W.B., Galigniana, M.D., Harrell, J.M., and DeFranco, D.B. 2004. Role of hsp90 and the hsp90-binding immunophilins in signalling protein movement. *Cell Signal* **16**(8): 857-872.

Preall, J.B., Czech, B., Guzzardo, P.M., Muerdter, F., and Hannon, G.J. 2012. shutdown is a component of the Drosophila piRNA biogenesis machinery. *Rna* **18**(8): 1446-1457.

Presente, A., Shaw, S., Nye, J.S., and Andres, A.J. 2002. Transgene-mediated RNA interference defines a novel role for notch in chemosensory startle behavior. *Genesis* **34**(1-2): 165-169.

Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B., and Radmark, O. 2002. Ribonuclease activity and RNA binding of recombinant human Dicer. *Embo J* **21**(21): 5864-5874.

Qi, H., Watanabe, T., Ku, H.Y., Liu, N., Zhong, M., and Lin, H. 2011. The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. *J Biol Chem* **286**(5): 3789-3797.

Rand, T.A., Petersen, S., Du, F., and Wang, X. 2005. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**(4): 621-629.

Rangan, P., Malone, C.D., Navarro, C., Newbold, S.P., Hayes, P.S., Sachidanandam, R., Hannon, G.J., and Lehmann, R. 2011. piRNA production requires heterochromatin formation in Drosophila. *Curr Biol* **21**(16): 1373-1379.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. 2002. MicroRNAs in plants. *Genes Dev* **16**(13): 1616-1626.

Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R., and Pillai, R.S. 2011. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480**(7376): 264-267.

Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A., and Pillai, R.S. 2009. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol* **16**(6): 639-646.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. 2002. Prediction of plant microRNA targets. *Cell* **110**(4): 513-520.

Ro, S., Park, C., Young, D., Sanders, K.M., and Yan, W. 2007. Tissue-dependent paired expression of miRNAs. *Nucleic Acids Res* **35**(17): 5944-5953.

Robine, N., Lau, N.C., Balla, S., Jin, Z., Okamura, K., Kuramochi-Miyagawa, S., Blower, M.D., and Lai, E.C. 2009. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* **19**(24): 2066-2076.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., and Bradley, A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res* **14**(10A): 1902-1910.

Rorth, P. 1998. Gal4 in the Drosophila female germline. *Mech Dev* **78**(1-2): 113-118.

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**(6): 1193-1207.

Ruby, J.G., Jan, C.H., and Bartel, D.P. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**(7149): 83-86.

Sabin, L.R., Zhou, R., Gruber, J.J., Lukinova, N., Bambina, S., Berman, A., Lau, C.K., Thompson, C.B., and Cherry, S. 2009. Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. *Cell* **138**(2): 340-351.

Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., and Siomi, M.C. 2009. A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. *Nature* **461**(7268): 1296-1299.

Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K.M., Siomi, H., and Siomi, M.C. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes Dev* **24**(22): 2493-2498.

Saito, K., Ishizuka, A., Siomi, H., and Siomi, M.C. 2005. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *PLoS Biol* **3**(7): e235.

Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev* **20**(16): 2214-2222.

Saito, K., Sakaguchi, Y., Suzuki, T., Siomi, H., and Siomi, M.C. 2007. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev* **21**(13): 1603-1608.

Sarot, E., Payen-Groschene, G., Bucheton, A., and Pelisson, A. 2004. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. *Genetics* **166**(3): 1313-1321.

Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., and Gvozdev, V. 2006. Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. *Genes Dev* **20**(3): 345-354.

Schüpbach, T. and Wieschaus, E. 1989. Female sterile mutations on the second chromosome of Drosophila melanogaster. I. Maternal effect mutations. *Genetics* **121**(1): 101-117.

Schüpbach, T. and Wieschaus, E. 1991. Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129**(4): 1119-1136.

Schwab, R., Ossowski, S., Riester, M., Warthmann, N., and Weigel, D. 2006. Highly specific gene silencing by artificial microRNAs in Arabidopsis. *Plant Cell* **18**(5): 1121-1133.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**(2): 199-208.

Seitz, H., Ghildiyal, M., and Zamore, P.D. 2008. Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA* strands in flies. *Curr Biol* **18**(2): 147-151.

Senti, K.A. and Brennecke, J. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet* **26**(12): 499-509.

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**(1): 65-77.

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**(5809): 244-247.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**(4): 246-258.

Slotkin, R.K. and Martienssen, R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**(4): 272-285.

Squires, J.E., Patel, H.R., Nousch, M., Sibbritt, T., Humphreys, D.T., Parker, B.J., Suter, C.M., and Preiss, T. 2012. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* **40**(11): 5023-5033.

Steiner, F.A., Hoogstrate, S.W., Okihara, K.L., Thijssen, K.L., Ketting, R.F., Plasterk, R.H., and Sijen, T. 2007. Structural features of small RNA precursors determine Argonaute loading in Caenorhabditis elegans. *Nat Struct Mol Biol* **14**(10): 927-933.

Szakmary, A., Reedy, M., Qi, H., and Lin, H. 2009. The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. *J Cell Biol* **185**(4): 613-627.

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**(7): 861-871.

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 Physiol* **49**(4): 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., and Hannon, G.J. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**(7194): 534-538.

Tanentzapf, G., Devenport, D., Godt, D., and Brown, N.H. 2007. Integrin-dependent anchoring of a stem-cell niche. *Nat Cell Biol* **9**(12): 1413-1418.

Theurkauf, W.E., Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., and Cook, H.A. 2006. rasiRNAs, DNA damage, and embryonic axis specification. *Cold Spring Harb Symp Quant Biol* **71**: 171-180.

Tolia, N.H. and Joshua-Tor, L. 2007. Slicer and the argonautes. *Nat Chem Biol* **3**(1): 36-43.

Tomari, Y., Du, T., and Zamore, P.D. 2007. Sorting of Drosophila small silencing RNAs. *Cell* **130**(2): 299-308.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. 2004. A protein sensor for siRNA asymmetry. *Science* **306**(5700): 1377-1380.

Tracey, W.D., Jr., Ning, X., Klingler, M., Kramer, S.G., and Gergen, J.P. 2000. Quantitative analysis of gene function in the Drosophila embryo. *Genetics* **154**(1): 273-284.

Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H., and Tomari, Y. 2011. Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nat Struct Mol Biol* **18**(10): 1153-1158.

Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., and Sharp, P.A. 1999. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev* **13**(24): 3191-3197.

Vagin, V.V., Klenov, M.S., Kalmykova, A.I., Stolyarenko, A.D., Kotelnikov, R.N., and Gvozdev, V.A. 2004. The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster. *RNA Biol* **1**(1): 54-58.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**(5785): 320-324.

Vagin, V.V., Wohlschlegel, J., Qu, J., Jonsson, Z., Huang, X., Chuma, S., Girard, A., Sachidanandam, R., Hannon, G.J., and Aravin, A.A. 2009. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev* **23**(15): 1749-1762.

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. 2004. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr Biol* **14**(4): 346-351.

Vert, J.P., Foveau, N., Lajaunie, C., and Vandenbrouck, Y. 2006. An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics* **7**: 520.

Wang, J., Czech, B., Crunk, A., Wallace, A., Mitreva, M., Hannon, G.J., and Davis, R.E. 2011. Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. *Genome Res* **21**(9): 1462-1477.

Wang, S.H. and Elgin, S.C. 2011. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc Natl Acad Sci U S A* **108**(52): 21164-21169.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. 2009. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**(7265): 754-761.

Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* **20**(13): 1732-1743.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M.A., Sakaki, Y., and Sasaki, H. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**(7194): 539-543.

Wehr, K., Swan, A., and Schupbach, T. 2006. Deadlock, a novel protein of Drosophila, is required for germline maintenance, fusome morphogenesis and axial patterning in oogenesis and associates with centrosomes in the early embryo. *Dev Biol* **294**(2): 406-417.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**(7385): 331-338.

Wightman, B., Ha, I., and Ruvkun, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**(5): 855-862.

Yan, W., Ma, L., Zilinski, C.A., and Matzuk, M.M. 2004. Identification and characterization of evolutionarily conserved pufferfish, zebrafish, and frog orthologs of GASZ. *Biol Reprod* **70**(6): 1619-1625.

Yan, W., Rajkovic, A., Viveiros, M.M., Burns, K.H., Eppig, J.J., and Matzuk, M.M. 2002. Identification of Gasz, an evolutionarily conserved gene expressed exclusively in germ cells and encoding a protein with four ankyrin repeats, a sterile-alpha motif, and a basic leucine zipper. *Mol Endocrinol* **16**(6): 1168-1184.

Yang, J.S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C. 2010. Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proc Natl Acad Sci U S A* **107**(34): 15163-15168.

Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R., and Nishikura, K. 2006. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* **13**(1): 13-21.

Yekta, S., Shih, I.H., and Bartel, D.P. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**(5670): 594-596.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**(24): 3011-3016.

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**(4): 747-757.

Yin, H. and Lin, H. 2007. An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature* **450**(7167): 304-308.

Yoda, M., Kawamata, T., Paroo, Z., Ye, X., Iwasaki, S., Liu, Q., and Tomari, Y. 2010. ATP-dependent human RISC assembly pathways. *Nat Struct Mol Biol* **17**(1): 17-23.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**(1): 25-33.

Zamparini, A.L., Davis, M.Y., Malone, C.D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G.J., and Lehmann, R. 2011. Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. *Development* **138**(18): 4039-4050.

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. *Embo J* **21**(21): 5875-5885.

Zhang, Z., Xu, J., Koppetsch, B.S., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W.E., and Zamore, P.D. 2011. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol Cell* **44**(4): 572-584.

Zhou, R., Czech, B., Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., and Hannon, G.J. 2009. Processing of Drosophila endo-siRNAs depends on a specific Loquacious isoform. *Rna* **15**(10): 1886-1895.

5 Indices

5.1 Index of Figures

Acknowledgements

The work presented in this thesis has been carried out in the laboratory of Prof. Dr. Greg Hannon at the *Cold Spring Harbor Laboratory*, NY, USA.

At first, I would like to thank my advisor, Greg Hannon, for giving me the possibility to carry out my thesis research in his lab. It has been a privilege working in this awesome and open-minded environment. I am deeply grateful for his brilliant ideas, his encouragement, and for his terrific support. Being a part of the Hannon lab certainly changed the way I think about science.

In equal parts, I would also like to thank my second advisor, Prof. Dr. Gerd Jürgens, who agreed to mentor me, allowed me to go aboard for my research, and represents this work at the University of Tübingen. I am very grateful for his support and advice.

I would like to thank Julius Brennecke for his fruitful supervision during the early stages of my doctoral work, and for countless scientific and non-scientific discussions during the course of my thesis projects.

Furthermore, I want to thank our collaborators as well as current and former members of the Hannon lab, who helped me with words and deeds, especially Alexei Aravin, Ilaria Falciatori, Angelique Girard, Paloma Guzzardo, Astrid Haase, Emily Hodges, Ted Karginov, Emily Lee, Colin Malone, Jon McGinn, Antoine Molaro, Maria Mosquera, Felix Mürdter, Jon Preall, Vasily Vagin, Elvin Wagenblast, Yang Yu and Rui Zhou. I also want to thank Yaniv Erlich, Assaf Gordon, Simon Knott, Ravi Sachidanandam and Oliver Tam for computational support. In particular, I am grateful to members of the small RNA team for their efforts and enthusiasm, which was of major importance to my research.

I would like to thank Jon McGinn, Alice Eberhardt, Charles Underwood, Elvin Wagenblast and especially Oliver Tam for critically reading this thesis.

I also would like to thank the Boehringer Ingelheim Fonds for financial support in form of a PhD fellowship.

Finally, I would like to thank my parents for their endless support.

Curriculum Vitae

Persönliche Informationen

Name:	Benjamin Czech
Geburtsdatum:	08. Februar 1982
Geburtsort:	Göppingen, Deutschland
Nationalität:	Deutsch

Schulischer Werdegang

08/1988 – 06/1992	Meerbach-Grundschule, Göppingen-Bartenbach
07/1992 – 06/2001	Freihof-Gymnasium, Göppingen
01/2002 – 10/2002	Zivildienst, Wilhelmhilfe e.V., Göppingen

Wissenschaftlicher Werdegang

10/2002 – 09/2003	Studium der Bioinformatik Eberhard Karls Universität Tübingen
10/2003 – 06/2007	Studium der Biologie Eberhard Karls Universität Tübingen Fächer: Entwicklungsgenetik, Pflanzenphysiologie, Biochemie, Zellbiologie
07/2007 – 05/2008	Diplomarbeit im Labor von Prof. Gregory J. Hannon, Cold Spring Harbor Laboratory, NY, USA Titel: An endogenous siRNA pathway in *Drosophila*
06/2008	Diplom in Biologie Eberhard Karls Universität Tübingen
08/2008 – 07/2012	Bearbeitung der vorliegenden Doktorarbeit unter Anleitung von Prof. Gregory J. Hannon, Cold Spring Harbor Laboratory, NY, USA und Prof. Gerd Jürgens, Entwicklungsgenetik Eberhard Karls Universität Tübingen

Stipendien

08/2007 – 01/2008	Halter eines Studentenstipendiums des Deutschen Akademischen Austausch Diensts (DAAD) zur Anfertigung der Diplomarbeit im Ausland
04/2009 – 08/2011	Halter eines PhD-Stipendiums des Boehringer Ingelheim Fonds (B.I.F.)

Accepted manuscripts:

1. Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform
    a) Main paper
    b) Supplementary information
2. Hierarchical rules for Argonaute loading in *Drosophila*
    a) Main paper
    b) Supplementary information
3. Small RNA sorting: matchmaking for Argonautes
    a) Main publication
4. A genome-scale shRNA resource for transgenic RNAi in *Drosophila*
    a) Main paper
    b) Supplementary information
5. *shutdown* is a component of the *Drosophila* piRNA biogenesis machinery
    a) Main paper
    b) Supplementary information

Submitted and prepared manuscripts:

6. A transcriptome-wide RNAi screen in *Drosophila* ovaries reveals novel factors of the germline piRNA pathway

Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform

Rui Zhou*, **Benjamin Czech***, Julius Brennecke, Ravi Sachidanandam, James A. Wohlschlegel, Norbert Perrimon and Gregory J. Hannon

* authors contributed equally to this work

RNA. 2009 Oct;15(10):1886-95. Epub 2009 Jul 27.

Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform

RUI ZHOU,[1,4] BENJAMIN CZECH,[2,4] JULIUS BRENNECKE,[2,5] RAVI SACHIDANANDAM,[2,6] JAMES A. WOHLSCHLEGEL,[3] NORBERT PERRIMON,[1] and GREGORY J. HANNON[2]

[1]Harvard Medical School, Department of Genetics, Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA
[2]Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA
[3]Department of Biological Chemistry, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California 90095, USA

ABSTRACT

***Drosophila melanogaster* expresses three classes of small RNAs, which are classified according to their mechanisms of biogenesis. MicroRNAs are ~22–23 nucleotides (nt), ubiquitously expressed small RNAs that are sequentially processed from hairpin-like precursors by Drosha/Pasha and Dcr-1/Loquacious complexes. MicroRNAs usually associate with AGO1 and regulate the expression of protein-coding genes. Piwi-interacting RNAs (piRNAs) of ~24–28 nt associate with Piwi-family proteins and can arise from single-stranded precursors. piRNAs function in transposon silencing and are mainly restricted to gonadal tissues. Endo-siRNAs are found in both germline and somatic tissues. These ~21-nt RNAs are produced by a distinct Dicer, Dcr-2, and do not depend on Drosha/Pasha complexes. They predominantly bind to AGO2 and target both mobile elements and protein-coding genes. Surprisingly, a subset of endo-siRNAs strongly depend for their production on the dsRNA-binding protein Loquacious (Loqs), thought generally to be a partner for Dcr-1 and a cofactor for miRNA biogenesis. Endo-siRNA production depends on a specific Loqs isoform, Loqs-PD, which is distinct from the one, Loqs-PB, required for the production of microRNAs. Paralleling their roles in the biogenesis of distinct small RNA classes, Loqs-PD and Loqs-PB bind to different Dicer proteins, with Dcr-1/Loqs-PB complexes and Dcr-2/Loqs-PD complexes driving microRNA and endo-siRNA biogenesis, respectively.**

Keywords: *Drosophila melanogaster*; Dicer; double-stranded RNA-binding proteins (dsRBPs); Loquacious; endo-siRNA processing; transposon silencing

INTRODUCTION

Drosophila melanogaster expresses a wide variety of small RNAs, which are classified based on their mechanism of biogenesis and the Argonaute proteins to which they bind. MicroRNAs (miRNAs) are a class of ubiquitously expressed small RNAs, typically, ~22–23 nucleotides (nt) in length. They are derived from endogenous transcripts capable of forming hairpin-like structures, which are sequentially processed by Drosha/Pasha and Dcr-1/Loqs complexes (Lee et al. 2003, 2004; Denli et al. 2004; Förstemann et al. 2005; Jiang et al. 2005; Saito et al. 2005). They predominantly associate with Argonaute-1 (AGO1) and regulate the expression of protein-coding genes (Bartel 2004; Bushati and Cohen 2007; Eulalio et al. 2008). Piwi-interacting RNAs (piRNAs), typically, ~24–28 nt in length, associate with Piwi-family proteins. The expression of piRNAs is mainly restricted to gonadal tissues, where they function in silencing of mobile elements and repeats (Aravin et al. 2007; Brennecke et al. 2007; Gunawardane et al. 2007; Klattenhoff and Theurkauf 2008). Recently, a third class of endogenous small RNAs was identified in both the germline and the soma of *Drosophila*: endogenous small interfering RNAs (endo-siRNAs) (Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008; Okamura et al. 2008). Endo-siRNAs are predominantly 21 nt in length and are derived either from long endogenous transcripts capable of forming extensive fold-back structures, or are processed

[4]These authors contributed equally to this work.

Present addresses: [5]IMBA–Institute of Molecular Biotechnology, Dr. Bohr-Gasse 3, 1030 Vienna, Austria; [6]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, 1425 Madison Avenue, New York, NY 10029, USA.

Reprint requests to: Gregory J. Hannon, Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA; e-mail: hannon@cshl.edu; fax: (516) 367-8874.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1611309.

from double-stranded regions formed by intermolecular hybridization of convergently transcribed mRNAs. Endo-siRNAs usually join Argonaute-2 (AGO2) and function in the regulation of gene expression and transposon silencing.

The biogenesis of endo-siRNAs and miRNAs depends on a number of protein complexes containing RNA processing enzymes and their dsRNA-binding protein (dsRBPs) partners. In the case of miRNA processing, the nuclear type III ribonuclease Drosha associates with the dsRBP, Pasha, and processes primary miRNA transcripts to pre-miRNAs (Lee et al. 2003; Denli et al. 2004; Gregory et al. 2004). In a second step, the cytoplasmic Dicer enzyme Dcr-1, assisted by the dsRBP Loquacious (Loqs/R3D1), further processes pre-miRNAs to mature miRNAs (Förstemann et al. 2005; Jiang et al. 2005; Saito et al. 2005). In contrast, processing of long dsRNA precursors into siRNA duplexes depends on a second *Drosophila* Dicer protein, Dcr-2 (Lee et al. 2004). The canonical Dcr-2 partner R2D2 seems not to be required for the production of siRNAs. Instead, it was found to impact the loading of siRNA duplexes into the RNA-induced silencing complex (RISC) and proper guide strand selection (Liu et al. 2003; Tomari et al. 2004). In general, it is believed that the dsRBPs contribute to the substrate specificity of their partner RNA processing enzymes.

The dsRNA binding protein Loquacious was identified in *Drosophila* as a component of a complex that also contains the type III RNase Dicer-1 (Dcr-1). Genetic experiments suggested that Loqs was required for efficient miRNA biogenesis (Förstemann et al. 2005; Jiang et al. 2005; Saito et al. 2005; Liu et al. 2007). Loss of *loqs* mainly impacted the final step of miRNA processing as indicated by the accumulation of pre-miRNAs, which are formed by Drosha/Pasha complexes. Mutations in *loqs* also reduced levels of a subset of mature miRNAs, consistent with the impacts of these lesions on *Drosophila* viability and fertility (Förstemann et al. 2005; Jiang et al. 2005; Saito et al. 2005; Liu et al. 2007; Park et al. 2007; Ye et al. 2007). Recently, it was found that loss of *loqs* strongly reduced levels of endogenous siRNAs (endo-siRNAs) derived from structured loci in both S2 cells and flies (Czech et al. 2008; Okamura et al. 2008).

In *Drosophila*, the alternative splicing of *loqs* transcripts was reported to produce three distinct isoforms: *loqs-RA*, *RB*, and *RC* (Förstemann et al. 2005; Jiang et al. 2005). These are translated into three protein isoforms, Loqs-PA, PB, and PC. *RB* is the isoform predominantly expressed in ovaries, whereas *RA* is the principal isoform found in males. The third mRNA isoform, *RC*, was detected only in *Drosophila* S2 cells (Förstemann et al. 2005). While Loqs-PB was sufficient to rescue the miRNA processing defects of $loqs^{KO}$ flies, Loqs-PA was incapable of restoring proper miRNA processing (Park et al. 2007), indicating that these Loqs isoforms had distinct functions during development.

Here we examined the roles of individual Loqs isoforms in different small RNA pathways and characterized the activity of a novel Loqs isoform, Loqs-PD. We show that coordinated depletion of all Loqs isoforms in cultured cells affects the biogenesis of both miRNAs and endo-siRNAs, whereas cells singly depleted of Loqs-PB or Loqs-PD show an impact only on the miRNA or on the endo-siRNA pathway, respectively. While the re-expression of Loqs-PD restored endo-siRNA levels in cultured cells that had been depleted of all Loqs isoforms, Loqs-PD was incapable of rescuing miRNA processing defects. Moreover, we show that Loqs-PD preferentially interacts with Dcr-2, the enzyme responsible for the processing of all endo-siRNA species. Considered together, our studies demonstrate that a single Loquacious isoform, Loqs-PD, is necessary and sufficient for the biogenesis of several types of endogenous siRNAs.

RESULTS AND DISCUSSION

Using quantitative proteomics of Loqs immunoprecipitates from flies and *Drosophila* S2 cells, we identified physical interactions of Loqs with both Dcr-1 and Dcr-2 (Czech et al. 2008). In order to identify the Loqs isoforms involved in these interactions, we analyzed our proteomics data from immunoprecipitates, prepared using an antibody specific to the N terminus of endogenous Loqs proteins, for isoform-specific peptides. We found no significant peptide evidence for Loqs-PC in cultured cells and flies, whereas both isoforms PA and PB were present. In addition, we detected peptides corresponding to an as-yet-uncharacterized form of Loquacious. These could be assigned to a novel isoform identified by Förstemann and colleagues and termed Loqs-PD (Fig. 1A, 1B; JV Hartig, S Esslinger, R Böttcher, K Saito, and Förstemann K, unpubl.).

To characterize the roles of individual Loqs isoforms in small RNA biogenesis, we examined isoform-specific effects on miRNA and endo-siRNA processing. We depleted various combinations of Loqs proteins using isoform-specific dsRNAs in *Drosophila* S2 cells (Fig. 1A). Treatment of S2 cells with dsRNAs targeting *loqs-RA*, *RB*, or *RC* either singly or in combination led to a reduction in steady-state levels of the corresponding transcripts, as measured by RT-PCR (data not shown). In addition, Western blotting confirmed the depletion of specific Loqs isoforms by the corresponding dsRNAs (Fig. 1C). Notably, while Loqs-PC and PD could not be efficiently resolved by PAGE, cells treated with dsRNAs specifically targeting the PD isoform were effectively depleted of the Loqs protein species migrating at the position of PC/PD (Fig. 1C, lane 11). This indicated that the Loqs-PC isoform was not detectably expressed and confirmed our findings from the quantitative proteomic analysis of Loqs immunoprecipitates.

We examined the impact of isoform-specific knockdowns on the biogenesis of a prevalent endo-siRNA (Fig. 1D, *esi-2.1*) and a miRNA (Fig. 1D, *miR-bantam*). Depletion of all isoforms upon treatment with dsRNAs targeting

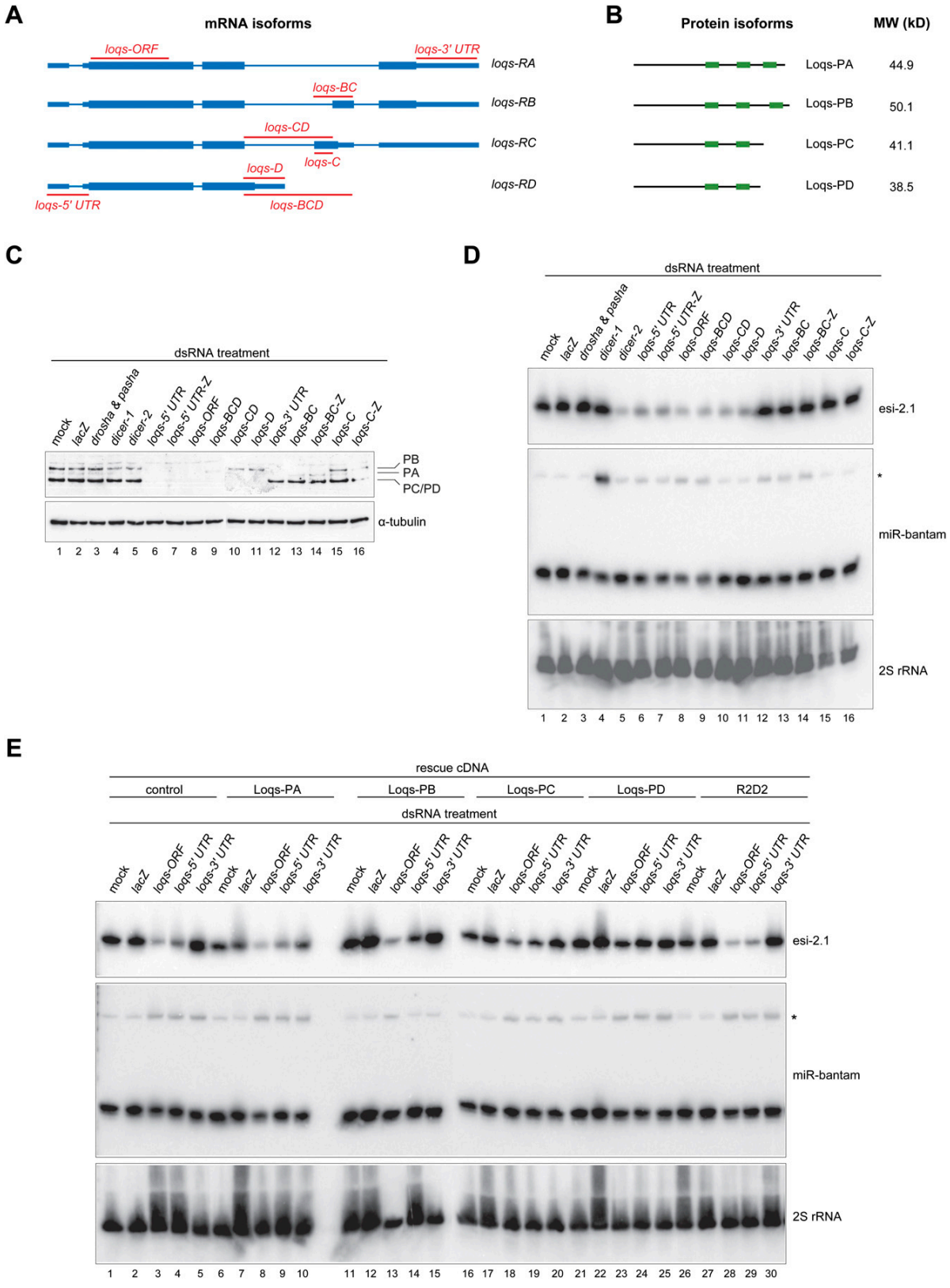


FIGURE 1. (Legend on next page)

common 5′-UTR or ORF sequences resulted in the accumulation of the *miR-bantam* precursor and in a strong reduction of *esi-2.1* levels (Fig. 1D, lanes 6–8), whereas depletion of isoforms PA, PB, and PC by dsRNAs targeting isoform-specific 3′-UTRs only affected miRNA processing (Fig. 1D, lane 12). We observed miRNA processing defects in all cells treated with dsRNAs that cotargeted Loqs-PB (Fig. 1D, lanes 6–9,12–14), whereas the biogenesis of endo-siRNA was affected only if cells were treated with dsRNAs that target Loqs-PD, either singly or together with other isoforms (Fig. 1D, lanes 6–11). Notably, depletion of Loqs-PD alone caused a strong reduction in *esi-2.1* levels, while miRNA processing was unaffected (Fig. 1D, lane 11).

To validate these findings, we depleted all Loqs isoforms in S2 cells using dsRNA targeting shared 5′-UTR sequences and tested whether the subsequent introduction of RNAi-resistant ORFs directing expression of individual Loqs isoforms was capable of rescuing defects in endo-siRNA or miRNA processing. In control cells, depletion of all Loqs isoforms led to a significant reduction in levels of *esi-2.1* (Fig. 1E, lanes 3,4), whereas dsRNAs targeting *loqs* 5′- or 3′-UTRs or targeting ORFs caused a moderate accumulation of the *miR-bantam* precursor (Fig. 1E, lanes 3–5). The expression of Loqs-PA, Loqs-PC, or R2D2 failed to rescue any observed biogenesis defect (Fig. 1E, lanes 6–10,16–20,26–30). In contrast, the re-expression of Loqs-PB effectively rescued miRNA-processing defects (Fig. 1E, cf. pre-*miR-bantam* levels in lanes 14,15 and lane 13). However, Loqs-PB failed to restore normal endo-siRNA levels (Fig. 1E, lane 14). The re-expression of Loqs-PD restored levels of *esi-2.1* in Loqs-depleted cells (Fig. 1E, lane 24) but was incapable of rescuing miRNA-processing defects (Fig. 1E, cf. lanes 24,25 and lane 23). Our observation that the expression of R2D2 was incapable of rescuing any small RNA-processing defect caused by depletion of Loqs strongly suggests that R2D2 and all Loqs isoforms cannot function in a redundant manner. Considered together, these data indicate that only the Loqs-PB isoform is required for the biogenesis of miRNAs and suggest that only Loqs-PD is essential for endo-siRNA production.

To support results emerging from cell culture, we also examined the roles of Loqs isoforms in vivo. Flies expressing the Loqs-PB isoform under the control of its endogenous regulatory elements could rescue both the pre-miRNA-processing defects and the pronounced phenotypes of *loqs*-null animals (Park et al. 2007). Strikingly, homozygous mutant $loqs^{KO}$ flies carrying a Loqs-PB transgene did not regain normal levels of the endo-siRNA, *esi-2.1*, but did express normal amounts of *miR-8* (Supplemental Fig. S1A). Similarly, the introduction of Loqs-PB also restored normal *miR-8* levels in $loqs^{f00791}$ homozygous mutant flies, whereas endo-siRNA biogenesis defects were not affected (Supplemental Fig. S1B). The expression of Loqs-PA in $loqs^{f00791}$ homozygous mutant flies was incapable of restoring either miRNA or endo-siRNA processing (Supplemental Fig. S1B). These results are consistent with the observation from cell-based studies showing that neither Loqs-PA nor PB is required for the endo-siRNA pathway. We conclude that Loqs-PB is required for miRNA biogenesis in multiple cell types, whereas Loqs-PD supports endo-siRNA biogenesis. Loqs-PC seems neither to be expressed at significant levels nor to impact small RNA biogenesis. The function of Loqs-PA, which is expressed both in cultured cells and in animals, remains elusive.

Given the unexpected requirement for Loqs in the biogenesis of endogenous siRNAs (endo-siRNAs) (Czech et al. 2008; Okamura et al. 2008) and the physical interaction between Loqs and both Dcr-1 and Dcr-2 (Czech et al. 2008), we sought to investigate whether individual Loqs isoforms might show specificity for either Dicer protein. Immunoprecipitation of Flag-tagged Dcr-2 followed by immunoblotting with an antibody specific to the N-terminus of endogenous Loqs, and thus recognizing all known Loqs isoforms, revealed a strong signal corresponding to the molecular weight of Loqs-PC/PD (Fig. 2A, lane 2). Since the levels of endogenous Loqs-PC in S2 cells are negligible, the predominant Dcr-2-interacting endogenous Loqs isoform appears to be Loqs-PD. Next, we examined the interactions between various Loqs isoforms and Dcr-2 by expressing Flag-tagged Dcr-2 together with T7-tagged Loqs isoforms in S2 cells. Cell extracts were subjected to anti-Flag immunoprecipitation, and coimmunoprecipitated proteins were detected using an anti-T7 antibody. All four tagged Loqs isoforms and the positive control T7-R2D2 were able to interact with Dcr-2 (Fig. 2B, lanes 7–11),

FIGURE 1. A specific Loqs isoform, Loqs-PD, is required for the biogenesis of endo-siRNAs derived from structured loci. (*A*) Shown is a scheme of the annotated genomic structure of the four isoforms of Loquacious. The regions targeted by isoform-specific dsRNAs are indicated. (Thin lines) Introns; (thin boxes) UTRs; (thick boxes) ORFs. (*B*) A scheme showing domain structures and molecular weights of the Loqs isoforms. (Small boxes) dsRNA-binding domains. (*C*) Western blots showing steady-state protein levels of Loqs isoforms upon treatment of S2 cells with the indicated dsRNAs. α-Tubulin served as a loading control. To increase the length of dsRNAs to facilitate dsRNA uptake by S2 cells, some dsRNAs were fused with a *lacZ* carrier sequence (indicated by "-Z"). (*D*) Northern blots probing levels of an endo-siRNA derived from a structured locus, *esi-2.1*, and the microRNA *miR-bantam* (pre-miRNA indicated by the asterisk) in S2 cells treated with dsRNA against the indicated genes (as in *C*). As a loading control, the membrane was stripped and re-probed for 2S rRNA. (*E*) Expression of Loqs-PD is sufficient to rescue endo-siRNA processing defects, while only Loqs-PB restores proper miRNA biogenesis. Control cells expressing the TAP epitope alone or those expressing various TAP-tagged proteins were treated twice with various dsRNAs (as indicated on *top* of the panel) for a total duration of 8 d. The expression of the transgenes was controlled by the basal activity of the *metallothionein* promoter. Total RNAs were prepared and subjected to Northern blotting using probes against *esi-2.1*, *miR-bantam*, and 2S rRNA.

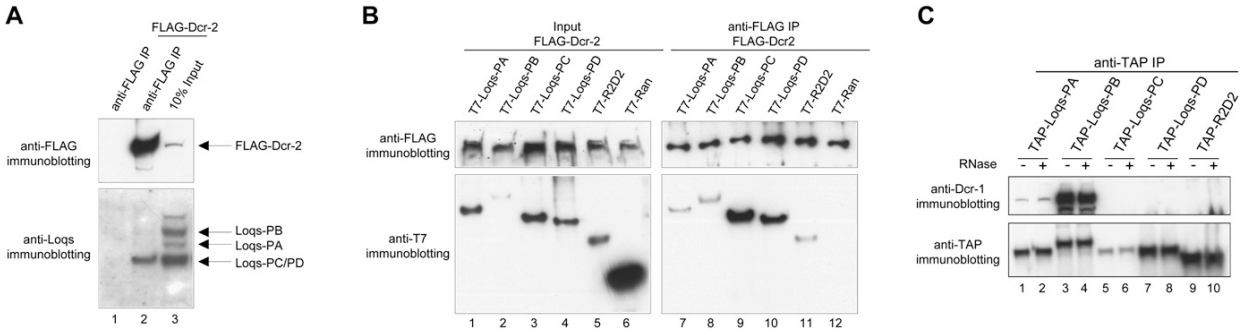


FIGURE 2. Interaction of Loqs isoforms with Dicer proteins. (*A*) Flag-tagged Dcr-2 coimmunoprecipitates endogenous Loqs-PD. Flag-Dcr-2 immunoprecipitates were subjected to immunoblotting using antibodies against the N terminus of endogenous Loqs and against the Flag epitope. (*B*) Flag-Dcr-2 was expressed in S2 cells together with T7-tagged R2D2 or various Loqs isoforms. T7-Ran served as a control. Flag-Dcr-2 was immunoprecipitated using an anti-Flag antibody, and the interaction with coexpressed proteins was examined by Western blotting using an anti-T7 antibody. (*C*) Coimmunoprecipitation of endogenous Dcr-1 with TAP-tagged Loqs isoforms. TAP-tagged proteins were expressed in S2 cells, immunoprecipitated using antibodies against the TAP epitope, and subjected to immunoblotting using antibodies against endogenous Dcr-1 and the TAP tag, respectively. RNase treatment (as indicated on *top* of the panel) revealed that the strong interaction between Dcr-1 and Loqs-PB as well as the weak interaction between Dcr-1 and Loqs-PA are RNA-independent.

while no interaction was detected for the negative control T7-Ran (Fig. 2B, lane 12). Therefore, Dcr-2 seems capable of interacting with all isoforms of Loqs if they are overexpressed. Interestingly, we also found that Loqs isoforms are capable of forming homo- and heterodimers (or even oligomers) with each other, as Flag-tagged Loqs isoforms PA, PB, and PC are capable of pulling down T7-tagged Loqs-PA in coimmunoprecipitation assays (Supplemental Fig. S3; data not shown). This could account for the apparent lack of specificity in the interaction between Dcr-2 with individual Loqs isoforms upon overexpression. We also probed the interaction between Loqs isoforms and Dcr-1. We detected robust interaction between endogenous Dcr-1 and TAP-Loqs-PB and a weak interaction between Dcr-1 and TAP-Loqs-PA (Fig. 2C, lanes 1–4). In contrast, no endogenous Dcr-1 was detected in the TAP-Loqs-PC, TAP-Loqs-PD, or TAP-R2D2 complexes.

To test whether the observed binding resulted from protein–protein interaction or from Loqs and Dicer proteins binding independently to the same RNA substrates, we treated immunoprecipitates with RNase. This had no significant effect on the observed interactions (Fig. 2C; data not shown). Overexpression of neither a pri-miRNA (pri-*miR-bantam*), nor a precursor of the endo-siRNA cluster, *esi-2* altered patterns of differential affinity (data not shown), indicating that our experiments detected RNA-independent protein–protein interactions. Therefore, we conclude that two Loqs isoforms, Loqs-PB and to a lesser extent Loqs-PA, are capable of interacting with Dcr-1, a result consistent with findings reported by Förstemann et al. (2005). In contrast, Dcr-2 predominantly interacts with endogenous Loqs-PD, while other isoforms interact with Dcr-2 to a significant extent only when overexpressed. Considering the restriction of *loqs-PC* transcripts to S2 cells (Förstemann et al. 2005) and the negligible protein levels of Loqs-PC, we conclude that endogenous Loqs-PD is the predominant isoform that interacts with Dcr-2.

In order to investigate the effects of Loqs depletion on small RNA profiles in more detail, we prepared small RNA libraries from S2 cells treated with *loqs-ORF* dsRNAs and from control knockdown cells, including *lacZ*, *dcr-1*, *dcr-2*, *r2d2*, and untreated cells (''mock''). These were deep-sequenced using the Illumina platform. Normalized cloning counts were plotted by read length to create size profiles (Fig. 3A). The indicated small RNA categories were isolated from the total library bioinformatically (Czech et al. 2008). Mature microRNAs populated a broad peak centered around 22–23 nt, and these were decreased most prominently in the *dcr-1* knockdown. MicroRNAs appeared to be slightly increased in *loqs-ORF* and *dcr-2* knockdowns as compared to *lacZ* (Fig. 3A), probably because of an artifact of library normalization due to the strong impact on endo-siRNA levels (Figs. 1D, 3B). Endo-siRNAs mapping to overlapping transcripts (exonic antisense) were strongly reduced in *dcr-2* and *loqs-ORF* knockdowns (Fig. 3A). A moderate reduction was also observed for endo-siRNAs derived from the *klarsicht* locus upon Loqs depletion, while *dcr-2* knockdowns had more prominent impacts. Finally, knockdowns of *dcr-2* or *loqs* caused a substantial (~50%) reduction in endo-siRNAs corresponding to repeat and transposon sequences, while the levels of these small RNAs remained unchanged in all other knockdowns tested (Fig. 3A).

We validated results from small RNA sequencing by Northern blotting. We saw decreased levels of three independent endo-siRNAs derived from structured loci (*esi-1.2*, *esi-2.1*, *esi-4.1*) in *dcr-2* and *loqs-ORF* knockdowns, while knockdown of *loqs-BC* had no similar impact (Fig. 3B). Consistent with these small RNA sequencing and Northern blotting results, depletion of all Loqs isoforms,

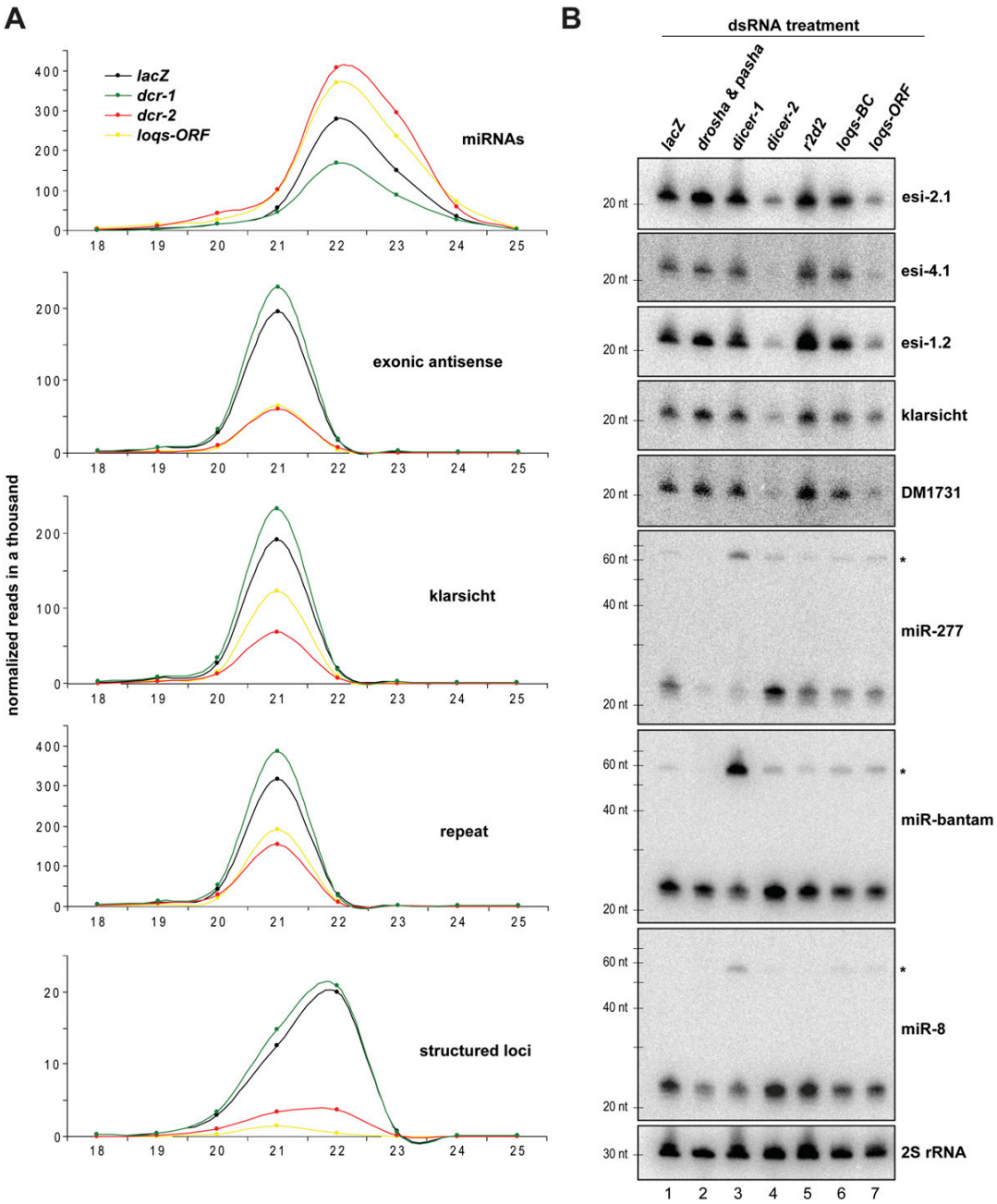


FIGURE 3. Loqs is involved in the biogenesis of several classes of endo-siRNAs. (*A*) Length profiles for small RNAs isolated from S2 cells treated with the indicated dsRNAs are shown. Known miRNAs and distinct classes of endo-siRNAs (as indicated) were split computationally, and normalized cloning counts were plotted over their length. (*B*) Northern blots show levels of miRNAs (pre-miRNAs indicated by asterisks), endo-siRNAs derived from structured loci, the LTR transposon *DM1731*, and the S2 cell-specific *klarsicht* locus in S2 cells treated with the indicated dsRNAs. 2S rRNA served as a control for equal loading.

but not depletion of Loqs-BC or Loqs-C, caused a de-repression of a *Renilla* luciferase sensor for *esi-2.1* (Supplemental Fig. S2). We also noted a reduction of *klarsicht* siRNAs and sequences derived from the transposon *DM1731* upon *loqs-ORF* knockdown, but not upon Loqs-BC depletion (Fig. 3B). Probing the same membrane for three miRNAs revealed a slight reduction in mature miRNA levels and a slight but detectable accumulation of pre-miRNAs in *dcr-1*, *loqs-ORF*, and *loqs-BC* knockdown cells (Fig. 3B). Considered together, these results support a

dependence of various categories of endo-siRNAs on the Loqs-PD isoform.

Given the effect of *loqs-ORF* knockdowns on repeat endo-siRNAs, we probed the potential impact of *loqs* mutations on transposon silencing. All endo-siRNAs that correspond to repeats and that were 21 nt in length were extracted bioinformatically, and heat maps showing their relative abundance were created for the indicated libraries (Fig. 4A). Knockdown of *dcr-2* caused a reduction of endo-siRNA sequences for the majority of transposable elements as compared to the number of reads in the *lacZ* library, while untreated cells (''mock''), *dcr-1* and *r2d2* knockdowns, showed only minor, if any, effects. Depletion of all Loqs isoforms reduced levels of repeat endo-siRNAs, which at least in part correlated with *dcr-2* depletion. Interestingly, we noted that knockdown of *r2d2* showed a weak impact on those repetitive elements that did not depend on Loqs depletion. Plotting the abundance of repeat siRNAs in Loqs or Dcr-2-depleted cells against those in the *lacZ* control library showed similar patterns, with some siRNAs appearing unaffected and others showing drastic reduction in *loqs-ORF* or *dcr-2* knockdowns (Fig. 4B,C). Plotting the normalized read number of endo-siRNAs matching to the transposable element *DM1731*, which shows the highest levels of transposon-derived endo-siRNAs in S2 cells, for the *lacZ*, *dcr-2*, and *loqs-ORF* libraries also shows clear dependence on *dcr-2* and *loqs* (Fig. 4D). Interestingly, depletion of Dcr-2 and Loqs does not show identical impacts on the patterns of individual sequences over this representative transposon (Fig. 4D). This could indicate multiple overlapping biogenesis pathways or differential affinities of one machinery for different processing sites.

Previous studies indicated that Dcr-1 preferentially binds Loqs-PB (Förstemann et al. 2005; Ye et al. 2007), and that Dcr-2 is capable of interacting with Loqs proteins, although isoforms were not examined in that study (Czech et al. 2008). Here we have characterized the role of a novel Loqs isoform, Loqs-PD, in the biogenesis of various categories of endo-siRNAs. We present evidence that Loqs-PD, in addition to the canonical partner R2D2 (Liu et al. 2003), is the predominant interacting partner for Dcr-2 in S2 cells. Our genetic studies clearly demonstrate the differential requirement for Loqs-PD and Loqs-PB in the biogenesis of endo-siRNAs and miRNAs, respectively. Thus, it is likely that besides the intrinsic substrate specificity of the individual Dicer proteins, the identity of Dicer interacting cofactors is also crucial for the biogenesis of different classes of small RNAs. Although small dsRNA-binding proteins are clearly important for small RNA production and act as cofactors for Dicer and Drosha enzymes, their precise biochemical roles remain unclear. Our studies add complexity to the landscape of the roles of dsRBPs in small RNA biogenesis and raise fundamental questions about why so many distinct forms of these proteins are required for these pathways to operate.

MATERIALS AND METHODS

DNA constructs

An ~500-base-pair DNA fragment encompassing the *miR-bantam* sequence was amplified by PCR and cloned into pRmHa-3 using BamHI/EcoRI sites to generate pRmHa-3-Bantam. A pair of oligos containing three imperfect binding sites for *miR-bantam* were annealed and cloned into pRmHa-3-*Renilla* (Zhou et al. 2008) using SalI to generate the sensor constructs for *miR-bantam* (pRmHa-3-*Renilla*-Bantam sensor). To generate epitope-tagged expression constructs, DNA fragments encoding epitope-tagged proteins were amplified by PCR and cloned into pRmHa-3 using the following restriction enzymes (SacI/KpnI for Flag-Dcr-2; EcoRI/BamHI for all T7-tagged Loqs isoforms; KpnI/BamHI for T7-R2D2; EcoRI/BamHI for T7-Ran). For TAP-tagged vectors, a pair of oligos was annealed and cloned into pMK33-NTAP (Veraksa et al. 2005) using XhoI and BamHI sites to generate pMK33-RZ-NTAP. DNA fragments encoding the corresponding proteins were either amplified by PCR and cloned into pMK33-RZ-NTAP using the following restriction enzymes (SacI/KpnI for Loqs-PA, PC, and PD), or subcloned from the Flag-tagged pRmHa-3 constructs using the following restriction enzymes (KpnI/BamHI for R2D2; EcoRI → blunt/BamHI-treated cDNA fragment cloned into SacI → blunt/BamHI-digested vector for Loqs-PB and Ran). DNA oligonucleotide sequences are listed in Supplemental Table S1.

Immunoprecipitation and immunoblotting

Cells were transfected with expression constructs for epitope-tagged proteins, induced with 500 μM $CuSO_4$ 2 d after transfection, and harvested another 24 h later. Immunoprecipitation and RNase treatment were performed as previously described (Czech et al. 2008; Zhou et al. 2008). The immunoprecipitated samples were resolved by SDS-PAGE, transferred to nitrocellulose membranes, and probed with antibodies against the T7 epitope (Novagen), the Flag-epitope (Sigma), the TAP-tag (Open Biosystems), or rabbit antibodies against Dcr-1 or Loqs. α-Tubulin was detected using an anti-α-tubulin antibody raised in mouse (Sigma).

Cell culture and dsRNA treatment

S2-NP cells were maintained in Schneider's medium (Invitrogen) supplemented with 10% FBS and 1% pen-strep (Invitrogen). For dsRNA treatment, ~3×10^6 S2-NP cells were soaked in 1.5 mL of serum-free Schneider's medium containing 10 μg of dsRNAs in 6-well plates, and 3 mL of serum-containing medium was added 45 min later. After 4 d of initial dsRNA treatment, cells were treated with dsRNAs for a second round and harvested another 4 d later. Sequences of the primers for generation the dsRNAs are listed in Supplemental Table S1.

RT-PCR

To measure levels of various *loqs* transcripts upon dsRNA treatment, total RNA from S2 cells was extracted using Trizol (Invitrogen). RNA was treated with RQ1 DNase (Promega) according to the manufacturer's instructions. cDNA was synthesized by reverse transcription using Quantiscript reverse transcriptase (QIAGEN) and a mixture of oligo-dT and random

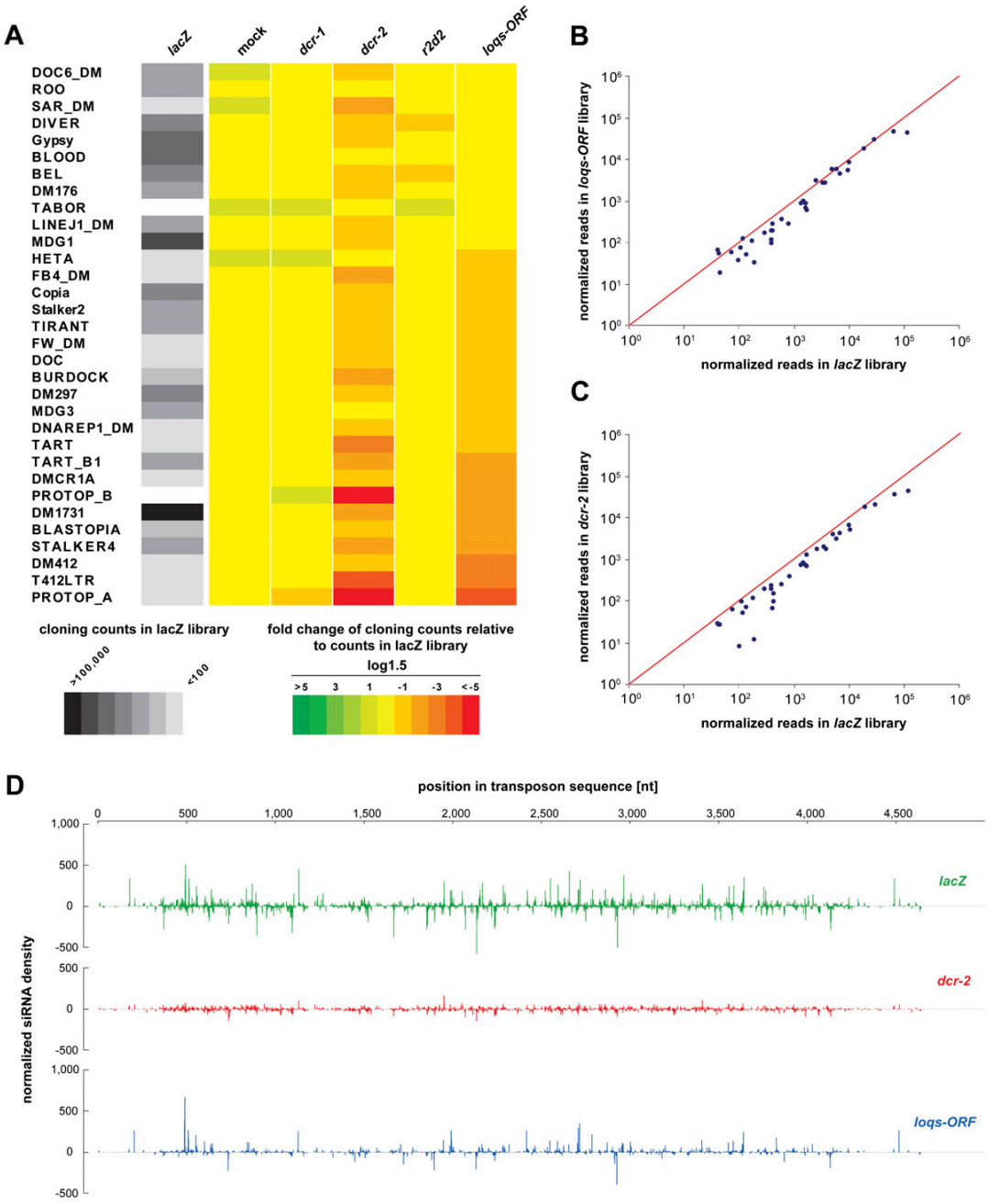


FIGURE 4. Depletion of Loqs in S2 cells results in reduced levels of endo-siRNAs derived from repeats and transposable elements. (*A*) Heat maps in grayscale show cloning frequencies of 21-nt-size siRNAs matching individual transposons in S2 cells treated with *lacZ* dsRNA. Fold changes of cloning counts relative to those in the *lacZ* library are shown in color (red-green scale) in $\log_{1.5}$ scale. Note that only transposons passing a cloning count threshold of 5 in a single library and 500 in all libraries together are shown. (*B*,*C*) All siRNAs matching to the transposons indicated in *A* with more than 50 reads in the *lacZ* library are plotted against the *loqs-ORF* and *dcr-2* libraries, respectively. (*D*) All 21-nt endo-siRNAs in the *lacZ*, *dcr-2*, and *loqs-ORF* libraries that match to the transposable element *DM1731* are plotted over the transposon sequence with their normalized cloning counts on the *y*-axis.

hexamer primers. cDNA was amplified using primers that amplify all three isoforms (products vary by size). The PCR products were separated by agarose gel electrophoresis. Analysis was carried out with two biological replicates per sample. The sequences of the DNA oligonucleotides used in this study are listed in Supplemental Table S1.

Northern blotting

Total RNA was isolated using Trizol (Invitrogen). Twenty micrograms of total RNA from S2 cells or 40 μg of total RNA from flies was separated on 15% denaturing poly acrylamide gels and transferred onto Hybond-N^+ membranes (Amersham Biosciences) in 0.5× TBE. RNA was UV cross-linked to the membrane and pre-hybridized in ULTRAhyb-Oligo buffer (Ambion) for 1 h. DNA probes (sequences are listed in Supplemental Table S1) complementary to the indicated endo-siRNAs, miRNAs, and 2S rRNA were 5′ radio-labeled and added to the hybridization buffer (hybridization overnight at 30°C). Membranes were washed four to six times in 1× SSC with 0.1% SDS at 30°C and exposed to PhosphorImager screens for 12–72 h. Probes were stripped by boiling the membrane twice in 0.2× SSC containing 0.1% SDS in a microwave.

Reporter assays

Transfections were performed in 384-well plate format. For each well, a total of ~100 ng of plasmid DNA was transfected. To measure *esi-2.1* sensor activity, 100 ng of pRmHa-3-*Renilla*-esi-2.1 sensor (Czech et al. 2008) and 2 ng of pRmHa-3-Firefly-long (Zhou et al. 2008) (as a control for transfection efficiency) were transfected. To measure the activity of endogenous *miR-bantam*, 2 ng of pRmHa-3-Firefly-long, 50 ng of pRmHa-3-*Renilla*-Bantam sensor, and 50 ng of pRmHa-3 (serving as carrier DNA) were transfected. The activity of overexpressed *miR-bantam* was measured by transfecting 2 ng of pRmHa-3-Firefly-long, 80 ng of pRmHa-3-*Renilla*-Bantam sensor, and 20 ng of pRmHa-3-Bantam. For each well, DNA was mixed with 0.8 μL of Enhancer in 15 μL of EC (QIAGEN) and incubated for 5 min at room temperature. Then 0.35 μL of Effectene reagent was added, and the mixture was immediately dispensed into each well containing ~80 ng of dsRNA. After incubation for 10 min at room temperature, 40 μL of S2-NP cells (10^6 cells/mL) were dispensed into the well. Cells were induced with 200 μM $CuSO_4$ 144 h post-transfection, and luciferase assays were performed 24 h later using DualGlo reagents (Promega). For each well, the reporter activity, referred to as relative luciferase units (RLU), was calculated as the ratio of *Renilla* luciferase to firefly luciferase. To calculate the effect of dsRNA treatment on the activity of specific sensors, the data points were first normalized against corresponding data points where pRmHa-3-*Renilla* was transfected (serving as no-site control), then normalized against samples transfected with dsRNA against LacZ.

Small RNA libraries

Small RNAs from total RNA were cloned as described (Brennecke et al. 2007) (detailed protocol available upon request). The following small RNA libraries from total RNA were prepared for this study:

19–24 nt from untreated S2 cells (“mock”);
19–24 nt from S2 cells treated with dsRNA against Dcr-1;
19–24 nt from S2 cells treated with dsRNA against Dcr-2;
19–24 nt from S2 cells treated with dsRNA against Loqs-ORF;
19–24 nt from S2 cells treated with dsRNA against R2D2; and
19–24 nt from S2 cells treated with dsRNA against LacZ.

Libraries were sequenced using the Illumina sequencing platform. Small RNA sequences were deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE17171.

Bioinformatics analysis of small RNA libraries

Small RNA libraries were analyzed as described (Czech et al. 2008). Small RNA sequences were matched to the *Drosophila* release 5 genome and genomes of *Drosophila* C virus, Flock house virus, and Cricket paralysis virus. Only reads perfectly matching the fly genome or matching to viral genomes with up to three mismatches were used for further analysis. For annotations we used FlyBase for protein-coding genes, UCSC for non-coding RNAs and transposons/repeats, and the most recent miRNA catalog (Ruby et al. 2007; Stark et al. 2007).

Fly stocks

$loqs^{KO}$ flies and flies expressing Loqs-PA and Loqs-PB isoforms were a kind gift of Qinghua Liu (University of Texas Southwestern Medical Center) (Park et al. 2007). The hypomorphic $loqs^{f00791}$ flies were obtained from Bloomington (stock #18371). $AGO2^{414}$ flies were a kind gift of Haruhiko Siomi (Keiko University School of Medicine) (Okamura et al. 2004), $dcr\text{-}2^{L811Fsx}$ flies were a kind gift of Richard Carthew (Northwestern University) (Lee et al. 2004), and $r2d2^{1}$ flies were a kind gift of Dean Smith (University of Texas Southwestern Medical Center) (Liu et al. 2003). Flies were double-balanced [double-balancer stock (*w*; *Sp*/*CyO*; *Dr*/*TM6C*,*Tb*), a gift from Phil Zamore (University of Massachusetts Medical Center)], then homozygous and heterozygotes flies were collected. Stock #2057 from Bloomington (Celera sequencing strain) was used as wild type.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

We thank J.V. Hartig and K. Förstemann and colleagues for communicating data prior to publication. We also thank Q. Liu, R. Carthew, H. Siomi, D. Smith, and P.D. Zamore for reagents. We are grateful to members of the Hannon and Perrimon laboratories for helpful discussion, and to M. Rooks and D. McCombie for help with deep sequencing. R.Z. is a Special Fellow of the Leukemia and Lymphoma Society. B.C. is supported by a PhD fellowship from the Boehringer Ingelheim Fonds. This work was supported in part by grants from the NIH to N.P. and G.J.H., and a gift from K.W. Davis (to G.J.H.). N.P. and G.J.H are investigators of the Howard Hughes Medical Institute.

Received February 19, 2009; accepted June 30, 2009.

REFERENCES

Aravin AA, Hannon GJ, Brennecke J. 2007. The Piwi–piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318:** 761–764.

Bartel DP. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116:** 281–297.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master

regulators of transposon activity in *Drosophila*. *Cell* **128:** 1089–1103.

Bushati N, Cohen SM. 2007. microRNA functions. *Annu Rev Cell Dev Biol* **23:** 175–205.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453:** 798–802.

Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the microprocessor complex. *Nature* **432:** 231–235.

Eulalio A, Huntzinger E, Izaurralde E. 2008. Getting to the root of miRNA-mediated gene silencing. *Cell* **132:** 9–14.

Förstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, Klattenhoff C, Theurkauf WE, Zamore PD. 2005. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* **3:** e236. doi: 10.1371/journal.pbio.0030236.

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320:** 1077–1081.

Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R. 2004. The microprocessor complex mediates the genesis of microRNAs. *Nature* **432:** 235–240.

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315:** 1587–1590.

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

Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. 2008. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453:** 793–797.

Klattenhoff C, Theurkauf W. 2008. Biogenesis and germline functions of piRNAs. *Development* **135:** 3–9.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425:** 415–419.

Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW. 2004. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117:** 69–81.

Liu Q, Rand TA, Kalidas S, Du F, Kim HE, Smith DP, Wang X. 2003. R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301:** 1921–1925.

Liu X, Park JK, Jiang F, Liu Y, McKearin D, Liu Q. 2007. Dicer-1, but not Loquacious, is critical for assembly of miRNA-induced silencing complexes. *RNA* **13:** 2324–2329.

Okamura K, Ishizuka A, Siomi H, Siomi MC. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes & Dev* **18:** 1655–1666.

Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC. 2008. The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453:** 803–806.

Park JK, Liu X, Strauss TJ, McKearin DM, Liu Q. 2007. The miRNA pathway intrinsically controls self-renewal of *Drosophila* germline stem cells. *Curr Biol* **17:** 533–538.

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* **17:** 1850–1864.

Saito K, Ishizuka A, Siomi H, Siomi MC. 2005. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol* **3:** e235. doi: 10.1371/journal.pbio.0030235.

Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, Hannon GJ, Kellis M. 2007. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res* **17:** 1865–1879.

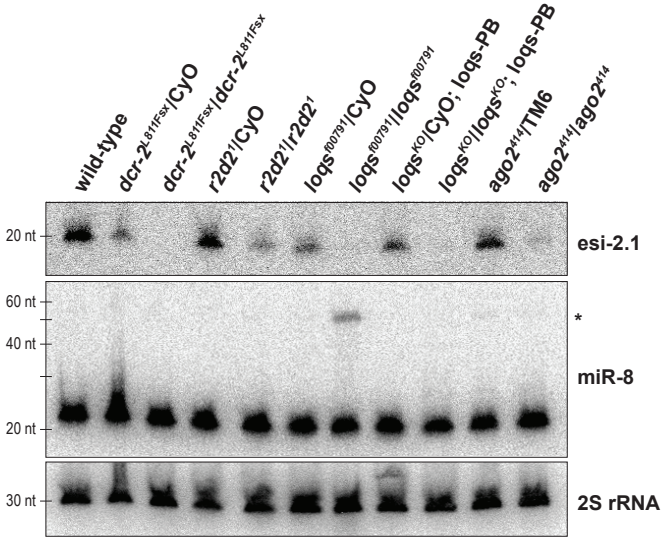
Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD. 2004. A protein sensor for siRNA asymmetry. *Science* **306:** 1377–1380.

Veraksa A, Bauer A, Artavanis-Tsakonas S. 2005. Analyzing protein complexes in *Drosophila* with tandem affinity purification-mass spectrometry. *Dev Dyn* **232:** 827–834.

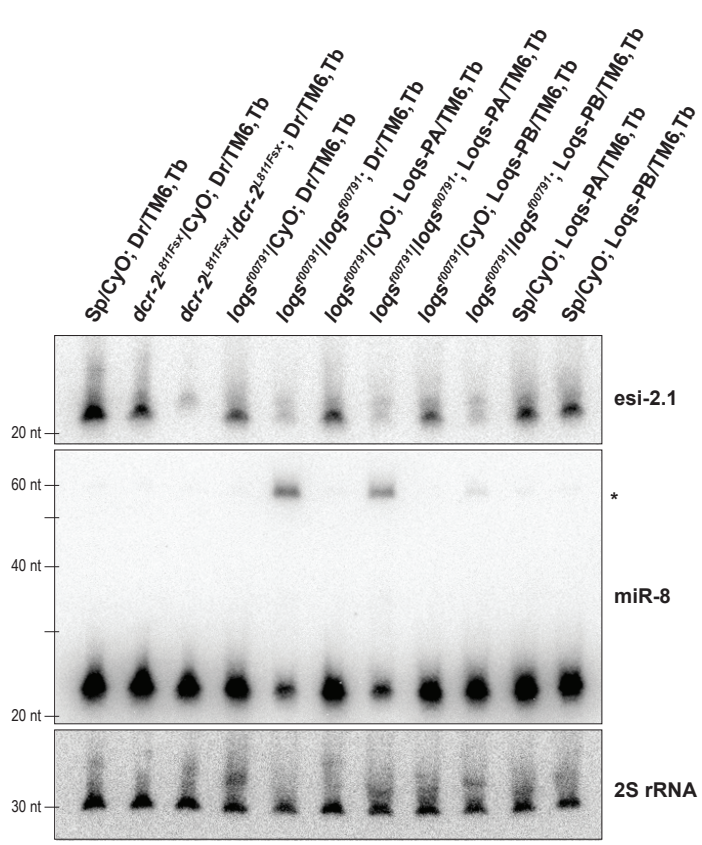
Ye X, Paroo Z, Liu Q. 2007. Functional anatomy of the *Drosophila* microRNA-generating enzyme. *J Biol Chem* **282:** 28373–28378.

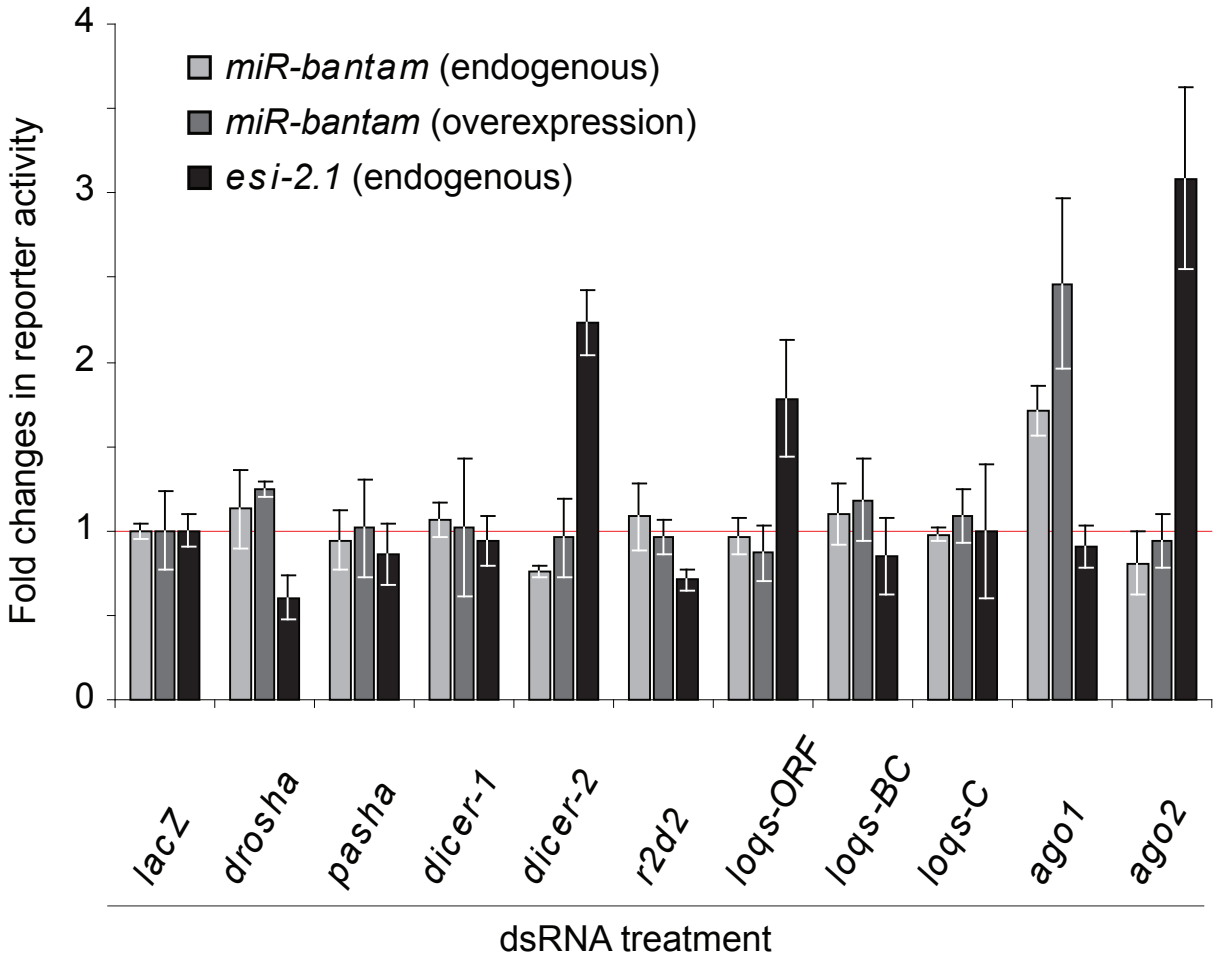
Zhou R, Hotta I, Denli AM, Hong P, Perrimon N, Hannon GJ. 2008. Comparative analysis of argonaute-dependent small RNA pathways in *Drosophila*. *Mol Cell* **32:** 592–599.

A


wild-type
dcr-2L811FsX/CyO
dcr-2L811FsX/dcr-2L811FsX
r2d21/CyO
r2d21/r2d21
loqsf00791/CyO
loqsf00791/loqsf00791
loqsKO/CyO; loqs-PB
loqsKO/loqsKO; loqs-PB
ago2414/TM6
ago2414/ago2414
20 nt
esi-2.1
60 nt
40 nt
20 nt
*
miR-8
30 nt
2S rRNA

B


Sp/CyO; Dr/TM6,Tb
dcr-2L811FsX/CyO; Dr/TM6,Tb
dcr-2L811FsX/dcr-2L811FsX; Dr/TM6,Tb
loqsf00791/CyO; Dr/TM6,Tb
loqsf00791/loqsf00791; Dr/TM6,Tb
loqsf00791/CyO; Loqs-PA/TM6,Tb
loqsf00791/loqsf00791; Loqs-PA/TM6,Tb
loqsf00791/CyO; Loqs-PB/TM6,Tb
loqsf00791/loqsf00791; Loqs-PB/TM6,Tb
Sp/CyO; Loqs-PA/TM6,Tb
Sp/CyO; Loqs-PB/TM6,Tb
20 nt
esi-2.1
60 nt
40 nt
20 nt
*
miR-8
30 nt
2S rRNA


4
3
2
1
0
Fold changes in reporter activity
miR-bantam (endogenous)
miR-bantam (overexpression)
esi-2.1 (endogenous)
lacZ
drosha
pasha
dicer-1
dicer-2
r2d2
loqs-ORF
loqs-BC
loqs-C
ago1
ago2
dsRNA treatment

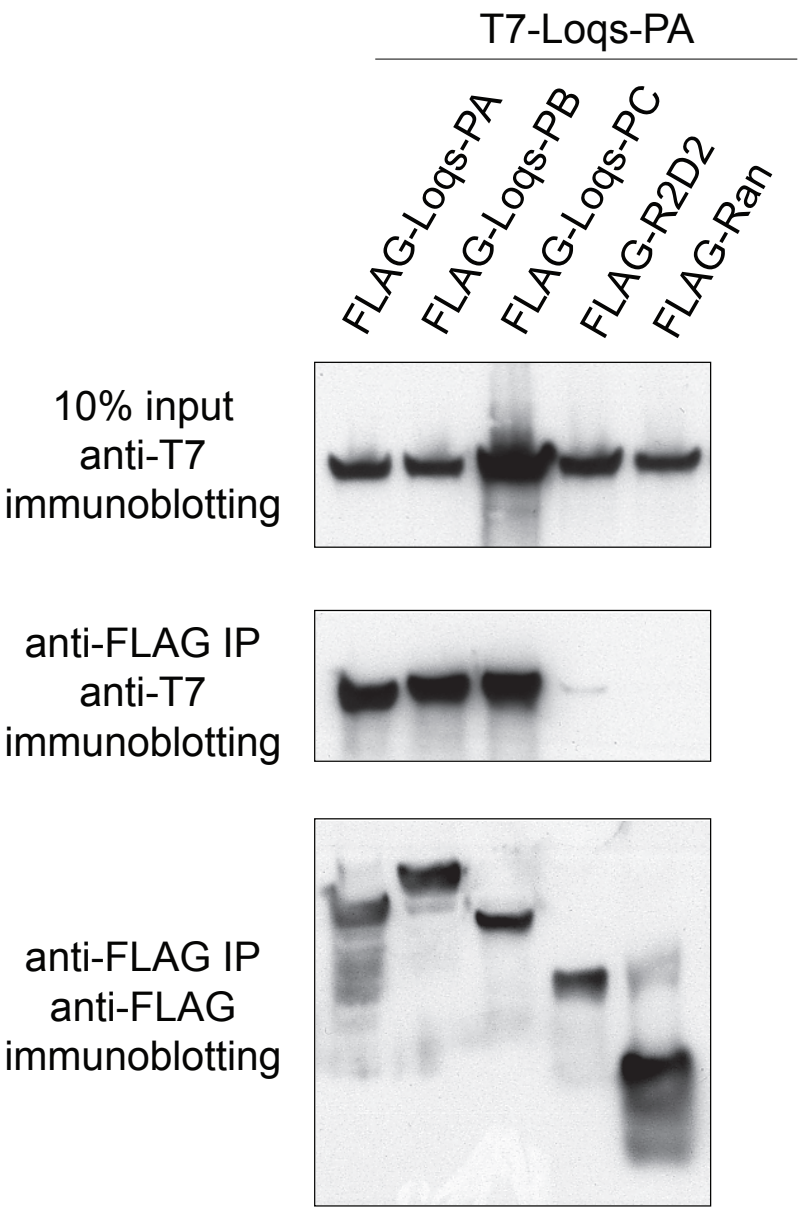

T7-Loqs-PA
FLAG-Loqs-PA
FLAG-Loqs-PB
FLAG-Loqs-PC
FLAG-R2D2
FLAG-Ran
10% input
anti-T7
immunoblotting
anti-FLAG IP
anti-T7
immunoblotting
anti-FLAG IP
anti-FLAG
immunoblotting

Table S1

DNA Oligonucleotides

dsRNA generating PCR primers:

T7-Dicer1-F	TAATACGACTCACTATAGGGTGCGACAACAATCTGC
T7-Dicer1-R	TAATACGACTCACTATAGGGTCAGTTGCTGCAGCTCAC
T7-Dicer2-F	TAATACGACTCACTATAGGGAAGATGTGGAAATCAAGCC
T7-Dicer2-R	TAATACGACTCACTATAGGGCCACGTTCGTAATTTC
T7-Drosha-F	TAATACGACTCACTATAGGGTGAATCAGGACTGGAACG
T7-Drosha-R	TAATACGACTCACTATAGGGAGCCATCGCTATCACTGC
T7-Ago1-F	TAATACGACTCACTATAGGGAACGGACAGACCGTAGAG
T7-Ago1-R	TAATACGACTCACTATAGGGTGGCGTACTTACAGAAGC
T7-Ago2-F	TAATACGACTCACTATAGGGAGCCACATCGACGAACG
T7-Aqo2-R	TAATACGACTCACTATAGGGCGAGGATCATCCTTGATC
T7-R2D2-F	TAATACGACTCACTATAGGGCATACACGGCTTGATGAAGGATTC
T7-R2D2-R	TAATACGACTCACTATAGGGTTGCTTGTGCTCGCTACTTGC
T7-Pasha-F	TAATACGACTCACTATAGGGACTTTGAAGTCCTACCCG
T7-Pasha-R	TAATACGACTCACTATAGGGCTCCTTGAACTCATAGG
T7-Loqs-ORF-F	TAATACGACTCACTATAGGGATGGACCAGGAGAATTTCC
T7-Loqs-ORF-R	TAATACGACTCACTATAGGGAAGGGCGTATCCTTGTC
T7-Loqs-BC268-F	TAATACGACTCACTATAGGGAACGAATCTGTAAAGCACCTTTTC
T7-Loqs-BC268-R	TAATACGACTCACTATAGGGCTGTAACTTAAGCAGTTTTTTGCC
Loqs-C-lacZ-R	CGGATGGTTCGGATAATGCTGTAAATAAGAGCGCAAAGTTTT
Loqs-C-lacZ-F	AAAACTTTGCGCTCTTATTTACAGCATTATCCGAACCATCCG
Loqs-BC-lacZ-R	CGGATGGTTCGGATAATGCTGTAACTTAAGCAGTTTTTTGC
Loqs-BC-lacZ-F	GCAAAAAACTGCTTAAGTTACAGCATTATCCGAACCATCCG
T7-LacZ-R-1580	TAATACGACTCACTATAGGGCCAGGTAGCGAAAGCCAT
T7-Loqs-C-R	TAATACGACTCACTATAGGGCTGTAAATAAGAGCGCAAAGTTTT
T7-Loqs-RD-F	TAATACGACTCACTATAGGGTGAGTATCATTCAAGACATCG
T7-Loqs-RD-R	TAATACGACTCACTATAGGGTAAGGTGTAAGCATTATGTTAATTATATGTATAATTATG
T7-LacZ-F	TAATACGACTCACTATAGGGCATTATCCGAACCATCC
T7-LacZ-R	TAATACGACTCACTATAGGGCAGAACTGGCGATCGTTCG
T7-loqs-3-UTR-F	TAATACGACTCACTATAGGGGTGCAGCCAACTGAATAG
T7-loqs-3-UTR-R	TAATACGACTCACTATAGGGAGTAGCGATTACCAAATACCG
T7-loqs-5-UTR-F	TAATACGACTCACTATAGGGGCAACCACAAATATCAGTAAAATAC
T7-loqs-5-UTR-R	TAATACGACTCACTATAGGGTGTTCTTGTTTTGCACG
Loqs-5-UTR-lacZ-R	CGGATGGTTCGGATAATGGTGTTCTTGTTTTGCAC
Loqs-5-UTR-lacZ-F	GTGCAAAACAAGAACACCATTATCCGAACCATCCG

siRNA sensor oligos:

Bam-si2-S2-2P-F	GATCCCAACAGTTTATTGGAGCGAACTTGTTGGAGTCAAAATGAACTGAGGGTGGAGCGAACTTGTTGGAGTCAAG
Sal-si2-S2-2P-R	TCGACTTGACTCCAACAAGTTCGCTCCACCCTCAGTTCATTTTGACTCCAACAAGTTCGCTCCAATAAACTGTTGG
miR-bantam_guide_bulged-F	TCGAATCAGCTTTCCTATGATCTCAAATGATATGAATCAGCTTTCCTATGATCTCAAATGAACTGAATCAGCTTTCCTATGATCTCA
miR_bantam_guide_bulged-R	TCGATGAGATCATAGGAAAGCTGATTCAGTTCATTTGAGATCATAGGAAAGCTGATTCATATCATTTGAGATCATAGGAAAGCTGAT

RT-PCR:

Loqs-F	GGCAAAATCAGCGACAGC
Loqs-R	CATCGGCAGCTGTTGGTC

Northern probes:

esi-2.1	GGAGCGAACTTGTTGGAGTCAA
esi-1.2	CATTTGATCCATAGTTTCCCG
esi-4.1	CCCTTCGCGTTTGGCCAACTC
miR-bantam	AATCAGCTTTCAAAATGATCTCA
miR-277	TGTCGTACCAGATAGTGCATTTA
miR-8	GACATCTTTACCTGACAGTATTA
DM1731.1	CGGGGTCTTCAGATCACCGCA
DM1731.2	CAACAGCTCTTCAGCTCTGCT
DM1731.3	CAACAAACGTTTGCTCAGCAA
DM1731.4	GCGCCCTAGGAGCTGCCTAAA
DM1731.5	AGACGTCTCTGGTAGAATTGA
DM1731.6	GCCCGTCCGAACAATCTTTGG
DM1731.7	CGAGCCAACCGCACGCTAGTA
DM1731.8	AGCATCCAGATGCGTTCAGTG
Klarsicht.1	GCCCAATCGATCAGAGCCAAT
Klarsicht.2	TGCCAACCGATTGTGGAGCAC
Klarsicht.3	CAGCTGCGGCTGCGATCGGGA
Klarsicht.4	GCCGCATCCGAGTCGAAATTG
Klarsicht.5	ACGAAACTCTGTAAAACTTAA
Klarsicht.6	GACACCCATTCGATAGAATCG
Klarsicht.7	GCACAGGAAGAAACGCTCGAA
Klarsicht.8	ACGCCAAGTTTGGTGTCAAAA
2S rRNA	TACAACCCTCAACCATATGTAGTCCAAGCA

Oligos for cloning:

Sac-Fla-Dcr2-F1	GCCGAGCTCATGGACTACAAGGACGACGATGACAAGATGGAAGATGTGGAAATCAAG
Kpn-Dcr2-R5169	GGGGTACCTTAGGCGTCGCATTTGCTTAG
EcoR-Fla-Loq-F1	GGAATTCATGGACTACAAGGACGACGATGACAAGATGGACCAGGAGAATTTCC
EcoR-T7-Loq-F1	GGAATTCATGGCTAGCATGACTGGTGGACAGCAAATGGGTATGGACCAGGAGAATTTCC
Bam-Loq-R-1398	GGGGATCCTACTTCTTGGTCATGATC
EcoR-Fla-Ran-F1	GGAATTCATGGACTACAAGGACGACGATGACAAGATGGCTCAGGAAGGTCAG
EcoR-T7-Ran-F1	GGAATTCATGGCTAGCATGACTGGTGGACAGCAAATGGGTATGGCTCAGGAAGGTCAG
Bam-Ran-R-651	GGGGATCCTTATAGCTCCTCGTCCTC
Kpn-Fla-R2D2-F1	GGGGTACCATGGACTACAAGGACGACGATGACAAGATGGATAACAAGTCAGCC
Kpn-T7-R2D2-F1	GGGGTACCATGGCTAGCATGACTGGTGGACAGCAAATGGGTATGGATAACAAGTCAGCC
Bam-R2D2-R-936	GGGGATCCTTAAATCAACATGGTGCGAAAATAG
new-MCS-NTAP-F	TCGAGGAGCTCGGTACCCCCGGGTCTAGAG
new-MCS-NTAP-R	GATCCTCTAGACCCGGGGGTACCGAGCTCC
Sac-Loqs-PA-F1	GCCGAGCTCATGGACCAGGAGAATTTCC
Kpn-Loqs-PA-R1260	GGGGTACCCTACTTCTTGGTCATGATCTTC
Sac-Fla-Loq-F1	GCCGAGCTCATGGACTACAAGGACGACGATGACAAGATGGACCAGGAGAATTTCC
Bam-Loqs-PC-R	GGGGATCCTACTGCGGGGCTGTAAATAAGAG
Bam-Loqs-PD-R	GGGGATCCTTAGATCTTGATGAACTCAAAATCTTTAGAG

Hierarchical Rules for Argonaute Loading in *Drosophila*

Benjamin Czech*, Rui Zhou*, Yaniv Erlich, Julius Brennecke, Richard Binari, Christians Villalta, Assaf Gordon, Norbert Perrimon and Gregory J. Hannon

* authors contributed equally to this work

Mol Cell. 2009 Nov 13;36(3):445-56.

Molecular Cell
Article

Hierarchical Rules for Argonaute Loading in *Drosophila*

Benjamin Czech,[1,3] Rui Zhou,[2,3] Yaniv Erlich,[1] Julius Brennecke,[1,4] Richard Binari,[2] Christians Villalta,[2] Assaf Gordon,[1] Norbert Perrimon,[2,*] and Gregory J. Hannon[1,*]
[1]Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[2]Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA
[3]These authors contributed equally to this work
[4]Present address: IMBA - Institute of Molecular Biotechnology, 1030 Vienna, Austria
*Correspondence: perrimon@receptor.med.harvard.edu (N.P.), hannon@cshl.edu (G.J.H.)
DOI 10.1016/j.molcel.2009.09.028

SUMMARY

Drosophila **Argonaute-1 and Argonaute-2 differ in function and small RNA content. AGO2 binds to siRNAs, whereas AGO1 is almost exclusively occupied by microRNAs. MicroRNA duplexes are intrinsically asymmetric, with one strand, the miR strand, preferentially entering AGO1 to recognize and regulate the expression of target mRNAs. The other strand, miR*, has been viewed as a byproduct of microRNA biogenesis. Here, we show that miR*s are often loaded as functional species into AGO2. This indicates that each microRNA precursor can potentially produce two mature small RNA strands that are differentially sorted within the RNAi pathway. miR* biogenesis depends upon the canonical microRNA pathway, but loading into AGO2 is mediated by factors traditionally dedicated to siRNAs. By inferring and validating hierarchical rules that predict differential AGO loading, we find that intrinsic determinants, including structural and thermodynamic properties of the processed duplex, regulate the fate of each RNA strand within the RNAi pathway.**

INTRODUCTION

The biogenesis of small RNAs derived from double-stranded or structured precursors requires the action of RNase III family proteins. In *Drosophila*, these small RNAs interact with the two AGO clade proteins, Argonaute-1 (AGO1) and Argonaute-2 (AGO2), and represent two major classes, microRNAs (miRNAs) and small interfering RNAs (siRNAs), respectively.

siRNAs are processed from exogenous dsRNAs by a dedicated Dicer protein, Dcr-2, and its cofactor, R2D2 (Lee et al., 2004b; Liu et al., 2003). Dcr-2 and R2D2 additionally function during siRNA loading into AGO2 (Tomari et al., 2004). In a mature complex, only one siRNA strand, the guide strand, is retained. The remaining strand, the passenger strand, is cleaved by AGO2 and ultimately degraded (Matranga et al., 2005; Miyoshi et al., 2005).

Endogenously encoded double-stranded RNAs can also form siRNAs, endo-siRNAs (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008a). These can be derived from dedicated noncoding transcripts that are extensively structured, from intermolecular hybrids of RNAs from convergently transcribed genes, or from transposon loci, which form dsRNA through unknown mechanisms. Endo-siRNAs are processed by Dcr-2 but lack a strong dependency on R2D2 (Czech et al., 2008; Okamura et al., 2008a). Instead, they rely upon a specific isoform of the dsRNA binding protein, Loquacious (Loqs-PD) (Czech et al., 2008; Hartig et al., 2009; Okamura et al., 2008a; Zhou et al., 2009). Both endo- and exo-siRNA primed AGO2 execute efficient small RNA-directed cleavage of complementary targets (Czech et al., 2008; Hammond et al., 2000). Moreover, all AGO2-bound guide strands become 2′-*O*-methyl modified at their 3′ termini by the methytransferase Hen1/Pimet (Horwich et al., 2007; Saito et al., 2007).

In contrast to AGO2, AGO1 principally hosts miRNAs. These are derived mainly from long RNA polymerase II transcripts through two site-specific cleavages. The first is catalyzed by Drosha/Pasha complexes (Denli et al., 2004; Gregory et al., 2004; Lee et al., 2003, 2004a) and the second by Dcr-1 in collaboration with another Loquacious isoform, Loqs-PB (Förstemann et al., 2005; Jiang et al., 2005; Park et al., 2007; Saito et al., 2005). The product of Dcr-1 cleavage is a duplex comprised of the miRNA (miR) and the miRNA-star (miR*) strands, with the miR corresponding to the guide strand and the miR* resembling the passenger strand. Loading of these duplexes into AGO1 followed by unwinding and degradation of the miR* strand leads to mature RISC. The miR strand guides AGO1 to mRNA targets, which are generally recognized by imperfect base-pairing interactions. Recognition by miRNAs generally leads to repression via reduction in protein synthesis. Although both AGO1 and AGO2 can act via this mechanism (Förstemann et al., 2007; Iwasaki et al., 2009), AGO1 seems biochemically optimized for cleavage-independent repression, while AGO2 is optimized as a multiturnover nuclease (Förstemann et al., 2007).

Based upon these observations, small RNAs in the RNAi pathway must be sorted in several ways. First, different types of small RNA duplexes are directed toward specific AGO complexes. Second, the individual strands of each small RNA duplex have a different probability of guiding mature RISC. As

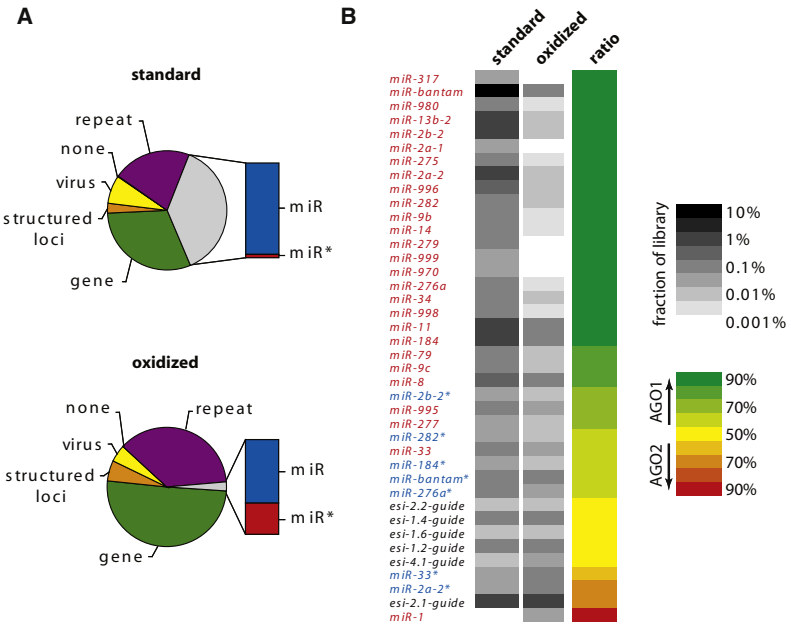


Figure 1. miR*s Have Modified 3′ Termini

(A) Pie charts represent the relative abundance of different endo-siRNA classes and miRNAs in 19–24 nt small RNA libraries from wild-type S2 cells. Results from a standard cloning protocol (upper diagram) and from a cloning strategy that enriches for small RNAs with modified 3′ termini (lower diagram) are shown. The fraction of miRs and miR*s is indicated for both libraries.

(B) Heat maps show the relative abundance of endo-siRNAs derived from structured loci, miRs, and miR*s in the indicated libraries (in grayscale). The ratio of normalized representation in the libraries indicates preferential association of small RNAs with either AGO1 (green) or AGO2 (red).

a consequence of coupled dicing and loading, selective incorporation into AGO1 or AGO2 could rely in part on the distinct enzymatic machinery underlying the biogenesis of siRNAs and miRNAs. However, at least one miRNA, *miR-277*, is substantially AGO2 loaded, although it is processed conventionally by Dcr-1 and Loqs (Förstemann et al., 2007). In contrast to many miRNA precursors, which contain several mismatches and bulges, the duplex precursor to *miR-277* has an unusual degree of perfect double-stranded character and therefore strongly resembles a siRNA precursor. Moreover, alterations in the extent of pairing in miRNA-mimetic siRNA duplexes allowed experimental direction to AGO1 or AGO2 preferentially (Tomari et al., 2007). The discrimination of miR and guide strands from miR* and passenger strands is proposed to rely upon the thermodynamic properties of the processed duplexes. In both cases, the strand with the less-stable 5′ end preferentially enters RISC.

Conventional wisdom holds that the passenger and miR* strands are simply byproducts of siRNA and miRNA biogenesis and RISC loading and are, therefore, discarded and degraded. However, in our studies of AGO2-bound small RNA species, we noted that a wide range of miR* strands represented some of the most abundant individual species in AGO2 RISC. This indicated that, following processing by Dcr-1, the miR:miR* duplex could be bifunctional, flowing down either the AGO1 or AGO2 loading pathway with the properties of each individual strand determining its destination. By studying the patterns of mismatches and thermodynamic stabilities of precursors to small RNAs resident within each complex and by selectively manipulating these characteristics, we find that a hierarchy of rules, depending both on duplex structure and thermodynamic properties, determines the fate of small RNAs in the RNAi pathway.

RESULTS

miR* Strands Often Bear 2′-*O*-Methylated 3′ Termini

We sought to investigate the fates of dsRNA-derived small RNAs and their flow through the RNAi pathway. We began by sequencing a 19–24 nt small RNA library from wild-type *Drosophila* S2 cells using our standard cloning protocol (“standard”) (Figure 1). In parallel, we analyzed a library enriched for small RNAs with 2′-*O*-methylated 3′ termini (“oxidized”) prepared using a modified cloning strategy (Seitz et al., 2008). After removing degradation products of abundant cellular RNAs, sequences were split into six categories: endo-siRNAs corresponding to (1) genes, (2) structured loci, (3) repeats, (4) viruses, and (5) genomic regions without annotation (“none”) and (6) miRNA (miR or miR*). Within the standard library, 62.6% of all sequences fell within different endo-siRNA classes. The remaining 37.4% corresponded to miRNA sequences, of which the vast majority derived from mature miRNA strands (Figure 1A). Consistent with previous reports that *Drosophila* miRNAs lack methylated 3′ termini (Horwich et al., 2007; Saito et al., 2007), miRNA species were significantly depleted in the oxidized library. There, 97.7% reads could be assigned endo-siRNAs, while only 2.3% corresponded to miRNA sequences. Within the remaining miRNA sequences, mature miRNA strands were strongly depleted, while levels of miRNA* strands did not change substantially. Specifically, ratios between miR and miR* strands changed from ~33:1 in the standard library to ~2:1 in the oxidized library, which corresponds to a 16-fold relative enrichment of miR*. Consistent with previous reports of siRNAs derived from the flock house virus (FHV) being only partially methylated (Aliyari et al., 2008; Flynt et al., 2009), viral siRNAs (more than 99% of our viral siRNAs matched to the FHV genome) were also reduced in the oxidized library. All other categories of endo-siRNAs were enriched by the modified cloning strategy (Figure 1A), consistent with the RNAs bearing modified 3′ termini (Chung et al., 2008; Kawamura et al., 2008; Okamura et al., 2008a). We plotted the cloning frequencies of the 40 most abundant sequences in each library corresponding to miRs (red text), miR*s (blue text), and endo-siRNAs from structured loci (black text) (Figure 1B). We calculated the relative

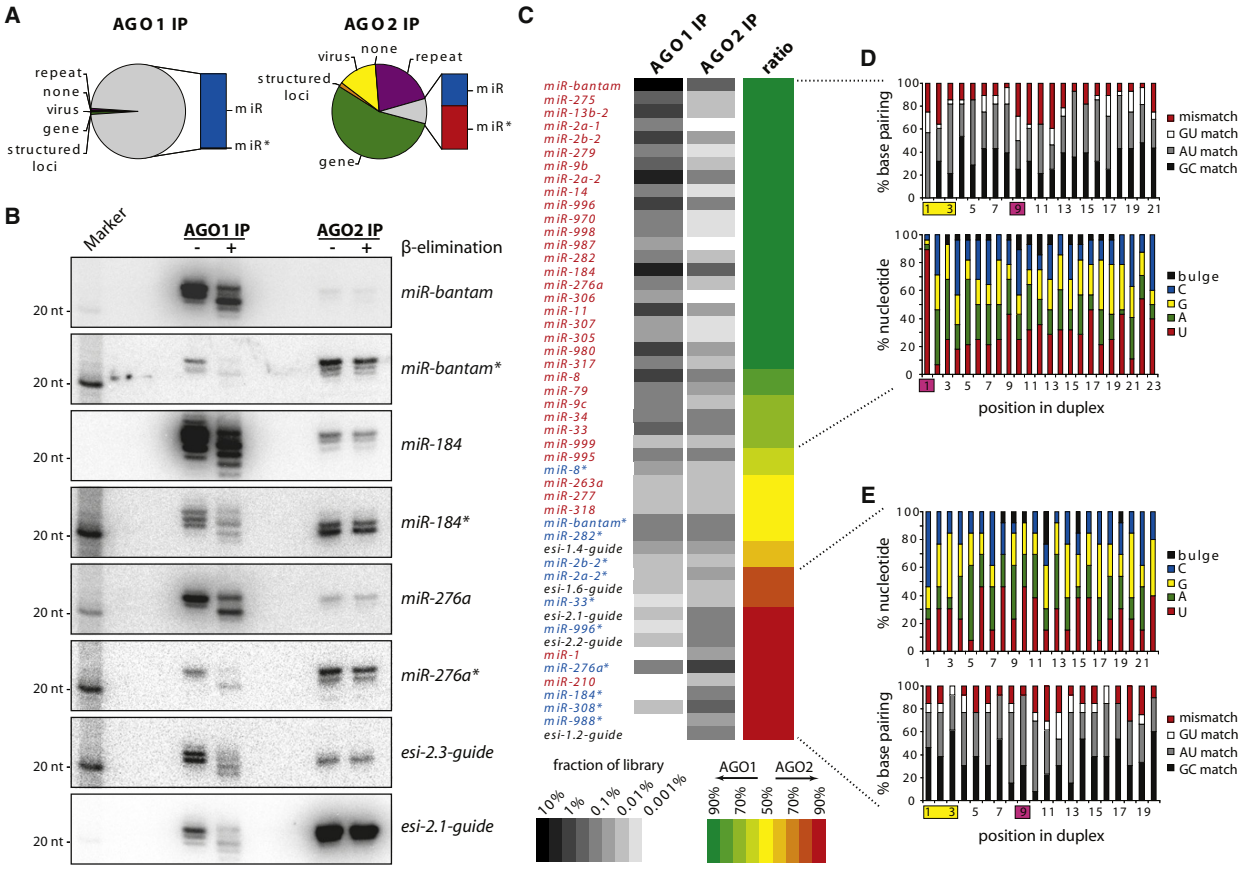


Figure 2. miR*s Are Preferentially Loaded into AGO2

(A) Pie charts show the relative abundance of endo-siRNA classes and miRNA libraries from AGO1 (left diagram) and AGO2 (right diagram) immunoprecipitates from S2 cells.

(B) Northern blots of RNA from AGO1 and AGO2 immunoprecipitates from S2 cells. AGO-bound small RNAs were untreated (−) or subjected to β-elimination (+) prior to gel electrophoresis. The same membrane was probed for three miRs, three miR*s, and two endo-siRNAs derived from structured loci.

(C) Heat maps showing the relative abundance of endo-siRNAs derived from structured loci, miRs, and miR*s in AGO1 and AGO2 libraries (grayscale). The relative association of small RNAs with AGO1 or AGO2 is indicated on a red/green scale.

(D) Median base-pairing (upper chart) and nucleotide composition (lower chart) of all sequences that show a relative association with AGO1 of 70% or more. Bulges on each strand were counted as mismatches.

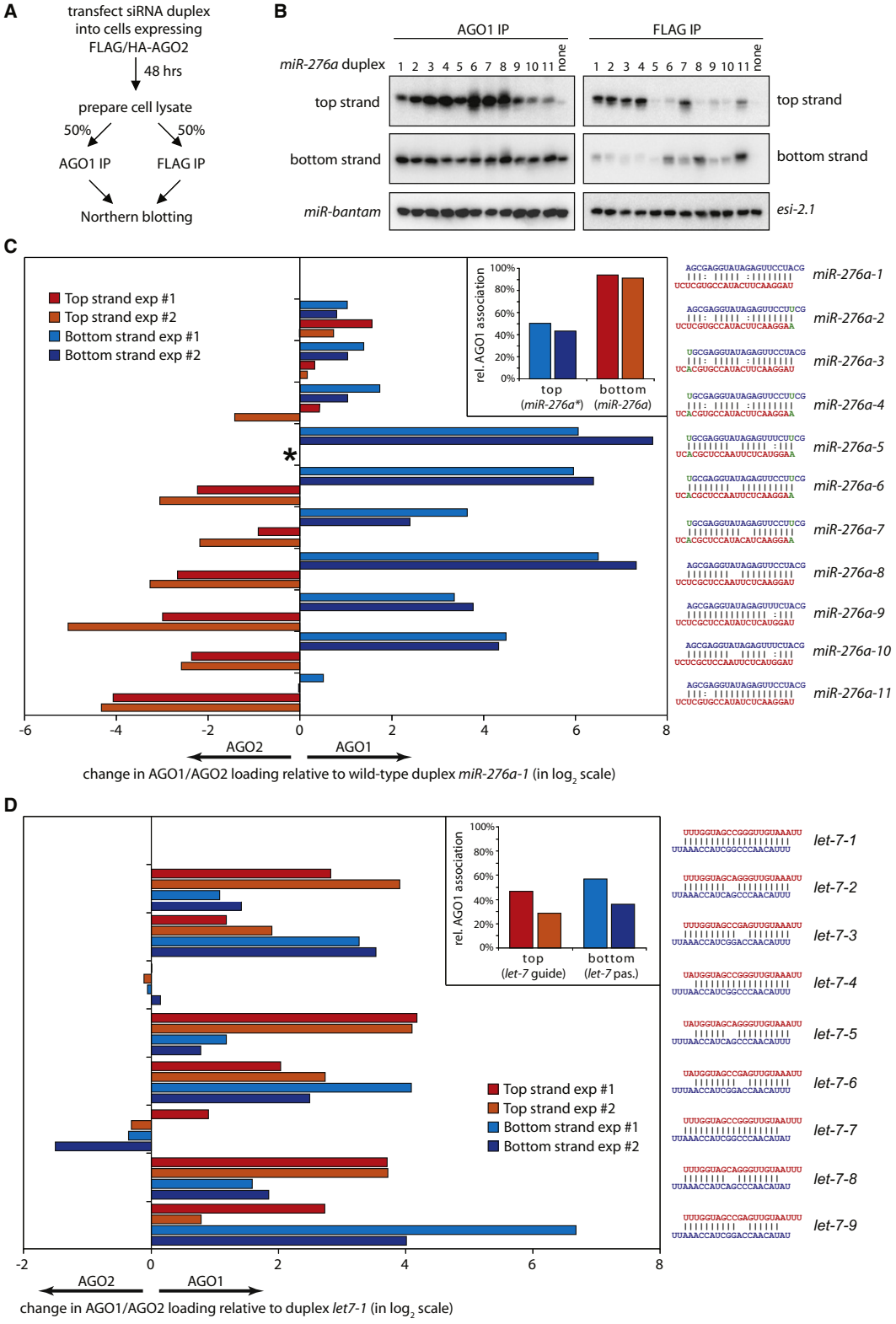
(E) Analysis as in (D) but with all sequences having a relative association of 70% or more with AGO2.

representation of each sequence in the two libraries and sorted by this ratio. Green bars indicate enrichment in the standard library, and red bars indicate enrichment in the oxidized library. Since 2′-*O*-methylation is characteristic of AGO2-loaded sequences, this ratio can also be taken as a rough surrogate for relative loading into AGO1 and AGO2 complexes. The results of this analysis are consistent with previous reports of miRNAs principally occupying AGO1 and endo-siRNAs occupying AGO2 (Figure 1B). Notably, these data also indicated that miR* strands were individually abundant within AGO2 complexes.

miR* Strands Primarily Associate with AGO2

To confirm the patterns of small RNA loading, we examined small RNA libraries from immunoprecipitates of AGO1 and AGO2 from *Drosophila* S2 cells (Czech et al., 2008), separating miRNA-related sequences into miR and miR* strands. Approximately 98% of all AGO1-associated reads match to annotated miRNAs, with 99% of these representing the miR strand. In contrast to recent reports, we did not observe significant loading of miR*s into AGO1 (Okamura et al., 2008b). The remaining AGO1-associated sequences comprised distinct classes of endo-siRNAs, including genic and viral sequences (Figure 2A). In contrast, AGO2 is predominantly loaded with all classes of endo-siRNAs. Approximately 8% of all reads in AGO2 immunoprecipitates match to miRNAs. Among the AGO2-associated miRNA sequences, only ~40% matched to the miR strand, while almost 60% represented miR* strands (Figure 2A).

To verify conclusions emerging from deep sequencing, we prepared total RNA from AGO1 and FLAG immunoprecipitates from a stable S2 line expressing FLAG/HA-AGO2 under its endogenous regulatory elements (Czech et al., 2008) (Figure S1A) and subjected a fraction of this material to β-elimination. Treated


A
transfect siRNA duplex into cells expressing FLAG/HA-AGO2
48 hrs
prepare cell lysate
50%
50%
AGO1 IP
FLAG IP
Northern blotting
B
AGO1 IP
FLAG IP
miR-276a duplex
1 2 3 4 5 6 7 8 9 10 11 none
top strand
bottom strand
miR-bantam
esi-2.1
C
Top strand exp #1
Top strand exp #2
Bottom strand exp #1
Bottom strand exp #2
rel. AGO1 association
top (miR-276a*)
bottom (miR-276a)
AGO2
AGO1
change in AGO1/AGO2 loading relative to wild-type duplex miR-276a-1 (in log2 scale)
miR-276a-1
miR-276a-2
miR-276a-3
miR-276a-4
miR-276a-5
miR-276a-6
miR-276a-7
miR-276a-8
miR-276a-9
miR-276a-10
miR-276a-11
D
rel. AGO1 association
top (let-7 guide)
bottom (let-7 pas.)
Top strand exp #1
Top strand exp #2
Bottom strand exp #1
Bottom strand exp #2
AGO2
AGO1
change in AGO1/AGO2 loading relative to duplex let7-1 (in log2 scale)
let-7-1
let-7-2
let-7-3
let-7-4
let-7-5
let-7-6
let-7-7
let-7-8
let-7-9

and untreated RNAs were blotted with probes specific to the miR and miR* strands of three miRNAs, *miR-bantam*, *miR-184*, and *miR-276a*. miR-strand probes for all three miRNAs generated strong signals in AGO1 immunoprecipitates and were only weakly, if at all, detected in AGO2 immunoprecipitates (Figure 2B). In contrast, all three miR* probes detected strong signals selectively in AGO2 immunoprecipitates. As expected, the endo-siRNA, *esi-2.1*, strongly associated with AGO2 (Czech et al., 2008). RNAs coimmunoprecipitated with AGO1 were sensitive to periodate treatment followed by β-elimination. However, all AGO2-associated RNAs, including the low-abundance AGO2-associated miR strands, were completely resistant to β-elimination (Figure 2B).

Patterns observed by northern blotting were also apparent in an analysis of the most abundant sequences derived from AGO1 and AGO2 complexes (Figure 2C). In AGO1 complexes, miR strands (red text) were strongly enriched, whereas miR*s (blue text) and endo-siRNAs (black text) were rare. In AGO2, miR*s and endo-siRNAs were cloned at higher frequencies. Consistent with a previous report (Förstemann et al., 2007), we also observed a significant proportion of *miR-277* in AGO2.

Our data imply that AGO1 and AGO2 loading rest on a more complex set of parameters than was previously supposed (Förstemann et al., 2007; Tomari et al., 2007). We therefore analyzed the properties of sequences that showed strong preferential (>70%) association with either AGO1 or AGO2 (Figures 2D and 2E). We assessed overall base-pairing patterns and the distributions of mismatches within miR:miR* and endo-siRNA guide:passenger duplexes and determined their positional nucleotide biases. In general, duplexes sorted to AGO1 contained slightly higher frequencies of mismatched bases than those sorted to AGO2, indicating that overall pairing is a minor determinant of small RNA sorting. Nucleotide biases were prominent for AGO1-loaded RNAs, with the previously noted strong enrichment for a 5′ U in miRNAs being easily observed (Figure 2D). Most of either AGO1- or AGO2-destined duplexes showed standard Watson-Crick base pairs across their first two residues, with rates reaching 80% for AGO2 but only 60% for AGO1 (Figures 2D and 2E). In AGO2-bound RNAs, there was an enrichment for a terminal C residue (~50% of sequences).

Strong differences were detected in the structure of the central regions of duplexes sorted to AGO1 and AGO2. In particular, the strand destined for AGO1 was often unpaired at position 9, while pairing at this position occurred in more than 90% of AGO2-associated strands. This pattern not only held for miR and miR* strands but also for the guide and passenger strands of endo-siRNAs. For example, both deep sequencing (data not shown) and northern blotting (Figure 2B) highlighted the guide strand of one endo-siRNA, *esi-2.3*, that acted anomalously, preferentially entering AGO1 rather than AGO2 complexes. Notably, in its precursor duplex, *esi-2.3* shows central mismatches characteristic of miR strands (Figure S2). Thus, a combination of sequence and structural determinants contributes to strand and small RNA sorting in the RNAi pathway, and these characteristics dominate over signals emanating from the upstream biogenesis pathways.

Validating Rules for Strand Sorting

To assess the relevance of our observations for small RNA strand sorting, S2 cells stably expressing FLAG/HA-AGO2 were transfected with altered *miRNA-276a* and *let-7* siRNA duplexes, and AGO1 and AGO2 complexes were subsequently recovered by immunoprecipitation (Figure 3A). Differential loading was probed by northern blotting (Figures 3B and S3). Levels of both top (miR* for *miR-276a*, guide for *let-7*) and bottom (miR for *miR-276a*, passenger for *let-7*) strands were normalized to nontransfected controls, and relative Argonaute loading indices for each strand were calculated compared to corresponding wild-type controls (Figures 3C and 3D). We found that both strands of the perfectly matched *let-7-1* duplex showed relatively strong association with AGO2 (Figure 3D). The insertion of central bulges or mismatches at the ends of *let-7* duplexes caused a general shift of both top (guide) and bottom (passenger) strands toward AGO1. We observed stronger effects on AGO1 loading for the strand featuring central bulges around position 9, as measured from its 5′ end (compare *let-7-4* and *let-7-7* with *let-7-2* and *let-7-3*). Introduction of mismatches at positions 9 and 10 caused a stronger preference for AGO1 loading than introduction of mismatches at positions 11 and 12 (compare the top strand with the bottom strand of *let-7-2* and *let-7-3*), in accord with our analysis of naturally AGO1-associated miRNA strands (Figure 2D). The combination of central bulges with unpaired terminal nucleotides in reciprocal configurations caused both strands to favor AGO1 (*let-7-5*, *let-7-6*, *let-7-8*, and *let-7-9*). However, the effects of central mismatches at positions 9 and 10 still showed a stronger impact than did alterations of duplex ends (compare *let-7-5* with *let-7-6* and *let-7-8* with *let-7-9*).

Generally, consistent results were obtained for sorting of *miR-276a* duplexes (Figure 3C). Changing the 5′ uracil of the miR strand (bottom, in red) to adenine did not extinguish AGO1 loading (*miR-276a-2*), while substitution of the 5′ adenine of the miR* strand (top, in blue) to uracil did cause a slight shift toward AGO1 (*miR-276a-3*). Modifying the terminal nucleotides of both strands at once failed to trigger more dramatic changes in AGO preference than did single substitutions, indicating that the observed nucleotide bias of miRNAs has a minor, if any,

Figure 3. Small RNA Duplexes Can Be Directed to AGO1 or AGO2

(A) Schematic drawing of the experimental procedure (Argonaute loading assay).

(B) Immunoprecipitation followed by northern blotting shows the loading of both top and bottom strands of various modified *miR-276a* duplexes into AGO1 or AGO2. *miR-bantam* and *esi-2.1* served as controls.

(C) Quantification of the Argonaute loading assay for modified *miR-276a* duplexes. The relative Argonaute loading index for each strand was normalized to that of the corresponding strand of duplex #1 (wild-type control); results were $\log_{(2)}$ transformed and plotted. Positive numbers indicate preferential loading into AGO1, whereas negative numbers indicate favored loading into AGO2. The asterisk indicates that the bottom strand of duplex 5 had low signal and could not be reliably quantified. The inset shows the loading pattern of both individual strands of duplex #1. Duplex structures are shown to the right.

(D) The relative Argonaute loading index for modified *let-7* duplexes as described in (C).

impact on sorting behavior (*miR-276a-4*). Next, we combined modification of terminal nucleotides with altered central bulges by inserting mismatches at positions 9 and 10 counted from the 5′ end of either the top or bottom strands. Alteration of the miR* strand combined with reversed terminal nucleotides (*miR-276a-6*) caused a dramatic shift of the miR* toward AGO1, while the miR strand was moderately shifted toward AGO2. Similar results were obtained if central mismatches only were introduced into the miR* strand (*miR-276a-8*) or if the central mismatches were combined with mismatches in the seed region of the miR strand (*miR-276a-5* and *miR-276a-10*). Sealing the central mismatches in the miR strand either alone (*miR-276a-11*) or in combination with a reversion of seed mismatches (*miR-276a-9*) biased the miR strand toward AGO2 as compared to the wild-type duplex. Considered together, we conclude that central mismatches are the dominant determinant for sorting of small RNAs among AGO1 and AGO2 complexes, while the overall pairing within the duplex also contributes, albeit to a lesser extent. Central mismatches also contribute to the decision of which strand is loaded, while thermodynamic properties become important for duplexes with relatively perfect dsRNA character.

Biogenesis of miRNA* Strands

Since our results pointed to bifunctionality within miRNA precursors, we wished to compare the requirements for processing and loading of miR and miR* strands. We depleted canonical components of the miRNA and endo-siRNA pathways in S2 cells and examined the impacts on levels of miRNAs, miR*s, and endo-siRNAs derived from structured loci. RNAs from the indicated knockdowns were split and subjected to β-elimination or left untreated prior to northern blotting. Knockdown of established miRNA pathway components generally had consistent effects on miR and miR* strands. Reduction of *drosha* and *pasha* together led to a decrease in both the miR and miR* strands, while endo-siRNA levels were not affected (Figure 4A). Depletion of Dcr-1 caused accumulation of pre-miRNAs and slightly reduced the levels of mature miRs and miR*s, while not affecting endo-siRNAs. In contrast, knockdown of some siRNA pathway components showed differential effects on miR and miR*. Knockdown of *dcr-2* or *loqs* had no effect on either miR or miR* levels, while endo-siRNAs were strongly reduced. However, depletion of Dcr-2 or R2D2 did cause significant band shifts for β-eliminated RNAs corresponding to miR*s. Upon AGO1 depletion, we noted a significant reduction in mature miRNA strands and an unexpected concomitant increase in the levels of *miR-bantam** and *miR-276a**. The latter resisted β-elimination, indicating proper loading into AGO2. Finally, depletion of AGO2 caused a reduction of endo-siRNA and miR* levels, while miRNA levels were unaffected. Consistent with the requirement of AGO2 binding for terminal methylation, miR*s remaining in *ago2* knockdowns had completely lost their resistance to β-elimination.

To probe the effects of AGO1 and AGO2 depletion more broadly, we sequenced small RNAs from knockdown cells (Figure 4B). By comparing individual sequences within these libraries, we could establish relative dependence on the two AGO proteins. miR*s and endo-siRNAs showed more dependence on AGO2, whereas miRNAs were more dependent on AGO1 (Figure 4B). We also examined the small RNA populations associated with AGO1 or AGO2 in cells depleted of Dcr-2 and observed a significant decrease in the miR* fraction within AGO2-bound miRNAs as compared to control samples (Figure S4). These results are consistent with miR*s being predominantly associated with AGO2 and depending upon components of the miRNA pathway for processing and components of the siRNA pathway for loading, stabilization, and 3′ end modification.

miR* Strands Can Silence Targets In Vitro

miR* strands show abundances in AGO2 RISC similar to those of endo-siRNAs, which are competent to silence target RNAs (Czech et al., 2008; Okamura et al., 2008b). We therefore tested whether AGO2-loaded miR*s could repress sensors carrying perfect complementary sites. Since a recent report employed AGO2 in the regulation of bulged target sites, we also probed the impact of miR* strands on sensors carrying imperfect sites (Iwasaki et al., 2009). We generated *Renilla* luciferase reporter constructs that carry multiple perfect or bulged binding sites for either the miR or miR* strand of *miR-276a* or *miR-bantam* (Figure 5A). These sensor constructs were transfected into S2 cells together with dsRNAs targeting canonical miRNA and siRNA pathway components, and the impact of depletion of these factors on reporter activity was examined. As expected, depletion of Drosha caused a consistent derepression of all sensors for the miR strand of *miR-276a* or *miR-bantam*. Importantly, Drosha depletion also led to a similar derepression of all sensors for miR* strands, indicating that these are also capable of repressing mRNA targets (Figures 5B and 5C). While depletion of Pasha or Dcr-1 caused a moderate derepression of sensors for endogenous miR or miR* strands, we observed a more consistent phenotype following overexpression of primary miRNAs (Figures 5B–5E). In addition, the sensor constructs for either *miR-276a** or *miR-bantam** in a "perfect match" configuration were derepressed upon depletion of AGO2, consistent with their acting in a complex with this protein (Figures 5B and 5C). This was dependent on Dcr-2 and R2D2, but not Loqs. Most notably, depletion of AGO1 enhanced the repression of the same set of sensors, in accord with the observed increase in miR* strands in knockdown cells (Figures 5B and 5C). Similar changes in sensor activity were observed when pri-miRNAs were overexpressed (Figures 5D and 5E). We therefore conclude that miR* strands are capable of silencing target transcripts carrying either perfect or imperfect complementary sites in cultured S2 cells and that the silencing of "perfect match" targets by miR* species depends on canonical siRNA pathway components.

miR* Strands Can Silence Targets In Vivo

To test whether the miR* strands also function in vivo, we generated transgenic sensor flies in which binding sites for either strand of *miR-276a* or *miR-bantam* in perfect or bulged configurations were placed within the 3′ UTR of an *EGFP* transgene. We tested silencing using clonal analyses in the developing wing disc. In homozygous *dcr-1* clones, GFP signals from sensors for the miRNA strand of *miR-276a* or *miR-bantam* (in both perfect and bulged configurations) increased as expected (Figures 6A,

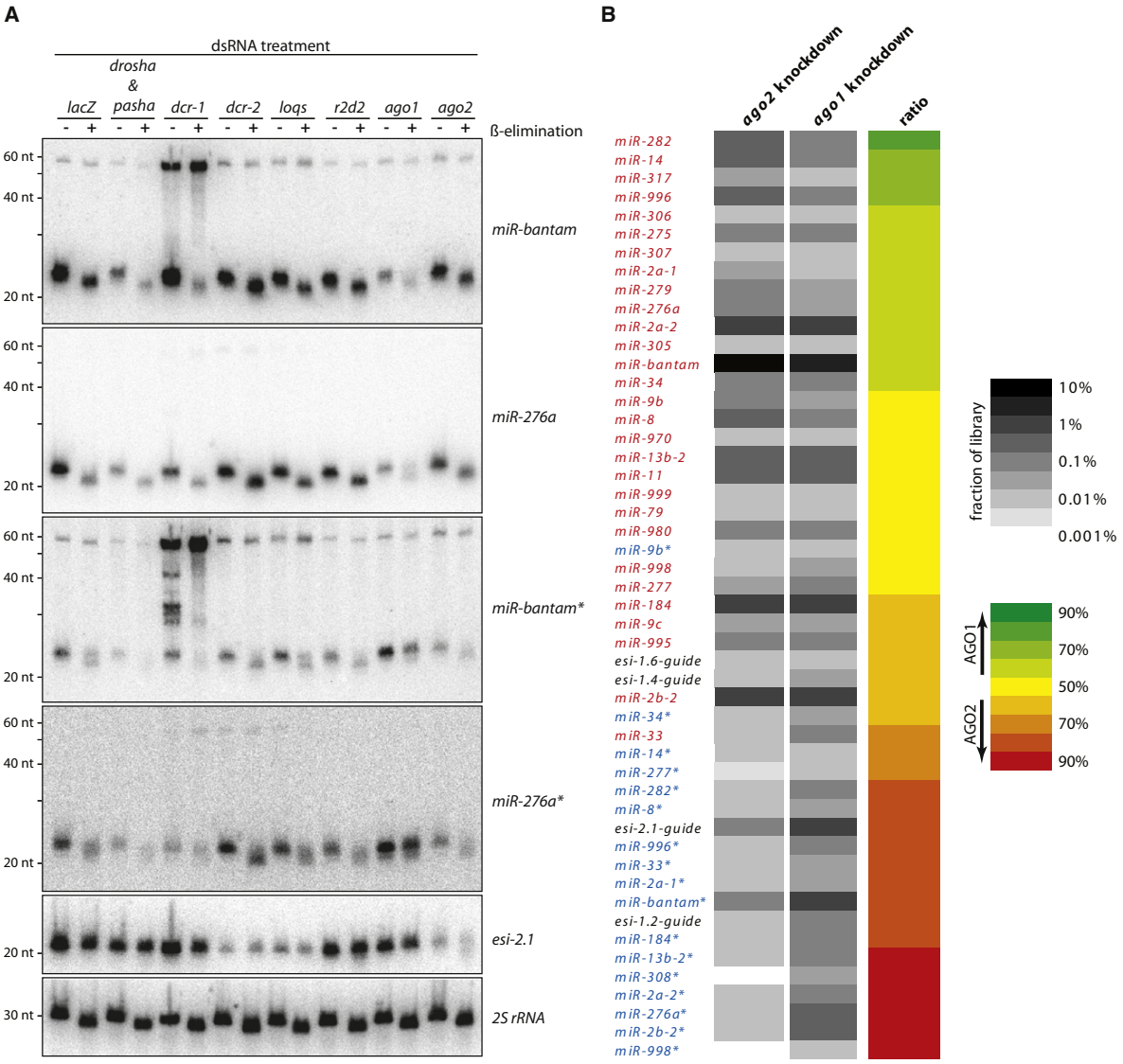


Figure 4. Requirements for Biogenesis and Loading of miR*s

(A) Northern blots were probed with two miRs, two miR*s, and an endo-siRNA derived from a structured locus. Total RNAs from the indicated RNAi knockdowns were untreated (−) or subjected to β-elimination (+) prior to gel electrophoresis. 2S rRNA served as loading control.

(B) Heat maps showing the relative abundance of miRs, miR*s, and endo-siRNAs derived from structured loci in total RNA libraries of samples treated with dsRNAs against AGO1 or AGO2 (in grayscale). Preferential dependence of small RNAs on AGO1 (green) or AGO2 (red) is shown to the right.

6B, S5A, and S5B). Sensors for the miR* strand of *miR-bantam* (in perfect and bulged configurations) were also derepressed in *dcr-1* clones (Figures 6C and 6D). We did not observe the same effect with sensors for the *miR-276a** strand, presumably due to its low endogenous levels in the wing disc (Figures S5C and S5D). We conclude that the *miR-bantam** strand is generated in a Dcr-1-dependent manner and is capable of repressing sensors carrying either perfect or bulged binding sites.

In *ago1* clones, perfectly complementary sensors for the miR strand of *miR-276a* or *miR-bantam* were derepressed (Figures 6E and S5E), as were sensors for the miR strand of *miR-276a* or *miR-bantam* in bulged configurations (Figures 6F and S5F). In *ago1* mutant clones, we found that perfect match sensors for the miR* strand of *miR-bantam* became hyper-repressed as compared to background tissue, which is heterozygous for the *ago1* mutation. We saw concomitant derepression in the twin spots, which carry two copies of the wild-type *ago1* gene (Figure 6G). The increase in silencing upon AGO1 depletion is consistent with effects of *ago1* knockdown in S2 cells (Figures 5B–5E). We were unable to detect significant derepression of

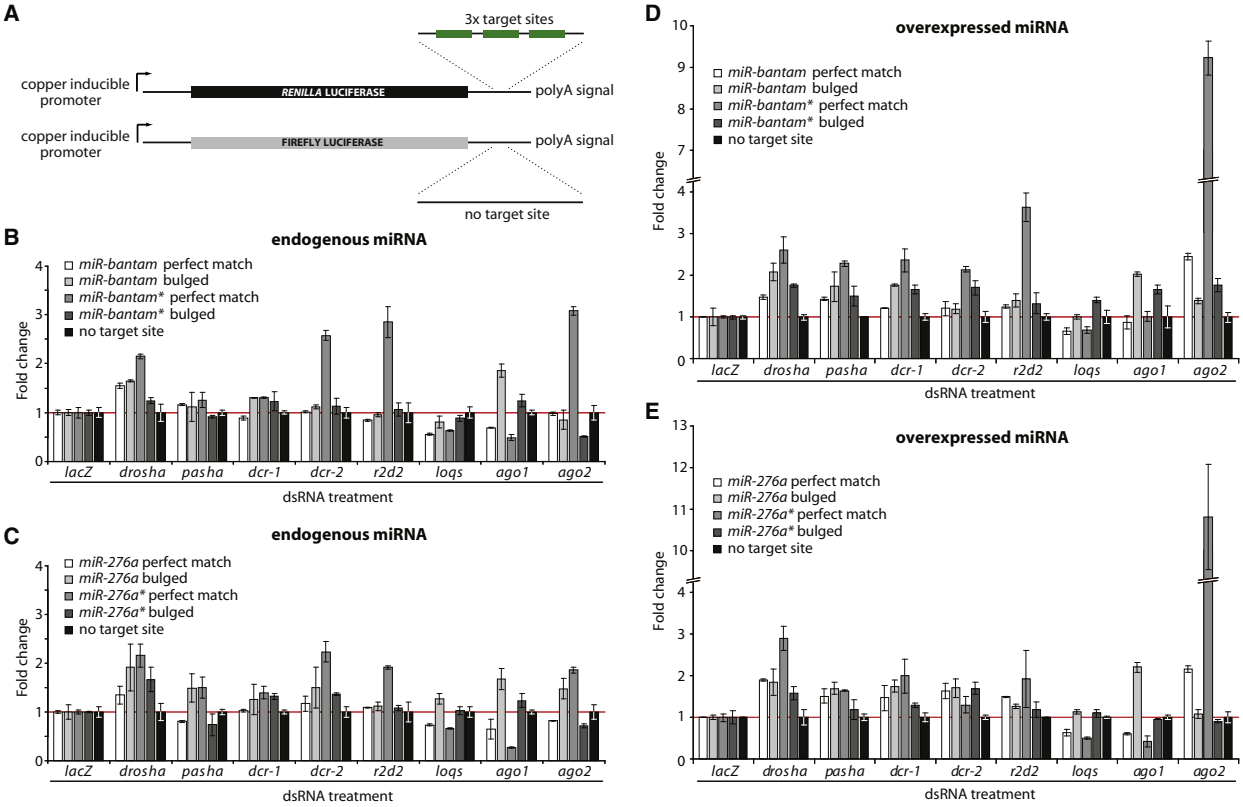


Figure 5. Silencing by miR and miR* Strands in S2 Cells

(A) Schematic diagram showing the configuration of the sensor constructs. Three perfect match or bulged target sites for the miR or miR* strands of *miR-bantam* and *miR-276a* were placed in the 3′ UTR of the *Renilla* luciferase gene. A firefly luciferase construct without target sites served as a normalization control.

(B) The indicated *Renilla* luciferase sensor constructs for *miR-bantam* or a control *Renilla* luciferase construct without target sites was cotransfected into S2 cells with a firefly luciferase construct. Cells were treated with dsRNAs targeting indicated RNAi pathway components. Fold changes in reporter activity were calculated as *Renilla*/firefly ratio normalized first against the control sample (cells treated with dsRNA targeting *lacZ*), then against cells transfected with the control construct without target sites. Shown is the average reporter activity (error bars indicate SD; n = 2).

(C) Sensor activities for *miR-276a* as described in (B).

(D) Sensor activities for overexpressed *miR-bantam*. Experiments were performed as described in (B), but in addition, an expression construct for *miR-bantam* was cotransfected with the reporter constructs.

(E) Sensor activities for overexpressed *miR-276a* as described in (D).

the sensors for the miR* strand of either *miR-bantam* or *miR-276a* in perfect configuration in *ago2* clones, possibly due to residual AGO2 protein in mutant clones (Figure S6). In fact, a sensor transgene for *esi-2.1*, a highly abundant endo-siRNA shown to be loaded to AGO2, was only mildly derepressed in *ago2* clones (Figure S10). Neither were obvious phenotypes observed in *loqs* clones (Figure S9). We did observe a moderate derepression of a perfect match sensor for the *miR-bantam** strand in *dcr-2* or *r2d2* clones (Figures S7G and S8G), consistent with their derepression following similar treatment of S2 cells (Figures 5B–5E).

Thermodynamic Properties of Endo-siRNAs and Strand Selection

Our data indicated that central bulges are the major determinant of sorting and strand selection in mismatch-containing duplexes. For these species, the thermodynamic properties of duplex ends impact sorting and strand selection to only a minor degree. To test the contribution of thermodynamic asymmetry for sorting and loading from perfect duplexes, we analyzed the energies of endo-siRNAs from the *klarsicht* locus and of viral siRNAs. These were almost absent from AGO1 immunoprecipitates but were loaded into AGO2 (Figure 2A). Only sequences where both the guide and passenger strands were cloned in libraries from AGO2 immunoprecipitates were considered for our analysis. We split siRNA duplexes into those showing strong asymmetry (strand bias of guide to passenger of 20:1 or higher) and weak asymmetry (strand bias of 5:1 or lower). We calculated the average thermodynamic energies of both ends, considering up to six terminal nucleotides. The average energies of guide-strand ends were divided by the average energies of passenger-strand ends, and the results were plotted (Figure 7A). Endo-siRNAs derived from the *klarsicht* locus that show stronger asymmetry (as indicated by the ratio of 20:1 or

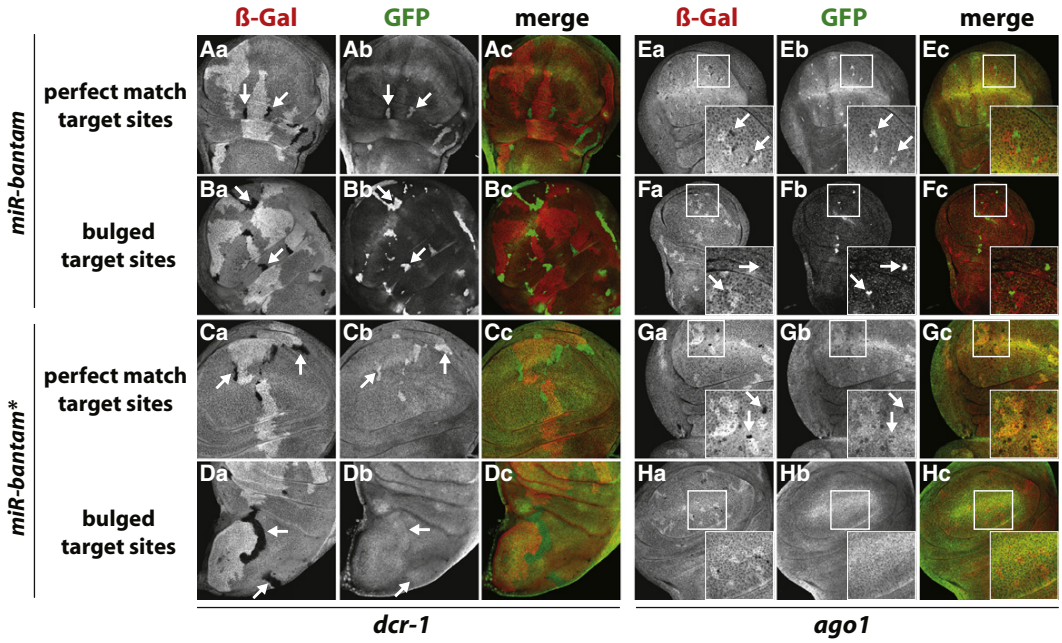


Figure 6. Silencing by miR and miR* Strands in Flies

(A–D) Shown are sensors for *miR-bantam* or *miR-bantam** containing perfectly matched or bulged target sites (as indicated to the left). Negative β-Gal staining (red channel in the merged images) indicates *dcr-1* mutant clones (also marked with arrows). Cells with strong β-Gal staining contain two wild-type *dcr-1* genes, while cells with intermediate staining are heterozygous for *dcr-1*. EGFP sensor activity is shown in green. The black and white panels indicate the separate channels for β-Gal and EGFP.

(E–H) Clonal analysis for *ago1*: details as in (A–D). Selected regions (enclosed in white boxes) were enlarged and shown as insets within each panel to display the smaller *ago1* clones.

higher) also show prominent differences in the end energy between guide and passenger strands for up to four terminal nucleotides. In contrast, *klarsicht* endo-siRNAs with low asymmetry (ratio of 5:1 or lower) show little if any energy differences between their ends. Similar results were obtained for siRNAs derived from viruses, although the magnitude of the overall energy differences was lower (Figure 7A).

DISCUSSION

miRNAs have been honed by evolution to selectively load one strand, the miR strand, into RISC and thus specifically regulate a set of targets that contain complementarity to its specific seed (Bushati and Cohen, 2007; Eulalio et al., 2008). The data presented herein suggest that miRNA precursors can be bifunctional, with individual strands adopting different fates within small RNA pathways. We find that miR* strands are not mere byproducts of miRNA biogenesis but can instead be loaded into demonstrably functional AGO complexes. Notably, this occurs despite miR and miR* strands being produced by precisely the same biogenesis mechanism involving Drosha/Pasha and Dcr-1/Loqs-PB complexes (Figure 4A). Current models incorporate coupled small RNA biogenesis and AGO loading in which Dicer-AGO interactions capture the energy of phosphodiester bond hydrolysis to facilitate incorporation of the small RNA into RISC. Results presented here seem at odds with this model unless Dcr-1 interacts simultaneously with AGO1 and AGO2 to drive the individual strands of a single duplex into separate RISCs. However, this seems unlikely, because depletion of either AGO tends to enrich rather than simultaneously deplete those RNAs present within the other complex. miR* strands persist but lose their terminal 2′-*O*-methylation in the absence of Dcr-2/R2D2, and the ratio of miR*/miR of AGO2-bound small RNA species significantly decreases under these conditions, indicating that this complex is required not for biogenesis but instead for successful and proper miR* loading into AGO2. Thus, we instead favor a model in which the miR:miR* duplex is released from Dcr-1 and subsequently recognized by Dcr-2/R2D2, which shepherds loading into AGO2 (Figure 7B). This release and rebinding has previously been proposed for strand selection within the siRNA pathway (Preall et al., 2006; Tomari and Zamore, 2005). Whether the proximate Dcr-1 product is ever released en route to miR strand loading into AGO1 remains an open question. In one scenario, loading of the miR strand could remain coupled to Dcr-1 cleavage, with those duplexes destined to produce miR*/AGO2 RISC being produced and released by Dcr-1 enzymes that had not formed a complex with AGO1 prior to pre-miRNA cleavage. However, even Dcr-1 complexes must somehow coordinate loading of miR strands, which lie on either the 5p or 3p arm of the precursor, perhaps suggesting that the AGO1 loading machinery might also rely on Dcr-1 product release prior to loading so that both strands can

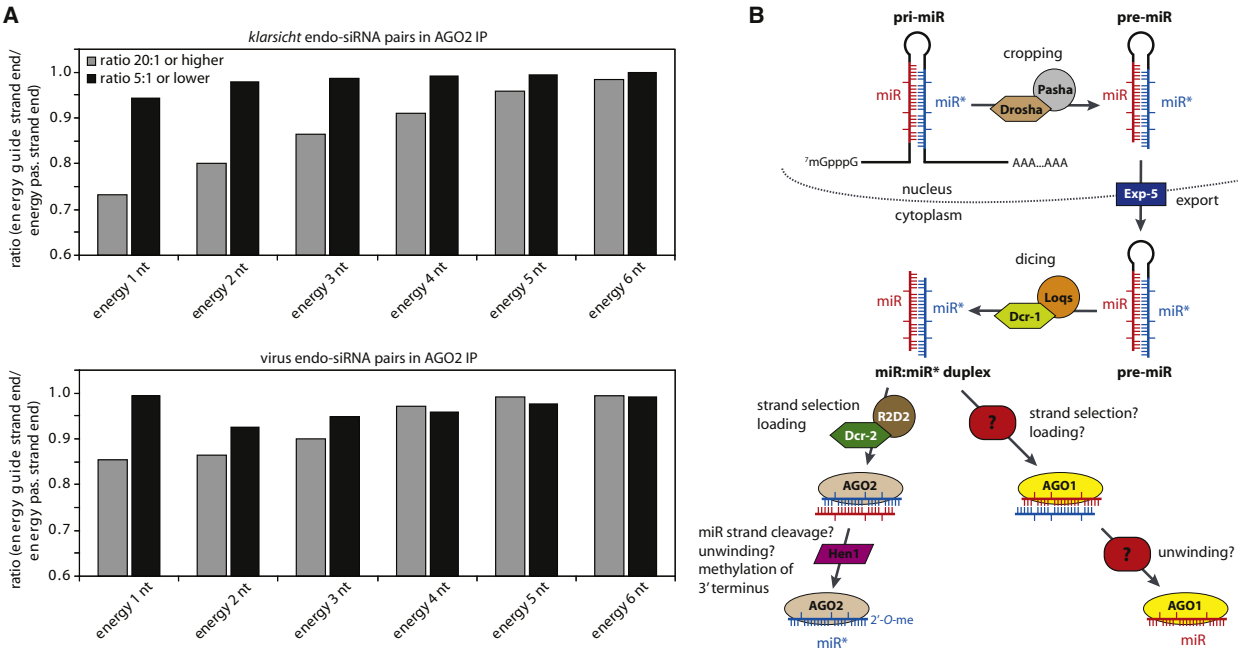


Figure 7. A Hierarchy of Rules for Small RNA Loading in Flies

(A) Thermodynamic properties of AGO2-associated endo-siRNAs matching the *klarsicht* locus (upper chart) and viral siRNAs (lower chart). All siRNA duplexes with both strands cloned were extracted bioinformatically, and ratios of cloning abundances between guide and passenger strands were calculated. Average energies for up to six terminal nucleotides were plotted for strongly asymmetric (strand bias of 20:1 or higher) and weakly asymmetric duplexes (strand bias of 5:1 or lower).

(B) Model for differential sorting of miRNA duplexes in flies.

be interrogated. This is further supported by the observation that the endo-siRNA *esi-2.3*, a Dcr-2 product, is preferentially loaded into AGO1 (Figure 2B).

In this regard, several lines of evidence suggest that the availability of AGO proteins influences the fate of the miR and miR* strands. The absence of AGO1 clearly impacts the abundance of miR* strands relative to other small RNAs, e.g., endo-siRNAs, that join AGO2 complexes. However, the strongest indications for coupling between AGOs and the fates of miR and miR* come from functional analysis of sensors in cell culture and in animals. A comparison of tissues containing 0, 1, or 2 copies of the *ago1* gene shows a graded ability to repress sensors for the miR* strands of *miR-bantam* or *miR-276a*. As compared to heterozygous cells, homozygous *ago1* clones hyper-repress miR* sensors, while cells with 2 copies of intact *ago1* show reduced repression as compared to heterozygous cells. Thus, either a true coupling remains between the biogenesis machinery and AGO proteins that determines the fate of small RNA duplexes, or the relative levels of proteins that will accept miR or miR* strands simply influence the availability of substrates for loading along each pathway.

Results presented herein incorporate several previously proposed rules for small RNA sorting in the *Drosophila* RNAi pathway, but refine some and place these within an overall hierarchy for selection of both the loaded strand and the destination AGO protein. For imperfect small RNA duplexes, the principal determinant seems to be the detection of paired or unpaired residues around the ninth position of the interrogated strand. Each strand of a precursor duplex seems to be assessed individually, since a single miRNA precursor can funnel one strand into AGO1 with the other independently flowing into AGO2. This is not specific to small RNAs generated by Dcr-1, since endo-siRNAs, which are Dcr-2 products, also follow this rule and can, based upon the pattern of interior bulges, select a particular strand for loading into AGO1. Analyses of natural miRNAs and of experimentally altered precursor duplexes indicate that this strand selection rule dominates thermodynamic asymmetry. For example, a number of miR* strands join AGO2 despite having a substantially more stable 5′ end than the miR strand. Previously proposed thermodynamic asymmetry rules (Khvorova et al., 2003; Schwarz et al., 2003) become dominant for perfectly paired small RNA duplexes, such as those arising from the *klarsicht* locus and from viruses. Thus, our studies not only begin to hierarchically integrate rules for small RNA selection in the RNAi pathway but also suggest that the pathways leading to the generation of miR-loaded AGO1 RISC and siRNA-loaded AGO2 RISC are perhaps not as separate as generally supposed.

EXPERIMENTAL PROCEDURES

Cell Culture, Transfection, and RNAi

S2-NP cells were maintained, transfected, and selected as previously described (see Supplemental Experimental Procedures).

DNA Constructs

DNA fragments (~500 bp) encompassing *miR-bantam* and *miR-276a* were amplified by PCR and cloned into pRmHa-3. Pairs of oligonucleotides containing three perfect or bulged target sites for *miR-bantam*, *miR-bantam**, *miR-276a*, or *miR-276a** were annealed and cloned into pRmHa-3-*Renilla* or pJB8 (*tubulin-EGFP* in pCaSpeR4) to generate sensor constructs. A pair of oligonucleotides containing two perfect sites for *esi-2.1* was annealed and cloned into pJB8 to generate an *esi-2.1* sensor. All these sensor constructs were used to generate transgenic flies using standard *P*-element-mediated transformation. See Table S1 for oligonucleotide sequences.

β-Elimination

The chemical structure of 3′ termini of small RNAs was analyzed as described (Vagin et al., 2006) (see Supplemental Experimental Procedures).

Immunoprecipitation

Cell extracts were prepared, evenly split and subjected to immunoprecipitation using antibodies against AGO1 (Abcam; Cambridge, MA) or the FLAG epitope (Sigma; St. Louis), respectively, as described (Czech et al., 2008; Zhou et al., 2008). RNAs were recovered from the immunoprecipitated samples using TRIzol (Invitrogen; Carlsbad, CA) and used for production of small RNA libraries or northern blotting.

Northern Blotting

Northern blotting was carried out as described (Czech et al., 2008; Zhou et al., 2009) (see Supplemental Experimental Procedures).

Small RNA Libraries

Small RNAs were cloned as described (Brennecke et al., 2007). A detailed description of small RNA libraries prepared or used in this study can be found in the Supplemental Experimental Procedures.

Bioinformatic Analysis of Small RNA Libraries

The analysis of small RNA libraries was performed similarly as described (Czech et al., 2008) (see Supplemental Experimental Procedures).

Fly Strains

Fly strains were maintained in standard media. All generated and used strains are listed in Table S2.

Clonal Analysis

Clonal analysis was performed as described (Brennecke et al., 2005). Briefly, developing larva were heat-shocked at 37°C for 1 hr at 50–60 hr of development for flies carrying mutations for *dcr-1*, *dcr-2*, *ago2*, *r2d2*, or *loqs*, except for *ago1* flies, which were heat-shocked at 96–108 hr of development. Wandering third-instar larva were dissected, and the imaginal wing discs were fixed in 4% formaldehyde-PBS at room temperature for 30 min and stained with monoclonal anti-β-Gal antibody (1:500; Promega; Madison, WI), rabbit anti-GFP antibody (1:1000; Molecular Probes; Carlsbad, CA), and secondary antibodies (Alexa 488-conjugated goat anti-rabbit and Alexa 594-conjugated goat anti-mouse; 1:500; Molecular Probes). A rat anti-HA antibody (1:1000; Roche; Indianapolis, IN) was employed to examine the expression pattern of FLAG/HA-AGO2 in the imaginal wing disc.

Argonaute Loading Assay

Cells expressing FLAG/HA-AGO2 (see above) were transfected with various siRNA or miRNA duplexes (Table S1) using HiPerFect (QIAGEN; Valencia, CA). Two days after transfection, cell lysates were prepared and evenly split, and each half was subjected to immunoprecipitation using antibodies against AGO1 and the FLAG tag, respectively (see above). RNAs were recovered from the immunoprecipitates and subjected to sequential northern blotting using a mixture of probes complementary to the top strands or to the bottom strands of the *miR-276a* or *let-7* series of duplexes and those against *miR-bantam* and the guide strand of *esi-2.1*. The intensity of the signals was quantified and normalized to those of *esi-2.1* and *miR-bantam* for AGO2 and AGO1 loading, respectively. The corresponding Argonaute loading index for each sample was calculated using the following equation. For example, the AGO1 loading index for the top strand of *miR-276a* duplex 1 is calculated as: [(*miR-276a* duplex #1 top strand$^{\textit{miR-276a}\ \text{duplex\ \#1\ tfxn\ AGO1\ IP}}$ − gel background) / (*miR-bantam*$^{\textit{miR-276a}\ \text{duplex\ \#1\ tfxn\ AGO1\ IP}}$ − gel background)] − [(*miR-276a* duplex #1 top strand$^{\text{nontransfection control AGO1 IP}}$ − gel background) / (*miR-bantam*$^{\text{nontransfection control AGO1 IP}}$ − gel background)]. To calculate the relative Argonaute loading index, the AGO1$^{\text{index}}$/AGO2$^{\text{index}}$ ratio for each strand of the duplex was determined. Finally, the relative Argonaute index for each strand was normalized to that of the corresponding strand of duplex 1, and the results were $\log_{(2)}$ transformed and plotted.

Thermodynamics Calculations

All 21 nt long reads within the wild-type AGO2 IP library matching to the *klarsicht* locus or viral genomes were extracted bioinformatically (Czech et al., 2008). Only those sequences corresponding to pairs of guide and passenger strands resembling perfect match duplexes with 2 nt overhangs at the 3′ termini were subjected to further analysis. The terminal energies of up to six nucleotides counted from both ends of those duplexes were calculated individually using UNAfold (Markham and Zuker, 2008). Sequences matching to both categories were next grouped into strong asymmetric duplexes (cloning count ratio of guide to passenger of 20:1 or higher) and weak asymmetric duplexes (strand bias of 5:1 or less). Average energies were computed for both groups, and energies of guide-strand ends were divided by energies for passenger-strand ends. To correlate the energies with the degree of asymmetry, the median results were plotted for all six nucleotides individually.

ACCESSION NUMBERS

Small RNA sequences generated in this study can be obtained at GEO using accession number GSE17734.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, two tables, and ten figures and can be found online at http://www.cell.com/molecular-cell/supplemental/S1097-2765(09)00688-1.

ACKNOWLEDGMENTS

We thank R. Carthew, Q. Liu, P. Jin, H. Siomi, P.D. Zamore, E. Lai, and the Bloomington Stock Center for fly strains. We are grateful to M. Kudla, O. Tam, M. Rooks, D. McCombie, and C. Pitsouli for technical and computational contributions. We thank R. Davis and J. Dover for kindly sequencing two Solexa libraries at University of Colorado, Denver. B.C. is supported by a PhD fellowship from the Boehringer Ingelheim Fonds. R.Z. is a Special Fellow of the Leukemia and Lymphoma Society. This work was supported in part by grants from the NIH to N.P. and G.J.H. and a gift from K.W. Davis (G.J.H.). N.P. and G.J.H. are investigators of the HHMI.

Received: July 21, 2009
Revised: August 20, 2009
Accepted: August 28, 2009
Published: November 12, 2009

REFERENCES

Aliyari, R., Wu, Q., Li, H.W., Wang, X.H., Li, F., Green, L.D., Han, C.S., Li, W.X., and Ding, S.W. (2008). Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in Drosophila. Cell Host Microbe *4*, 387–397.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol. *3*, e85.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089–1103.

Bushati, N., and Cohen, S.M. (2007). microRNA functions. Annu. Rev. Cell Dev. Biol. *23*, 175–205.

Chung, W.J., Okamura, K., Martin, R., and Lai, E.C. (2008). Endogenous RNA interference provides a somatic defense against Drosophila transposons. Curr. Biol. *18*, 795–802.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., et al. (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798–802.

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.

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

Flynt, A., Liu, N., Martin, R., and Lai, E.C. (2009). Dicing of viral replication intermediates during silencing of latent Drosophila viruses. Proc. Natl. Acad. Sci. USA *106*, 5270–5275.

Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol. *3*, e236.

Förstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287–297.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., and Zamore, P.D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077–1081.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235–240.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293–296.

Hartig, J.V., Esslinger, S., Böttcher, R., Saito, K., and Förstemann, K. (2009). Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. EMBO J. *28*, 2932–2944.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr. Biol. *17*, 1265–1272.

Iwasaki, S., Kawamata, T., and Tomari, Y. (2009). Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression. Mol. Cell *34*, 58–67.

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 Dev. *19*, 1674–1679.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature *453*, 793–797.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209–216.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. EMBO J. *23*, 4051–4060.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004b). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69–81.

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921–1925.

Markham, N.R., and Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. Methods Mol. Biol. *453*, 3–31.

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.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev. *19*, 2837–2848.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008a). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803–806.

Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T., and Lai, E.C. (2008b). The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. Nat. Struct. Mol. Biol. *15*, 354–363.

Park, J.K., Liu, X., Strauss, T.J., McKearin, D.M., and Liu, Q. (2007). The miRNA pathway intrinsically controls self-renewal of Drosophila germline stem cells. Curr. Biol. *17*, 533–538.

Preall, J.B., He, Z., Gorra, J.M., and Sontheimer, E.J. (2006). Short interfering RNA strand selection is independent of dsRNA processing polarity during RNAi in Drosophila. Curr. Biol. *16*, 530–535.

Saito, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2005). Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. PLoS Biol. *3*, e235.

Saito, K., Sakaguchi, Y., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2′-O-methylation of Piwi- interacting RNAs at their 3′ ends. Genes Dev. *21*, 1603–1608.

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199–208.

Seitz, H., Ghildiyal, M., and Zamore, P.D. (2008). Argonaute loading improves the 5′ precision of both MicroRNAs and their miRNA strands in flies. Curr. Biol. *18*, 147–151.

Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. Genes Dev. *19*, 517–529.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004). A protein sensor for siRNA asymmetry. Science *306*, 1377–1380.

Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell *130*, 299–308.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320–324.

Zhou, R., Hotta, I., Denli, A.M., Hong, P., Perrimon, N., and Hannon, G.J. (2008). Comparative analysis of argonaute-dependent small RNA pathways in Drosophila. Mol. Cell *32*, 592–599.

Zhou, R., Czech, B., Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., and Hannon, G.J. (2009). Processing of Drosophila endo-siRNAs depends on a specific Loquacious isoform. RNA *15*, 1886–1895.

Supplemental Data

Molecular Cell, *Volume 36*

Hierarchical Rules for Argonaute Loading in *Drosophila*

Benjamin Czech, Rui Zhou, Yaniv Erlich, Julius Brennecke, Richard Binari, Christians Villalta, Assaf Gordon, Norbert Perrimon, and Gregory J. Hannon

Supplemental Experimental Procedures

Cell culture, transfection, and RNAi

S2-NP cells were maintained in Schneider's medium (Invitrogen, Carlsbad, CA) supplemented with 10% FBS and 1% Pen-Strep (Invitrogen, Carlsbad, CA). Cells were transfected with an expression construct for FLAG/HA-tagged AGO2 along with the selection marker plasmid pMK33-NTAP using Effectene (Qiagen, Valencia, CA). To generate a stable cell line, selection was carried out with medium containing 150 µg/mL Hygromycin B (Calbiochem, La Jolla, CA). For dsRNA treatments, ~3X10^6 cells were soaked in 1.5 mL serum-free Schneider's medium containing 10 µg dsRNAs in 6-well plates, and 3 mL serum-containing medium was added 45 minutes later. After 4 days of initial dsRNA treatment, cells were treated with dsRNAs for a second round and harvested another 4 days later as described before (Czech et al., 2008; Zhou et al., 2008). For *Renilla* luciferase reporter assays, transfection was performed in 384-well format using Effectene (Qiagen, Valencia, CA) as described before (Czech et al., 2008).

β-elimination

The chemical structure of 3' termini of small RNAs was analyzed as described (Vagin et al., 2006). In brief, RNAs from immunoprecipitates or 25 µg of total RNA from S2 cells treated with indicated dsRNAs (17 µL total volume for each sample) were incubated at room temperature for 30 min with 5 µL 5x borate buffer (148 mM borax,148 mM boric acid, pH 8.6) supplemented with 3 µL freshly prepared 200 mM $NaIO_4$. 5 µL 50% glycerol was added to quench non-reacted

sodium periodate by incubating for additional 15 min at room temperature. Samples were subsequently vacuum dried and dissolved in 60 µL 1x borax buffer (30 mM borax, 30mM boric acid, 50 mM NaOH, pH 9.5). β -elimination was carried out by incubation for 2 hours at 45°C. RNAs were Ethanol-precipitated and resolved in 1x gel loading buffer.

Northern Blotting

Northern blotting was carried out as described (Czech et al., 2008; Zhou et al., 2009). In brief, total RNAs from knockdown cells were isolated using Trizol (Invitrogen, Carlsbad, CA). 30 µg total RNAs from cultured cells (with or without β-elimination) or RNAs from immunoprecipitations were separated on 15% denaturing poly acrylamide gels and transferred to Hybond-N+ membranes (Amersham Biosciences) in 1x TBE buffer. Small RNAs were UV cross-linked to the membrane and pre-hybridized in ULTRAhyb^{TM}-Oligo buffer (Ambion, Austin, TX) for one hour. DNA probes (sequences are shown in Table S1) complementary to the indicated strands were 5' radio-labeled and added to the hybridization buffer (hybridization for 6 hours at 30°C). Membranes were washed 4 times in 1x SSC with 0.1% SDS at 30°C and exposed to PhosphorImager screens for 12-48 hours. Membranes were stripped by heating in 0.2x SSC containing 0.1% SDS in a microwave twice.

Small RNA libraries

Small RNAs were cloned as described (Brennecke et al., 2007). For this study, the following small RNA libraries from total RNAs were prepared:

19- to 24-nt from wild-type S2 cells subjected to a modified cloning strategy ("oxidized") described in (Seitz et al., 2008),

19- to 24-nt from S2 cells treated with dsRNA against *ago1*,

19- to 24-nt from S2 cells treated with dsRNA against *ago2*,

AGO1 immunoprecipitates from S2 cells treated with dsRNA against *dcr-2*, and

AGO2 immunoprecipitates from S2 cells treated with dsRNA against *dcr-2*.

Libraries were sequenced in-house or at the University of Colorado, Denver (courtesy of J. Dover and R. E. Davis) using the Illumina GAII sequencing platform. Small RNA sequences were deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE17734. In addition, we used the following published small RNA libraries for our analyses:

19- to 24-nt from wild-type S2 cells ("standard") (GSE17171) (Zhou et al., 2009),

AGO1 immunoprecipitates from S2 cells (GSE11086) (Czech et al., 2008), and

AGO2 immunoprecipitates from S2 cells (GSE11086) (Czech et al., 2008).

Bioinformatic analysis of small RNA libraries

The analysis of small RNA libraries was performed similar as described (Czech et al., 2008). Illumina reads were stripped of the 3' linker, collapsed, and the resulting small RNA sequences were matched without mismatches to the *Drosophila* release 5 genome, and to the genomes of *Drosophila* C virus, Flock house virus and Cricket paralysis virus with up to 3 mismatches. Only reads that met these conditions were subjected to further analyses. For annotations we used a combination of UCSC, miRBase, and Flybase tracks for protein coding genes, repeats/transposons, non-coding RNAs and microRNAs, as well as custom tracks (for synthetic markers, endo-siRNAs from structured loci, miR and miR* strands) with different priorities (annotation priority list available upon request). For comparison of small RNA counts between libraries, reads were normalized to the same total number after bioinformatic removal of those sequences that matched to synthetic cloning markers or assumed degradation products of abundant cellular RNAs (rRNAs, snoRNAs and tRNAs). Heatmaps were created by plotting the abundance of individual sequences within each library and by plotting the calculated relative association between the two analyzed libraries.

Supplemental Figures and Tables

Table S1. Sequences of DNA/RNA oligonucleotides

siRNA/miRNA sensor oligonucleotides	*sequence (5'->3')*
miR-276a_miR_perfect_for	TCGAAGAGCACGGTATGAAGTTCCTAAATGATATGAAGAGCACGGTATGAAGTTCCTAAATGAACTGAAGAGCACGGTATGAAGTTCCTA
miR-276a_miR_perfect_rev	TCGATAGGAACTTCATACCGTGCTCTTCAGTTCATTTAGGAACTTCATACCGTGCTCTTCATATCATTTAGGAACTTCATACCGTGCTCT
miR-276a_miR_bulged_for	TCGAAGAGCACGGTGTAAGTTCCTAAATGATATGAAGAGCACGGTGTAAGTTCCTAAATGAACTGAAGAGCACGGTGTAAGTTCCTA
miR-276a_miR_bulged_rev	TCGATAGGAACTTACACCGTGCTCTTCAGTTCATTTAGGAACTTACACCGTGCTCTTCATATCATTTAGGAACTTACACCGTGCTCT
miR-276a_star_perfect_for	TCGACGTAGGAACTCTATACCTCGCTAATGATATGACGTAGGAACTCTATACCTCGCTAATGAACTGACGTAGGAACTCTATACCTCGCT
miR-276a_star_perfect_rev	TCGAAGCGAGGTATAGAGTTCCTACGTCAGTTCATTAGCGAGGTATAGAGTTCCTACGTCATATCATTAGCGAGGTATAGAGTTCCTACG
miR-276a_star_bulged_for	TCGACGTAGGAACTACTACCTCGCTAATGATATGACGTAGGAACTACTACCTCGCTAATGAACTGACGTAGGAACTACTACCTCGCT
miR-276a_star_bulged_rev	TCGAAGCGAGGTAGTAGTTCCTACGTCAGTTCATTAGCGAGGTAGTAGTTCCTACGTCATATCATTAGCGAGGTAGTAGTTCCTACG
miR-bantam_miR_perfect_for	TCGAATCAGCTTTCAAAATGATCTCAAATGATATGAATCAGCTTTCAAAATGATCTCAAATGAACTGAATCAGCTTTCAAAATGATCTCA
miR-bantam_miR_perfect_rev	TCGATGAGATCATTTTGAAAGCTGATTCAGTTCATTTGAGATCATTTTGAAAGCTGATTCATATCATTTGAGATCATTTTGAAAGCTGAT
miR-bantam_miR_bulged_for	TCGAATCAGCTTTCCTATGATCTCAAATGATATGAATCAGCTTTCCTATGATCTCAAATGAACTGAATCAGCTTTCCTATGATCTCA
miR-bantam_miR_bulged_rev	TCGATGAGATCATAGGAAAGCTGATTCAGTTCATTTGAGATCATAGGAAAGCTGATTCATATCATTTGAGATCATAGGAAAGCTGAT
miR-bantam_star_perfect_for	TCGAGTCAAACCAAATCGAAAACCGGAATGATATGAGTCAAACCAAATCGAAAACCGGAATGAACTGAGTCAAACCAAATCGAAAACCGG
miR-bantam_star_perfect_rev	TCGACCGGTTTTCGATTTGGTTTGACTCAGTTCATTCCGGTTTTCGATTTGGTTTGACTCATATCATTCCGGTTTTCGATTTGGTTTGAC
miR-bantam_star_bulged_for	TCGAGTCAAACCAAGAGAAAACCGGAATGATATGAGTCAAACCAAGAGAAAACCGGAATGAACTGAGTCAAACCAAGAGAAAACCGG
miR-bantam_star_bulged_rev	TCGACCGGTTTTCTCTTGGTTTGACTCAGTTCATTCCGGTTTTCTCTTGGTTTGACTCATATCATTCCGGTTTTCTCTTGGTTTGAC
Xba-si2-S2-2P-F	CTAGACAACAGTTTATTGGAGCGAACTTGTTGGAGTCAAAATGAACTGAGGGTGGAGCGAACTTGTTGGAGTCAAC
Xho-si2-S2-2P-R	TCGAGTTGACTCCAACAAGTTCGCTCCACCCTCAGTTCATTTTGACTCCAACAAGTTCGCTCCAATAAACTGTTGT

Oligonucleotides for pri-miRNA cloning	*sequence (5'->3')*
EcoR-Bantam-F	GGAATTCTTTTGAGTCTCTCGTTC
Bam-Bamtam-R	GGGGATCCTCAAAACCTTTGAACCTTTTAAATCC
EcoR-miR-276a-F	GGAATTCTGAAGAGAATCCCATCAGTC
Bam-miR-276a-R	GGGGATCCAAATATCATCGCATTTCGCTC

Probes for Northern blotting	*sequence (5'->3')*	*Notes*
probe-miR276a-WT-F	CGTAGGAACTCTATACCTCGCT	probe *miR-276a-1* top strand
probe-miR276a-WT-R	AGAGCACGGTATGAAGTTCCTA	probe *miR-276a-1* bottom strand
probe-miR276a-guide-5A-F	CGAAGGAACTCTATACCTCGCT	probe *miR-276a-2* top strand
probe-miR276a-guide-5A-R	AGAGCACGGTATGAAGTTCCTT	probe *miR-276a-2* bottom strand
probe-miR276a-star-5U-F	CGTAGGAACTCTATACCTCGCA	probe *miR-276a-3* top strand
probe-miR276a-star-5U-R	AGTGCACGGTATGAAGTTCCTA	probe *miR-276a-3* bottom strand
probe-miR276a-guide-5A-star-5U-F	CGAAGGAACTCTATACCTCGCA	probe *miR-276a-4* top strand
probe-miR276a-guide-5A-star-5U-R	AGTGCACGGTATGAAGTTCCTT	probe *miR-276a-4* bottom strand
probe-miR276a-central-seed-guide-5A-star-5U-F	CGAAGAAACTCTATACCTCGCA	probe *miR-276a-5* top strand
probe-miR276a-central-seed-guide-5A-star-5U-R	AGTGCGAGGTTAAGAGTACCTT	probe *miR-276a-5* bottom strand
probe-miR276a-central-guide-5A-star-5U-F	CGAAGGAACTCTATACCTCGCA	probe *miR-276a-6* top strand
probe-miR276a-central-guide-5A-star-5U-R	AGTGCGAGGTTAAGAGTTCCTT	probe *miR-276a-6* bottom strand
probe-miR276a-seed fixed-guide-5A-star-5U-F	CGAAGGAACTCTATACCTCGCA	probe *miR-276a-7* top strand
probe-miR276a-seed fixed-guide-5A-star-5U-R	AGTGCGAGGTATGTAGTTCCTT	probe *miR-276a-7* bottom strand
probe-miR276a-central-F	CGTAGGAACTCTATACCTCGCT	probe *miR-276a-8* top strand
probe-miR276a-central-R	AGAGCGAGGTTAAGAGTTCCTA	probe *miR-276a-8* bottom strand
probe-miR276a-seed-F	CGTAGAAACTCTATACCTCGCT	probe *miR-276a-9* top strand
probe-miR276a-seed-R	AGAGCGAGGTATAGAGTACCTA	probe *miR-276a-9* bottom strand
probe-miR276a-central-seed-F	CGTAGAAACTCTATACCTCGCT	probe *miR-276a-10* top strand
probe-miR276a-central-seed-R	AGAGCGAGGTTAAGAGTACCTA	probe *miR-276a-10* bottom strand
probe-miR276a-seed only-F	CGTAGGAACTCTATACCTCGCT	probe *miR-276a-11* top strand
probe-miR276a-seed only-R	AGAGCACGGTATAGAGTTCCTA	probe *miR-276a-11* bottom strand
probe-mod(let7)_same ends-F	AATTTACAACCCGGCTACCAAA	probe *let-7-1* top strand
probe-mod(let7)_same ends-R	AATTTGGTAGCCGGGTTGTAAA	probe *let-7-1* bottom strand
probe-mod(let7)_same ends+guide bulge-F	AATTTACAACCCTGCTACCAAA	probe *let-7-2* top strand
probe-mod(let7)_same ends+guide bulge-R	AATTTGGTAGTCGGGTTGTAAA	probe *let-7-2* bottom strand
probe-mod(let7)_same ends+pas bulge-F	AATTTACAACTCGGCTACCAAA	probe *let-7-3* top strand
probe-mod(let7)_same ends+pas bulge-R	AATTTGGTAGCCTGGTTGTAAA	probe *let-7-3* bottom strand
probe-mod(let7)_unstable guide-F	AATTTACAACCCGGCTACCATA	probe *let-7-4* top strand
probe-mod(let7)_unstable guide-R	AAATTGGTAGCCGGGTTGTAAA	probe *let-7-4* bottom strand
probe-mod(let7)_unstable guide+guide bulge-F	AATTTACAACCCTGCTACCATA	probe *let-7-5* top strand
probe-mod(let7)_unstable guide+guide bulge-R	AAATTGGTAGTCGGGTTGTAAA	probe *let-7-5* bottom strand
probe-mod(let7)_unstable guide+pas bulge-F	AATTTACAACTCGGCTACCATA	probe *let-7-6* top strand
probe-mod(let7)_unstable guide+pas bulge-R	AAATTGGTAGCCTGGTTGTAAA	probe *let-7-6* bottom strand
probe-mod(let7)_unstable pas-F	AAATTACAACCCGGCTACCAAA	probe *let-7-7* top strand
probe-mod(let7)_unstable pas-R	AATTTGGTAGCCGGGTTGTATA	probe *let-7-7* bottom strand
probe-mod(let7)_unstable pas+guide bulge-F	AAATTACAACCCTGCTACCAAA	probe *let-7-8* top strand
probe-mod(let7)_unstable pas+guide bulge-R	AATTTGGTAGTCGGGTTGTATA	probe *let-7-8* bottom strand
probe-mod(let7)_unstable pas+pas bulge-F	AAATTACAACTCGGCTACCAAA	probe *let-7-9* top strand
probe-mod(let7)_unstable pas+pas bulge-R	AATTTGGTAGCCTGGTTGTATA	probe *let-7-9* bottom strand
miR-bantam	AATCAGCTTTCAAAATGATCTCA	
miR-bantam-star	AGTCAAACCAAATCGAAAACCGG	
miR-184	GCCCTTATCAGTTCTCCGTCCA	
miR-184-star	ACGGGGCGAGAGAATGATAAGG	
miR-276a	AGAGCACGGTATGAAGTTCCTA	
miR-276a-star	CGTAGGAACTCTATACCTCGCT	
esi-2.1-quide	GGAGCGAACTTGTTGGAGTCAA	
esi-2.3-guide	TGACTTCCGGCGGTTAAGATTT	
2S rRNA	TACAACCCTCAACCATATGTAGTCCAAGCA	

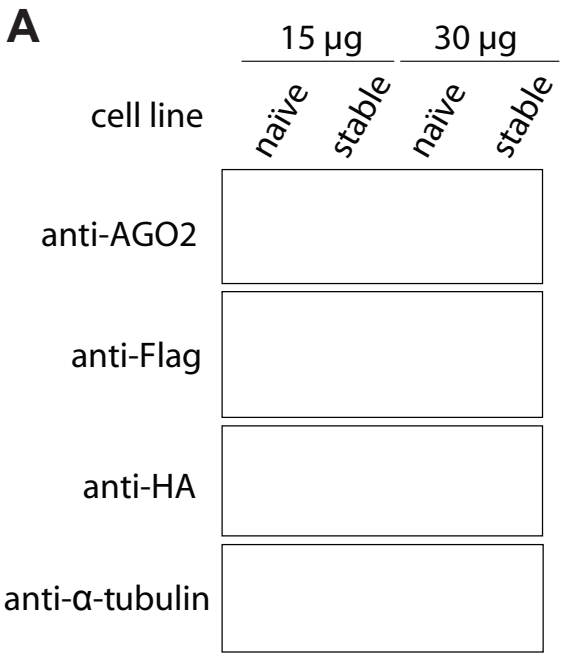
siRNAs for Argonaute loading assay	*sequence (5'->3')*	*Notes*
miR276a-WT-F	agcgaggauaagaguuccuacg	*miR-276a-1* top strand
miR276a-WT-R	uaggaacuucauaccgugcucu	*miR-276a-1* bottom strand
miR276a-guide-5A-F	agcgaggauaagaguuccuUcg	*miR-276a-2* top strand
miR276a-guide-5A-R	Aaggaacuucauaccgugcucu	*miR-276a-2* bottom strand
miR276a-star-5U-F	Ugcgaggauaagaguuccuacg	*miR-276a-3* top strand
miR276a-star-5U-R	uaggaacuucauaccgugcAcu	*miR-276a-3* bottom strand
miR276a-guide-5A-star-5U-F	UgcgaggauaagaguuccuUcg	*miR-276a-4* top strand
miR276a-guide-5A-star-5U-R	AaggaacuucauaccgugcAcu	*miR-276a-4* bottom strand
miR276a-central-seed-guide-5A-star-5U-F	UgcgaggauaagaguUcuUcg	*miR-276a-5* top strand
miR276a-central-seed-guide-5A-star-5U-R	AaggUacuCUUAaccUCgcAcu	*miR-276a-5* bottom strand
miR276a-central-guide-5A-star-5U-F	UgcgaggauaagaguuccuUcg	*miR-276a-6* top strand
miR276a-central-guide-5A-star-5U-R	AaggaacuCUUAaccUCgcAcu	*miR-276a-6* bottom strand
miR276a-seed fixed-guide-5A-star-5U-F	UgcgaggauaagaguuccuUcg	*miR-276a-7* top strand
miR276a-seed fixed-guide-5A-star-5U-R	AaggaacuacauaccucgcAcu	*miR-276a-7* bottom strand
miR276a-central-F	agcgaggauaagaguuccuacg	*miR-276a-8* top strand
miR276a-central-R	uaggaacuCUUAaccUCgcucu	*miR-276a-8* bottom strand
miR276a-seed-F	agcgaggauaagaguuUcuacg	*miR-276a-9* top strand
miR276a-seed-R	uaggUacuCUauaccUCgcucu	*miR-276a-9* bottom strand
miR276a-central-seed-F	agcgaggauaagaguuUcuacg	*miR-276a-10* top strand
miR276a-central-seed-R	uaggUacuCUUAaccUCgcucu	*miR-276a-10* bottom strand
miR276a-seed only-F	agcgaggauaagaguuccuacg	*miR-276a-11* top strand
miR276a-seed only-R	uaggaacuCUauaccgugcucu	*miR-276a-11* bottom strand
mod(let7)_same ends-F	UUUGGUAGCCGGGUUGUAAAUU	*let-7-1* top strand
mod(let7)_same ends-R	UUUACAACCCGGCUACCAAAUU	*let-7-1* bottom strand
mod(let7)_same ends+guide bulge-F	UUUGGUAGCAGGGUUGUAAAUU	*let-7-2* top strand
mod(let7)_same ends+guide bulge-R	UUUACAACCCGACUACCAAAUU	*let-7-2* bottom strand
mod(let7)_same ends+pas bulge-F	UUUGGUAGCCGAGUUGUAAAUU	*let-7-3* top strand
mod(let7)_same ends+pas bulge-R	UUUACAACCAGGCUACCAAAUU	*let-7-3* bottom strand
mod(let7)_unstable guide-F	UAUGGUAGCCGGGUUGUAAAUU	*let-7-4* top strand
mod(let7)_unstable guide-R	UUUACAACCCGGCUACCAAUUU	*let-7-4* bottom strand
mod(let7)_unstable guide+guide bulge-F	UAUGGUAGCAGGGUUGUAAAUU	*let-7-5* top strand
mod(let7)_unstable guide+guide bulge-R	UUUACAACCCGACUACCAAUUU	*let-7-5* bottom strand
mod(let7)_unstable guide+pas bulge-F	UAUGGUAGCCGAGUUGUAAAUU	*let-7-6* top strand
mod(let7)_unstable guide+pas bulge-R	UUUACAACCAGGCUACCAAUUU	*let-7-6* bottom strand
mod(let7)_unstable pas-F	UUUGGUAGCCGGGUUGUAAUUU	*let-7-7* top strand
mod(let7)_unstable pas-R	UAUACAACCCGGCUACCAAAUU	*let-7-7* bottom strand
mod(let7)_unstable pas+guide bulge-F	UUUGGUAGCAGGGUUGUAAUUU	*let-7-8* top strand
mod(let7)_unstable pas+guide bulge-R	UAUACAACCCGACUACCAAAUU	*let-7-8* bottom strand
mod(let7)_unstable pas+pas bulge-F	UUUGGUAGCCGAGUUGUAAUUU	*let-7-9* top strand
mod(let7)_unstable pas+pas bulge-R	UAUACAACCAGGCUACCAAAUU	*let-7-9* bottom strand

Table S2. Fly strains used or produced in this study

genotype	*source*	*parental stock*
yw,hs-flp;arm-LacZ,FRT40A/CyO	This study	Bloomington
yw,hs-flp;arm-LacZ,FRT42D	This study	Bloomington
yw,hs-flp;arm-LacZ,FRT80B	Bloomington	
yw,hs-flp;arm-LacZ,FRT82B/TM6,Tb	This study	Bloomington
yw ey-flp;FRT82B dcr-1^{Q1147X}/TM3,Sb,Ser	(Lee et al., 2004)	
yw ey-flp;FRT42D dcr-2^{L811fsX}	(Lee et al., 2004)	
w;FRT80B,ago2^{414}	This study	(Okamura et al, 2004)
FRT42D,ago1^{EMS1}/CyO	This study	(Yang et al., 2007)
FRT40A,loqs^{KO}/CyO	(Park et al., 2007)	
FRT40A,r2d2^{1}/CyO	This study	(Liu et al., 2003)
tub-EGFP-bantam-guide-perfect	(Brennecke et al., 2003)	
tub-EGFP-bantam-guide-bulge	This study	
tub-EGFP-bantam-star-perfect	This study	
tub-EGFP-bantam-star-bulge	This study	
tub-EGFP-miR276a-guide-perfect	This study	
tub-EGFP-miR276a-guide-bulge	This study	
tub-EGFP-miR276a-star-perfect	This study	
tub-EGFP-miR276a-star-bulge	This study	
w;[FLAG/HA-AGO2]/CyO	(Czech et al., 2008)	

Figure S1. AGO2 expression in naïve and FLAG/HA-AGO2 cells and in the wing imaginal disc.

(A) Cell extracts from naïve S2 cells or FLAG/HA-AGO2 stable cells expressing tagged AGO2 under its endogenous regulatory elements (Czech et al., 2008) were subjected to Western blotting using antibodies against AGO2, the FLAG and HA epitope and α-tubulin as control for equal loading. Note that each panel was from identical sets of samples processed in parallel using the corresponding antibodies. Our stable cell line expresses less total AGO2 than do naïve S2 cells. **(B)** Expression pattern of FLAG/HA-AGO2 in the wing imaginal disc of transgenic flies was examined by immunofluorescence using an anti-HA antibody. FLAG/HA-AGO2 shows uniform, broad cytoplasmic expression.


A
15 µg
30 µg
cell line
naive
stable
naive
stable
anti-AGO2
anti-Flag
anti-HA
anti-α-tubulin

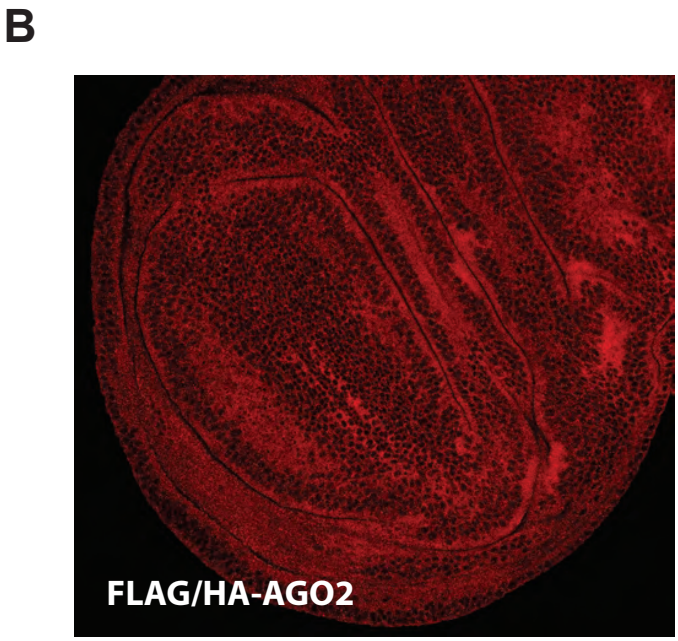

B
FLAG/HA-AGO2

Figure S2. Duplex structures for miRNAs and endo-siRNAs.

Shown are the miRNA and endo-siRNA sequences and their duplex structures that were detected by Northern Blotting in Figure 2B. The miR or guide strand (shown in red) is annealed to its corresponding miR* or passenger strand (shown in blue) with Watson-Crick base pairs indicated by lines and GU wobble base pairs as colons.

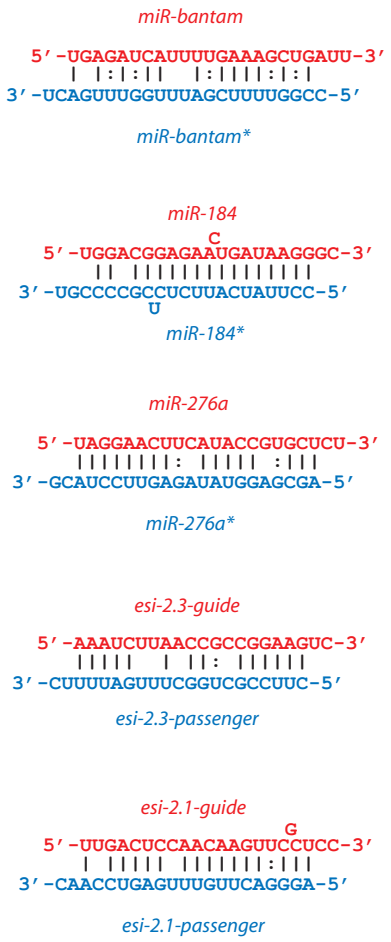

miR-bantam
5′-UGAGAUCAUUUUGAAAGCUGAUU-3′
3′-UCAGUUUGGUUUAGCUUUUGGCC-5′
miR-bantam*
miR-184
C
5′-UGGACGGAGAAUGAUAAGGGC-3′
3′-UGCCCCGCCUCUUACUAUUCC-5′
U
miR-184*
miR-276a
5′-UAGGAACUUCAUACCGUGCUCU-3′
3′-GCAUCCUUGAGAUAUGGAGCGA-5′
miR-276a*
esi-2.3-guide
5′-AAAUCUUAACCGCCGGAAGUC-3′
3′-CUUUUAGUUUCGGUCGCCUUC-5′
esi-2.3-passenger
esi-2.1-guide
G
5′-UUGACUCCAACAAGUUCCUCC-3′
3′-CAACCUGAGUUUGUUCAGGGA-5′
esi-2.1-passenger

Figure S3. AGO loading pattern of individual strands of various modified *let-7* duplexes.

Immunoprecipitation followed by Northern blotting shows the loading of both top (guide) and bottom (passenger) strands for various modified *let-7* duplexes into AGO1 or AGO2. *miR-bantam* and *esi-2.1* served as controls for AGO1 and FLAG-AGO2 immunoprecipitation, respectively.

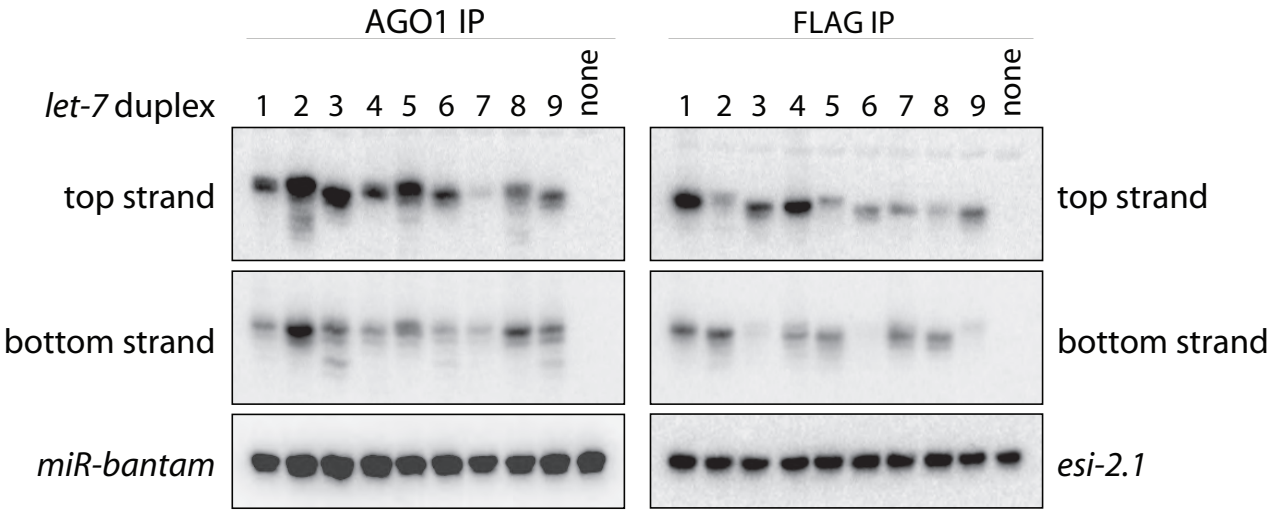

AGO1 IP
FLAG IP
let-7 duplex
1 2 3 4 5 6 7 8 9 none
1 2 3 4 5 6 7 8 9 none
top strand
top strand
bottom strand
bottom strand
miR-bantam
esi-2.1

Figure S4. The impact of *dcr-2* knockdown on miR* loading into AGO2.

The fraction of miR strands within all AGO1-associated miRNAs was analyzed from small RNA libraries of untreated cells or cells depleted of Dcr-2 (grey). In parallel, the percentage of AGO2-bound miR* species within all miRNA reads was plotted for the same conditions (black).

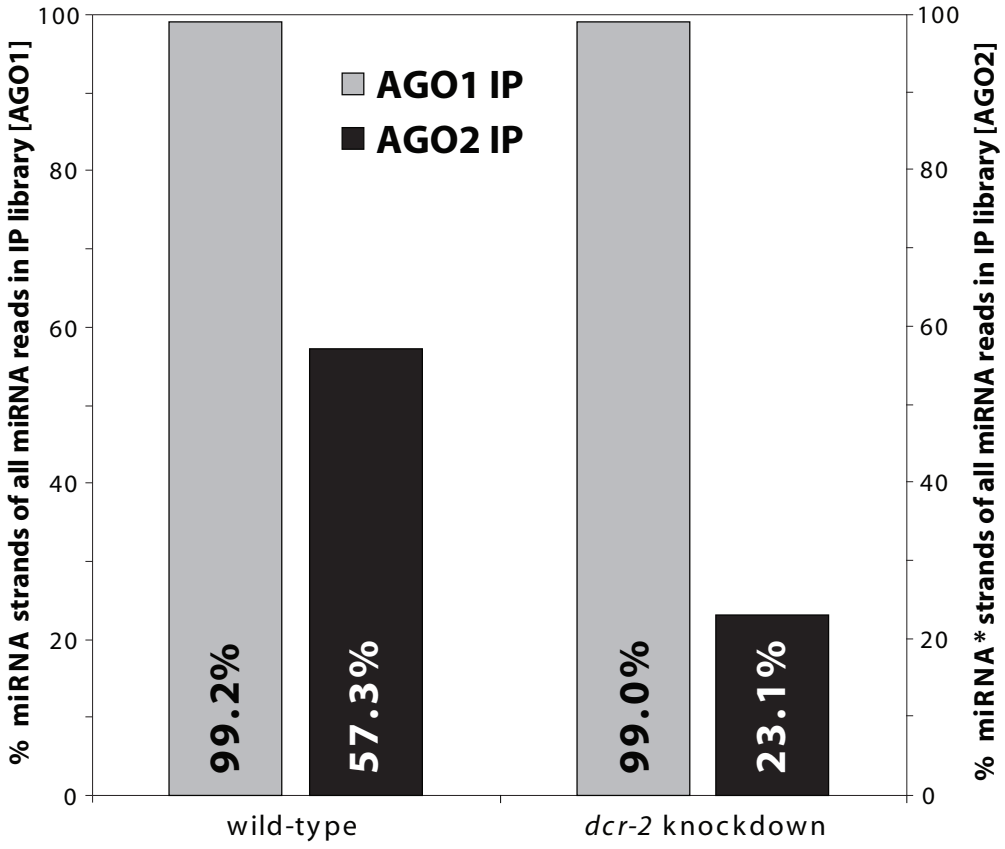

AGO1 IP
AGO2 IP
% miRNA strands of all miRNA reads in IP library [AGO1]
% miRNA* strands of all miRNA reads in IP library [AGO2]
100
80
60
40
20
0
99.2%
57.3%
99.0%
23.1%
wild-type
dcr-2 knockdown

Figure S5. Effects of Dcr-1 or AGO1 depletion on *miR-276a* and *miR-276a sensors in flies.**

Shown are sensors for *miR-276a* or *miR-276a** featuring perfectly matched or bulged target sites (as indicated to the left). **(A-D)**. Negative β-Gal staining (red in the merged images) indicates *dcr-1* mutant clones (also marked with arrows). Cells with strong β-Gal staining contain two wild-type *dcr-1* genes, while cells with intermediate staining are heterozygous for *dcr-1*. EGFP sensor activity is shown in green. The black and white panels indicate the separate channels for β-Gal and EGFP. (**E-H**) Clonal analysis for *ago1*: Details as in **(A-D)**. Selected regions (enclosed in white boxes) were zoomed in and shown as insets within each panel to visualize the smaller *ago1* clones.

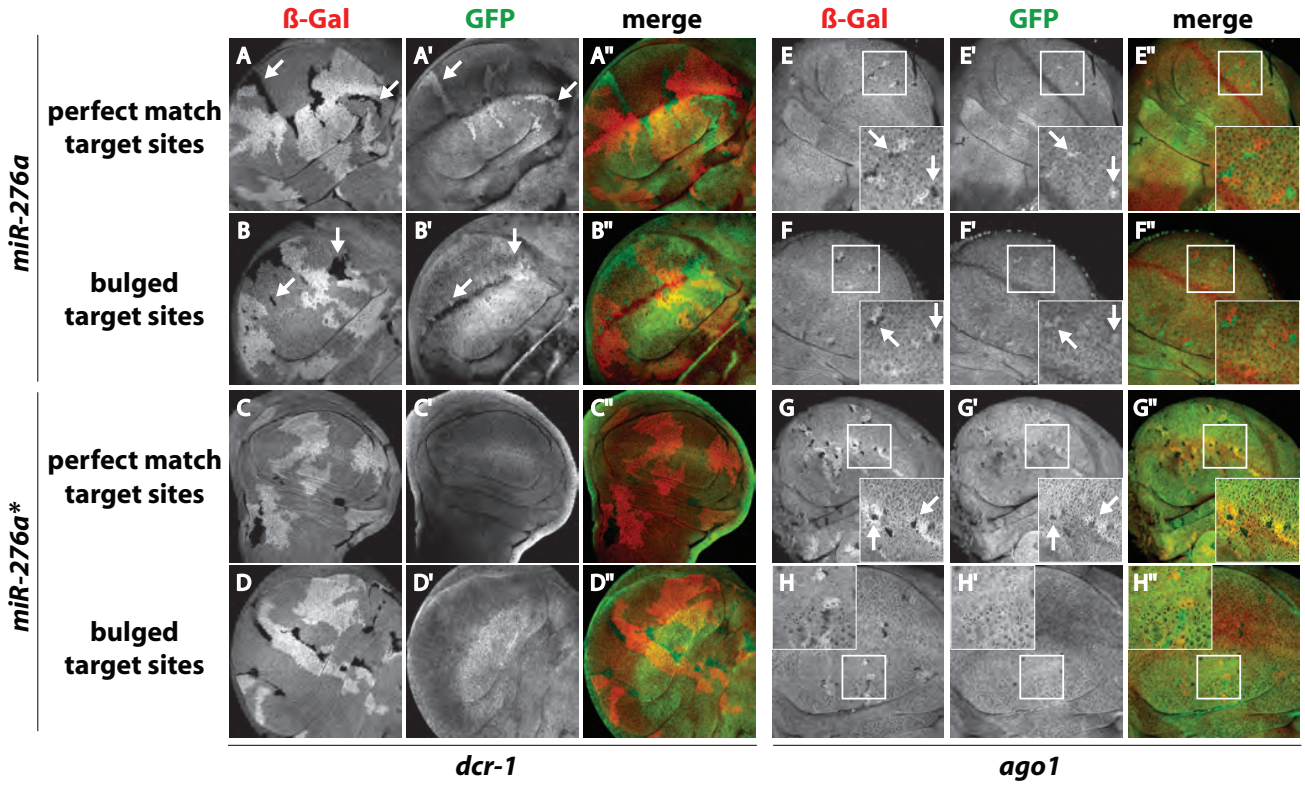

ß-Gal
GFP
merge
ß-Gal
GFP
merge
miR-276a
perfect match target sites
bulged target sites
miR-276a*
perfect match target sites
bulged target sites
A
A'
A"
B
B'
B"
C
C'
C"
D
D'
D"
E
E'
E"
F
F'
F"
G
G'
G"
H
H'
H"
dcr-1
ago1

Figure S6. Effects of AGO2 depletion on microRNA and microRNA* sensors in flies.

Shown are sensors for the miRNA strand or miRNA* strand of *miR-276a* **(A-D)** or *miR-bantam* (**E-H**) featuring perfectly matched or bulged target sites (as indicated to the left) in *ago2* clones.

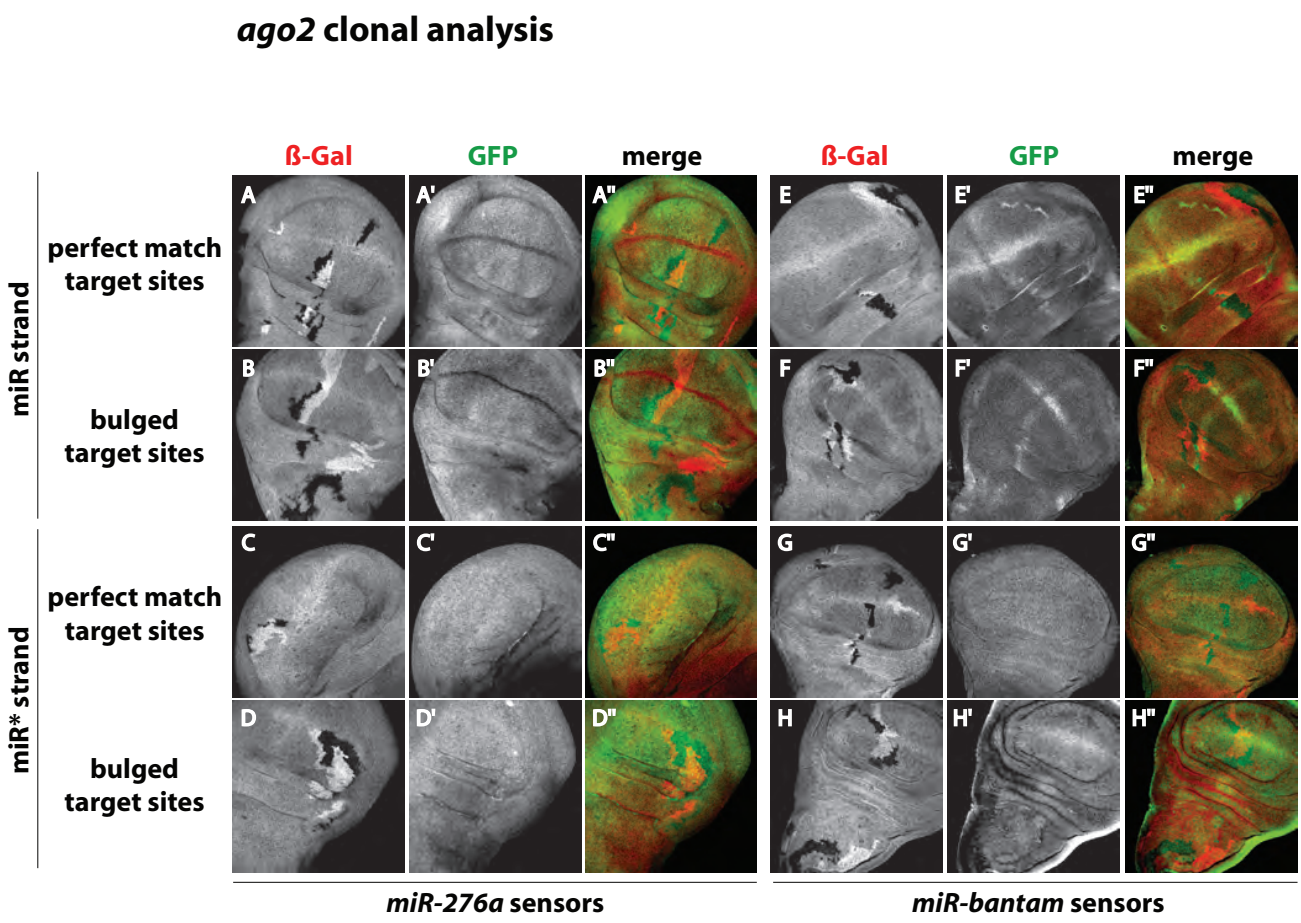

ago2 clonal analysis
ß-Gal
GFP
merge
ß-Gal
GFP
merge
miR strand
perfect match target sites
bulged target sites
miR* strand
perfect match target sites
bulged target sites
A
A'
A"
B
B'
B"
C
C'
C"
D
D'
D"
E
E'
E"
F
F'
F"
G
G'
G"
H
H'
H"
miR-276a sensors
miR-bantam sensors

Figure S7. Effects of Dcr-2 depletion on microRNA and microRNA* sensors in flies.

Shown are sensors for the miRNA strand or miRNA* strand of *miR-276a* **(A-D)** or *miR-bantam* (**E-H**) featuring perfectly matched or bulged target sites (as indicated to the left) in *dcr-2* clones.

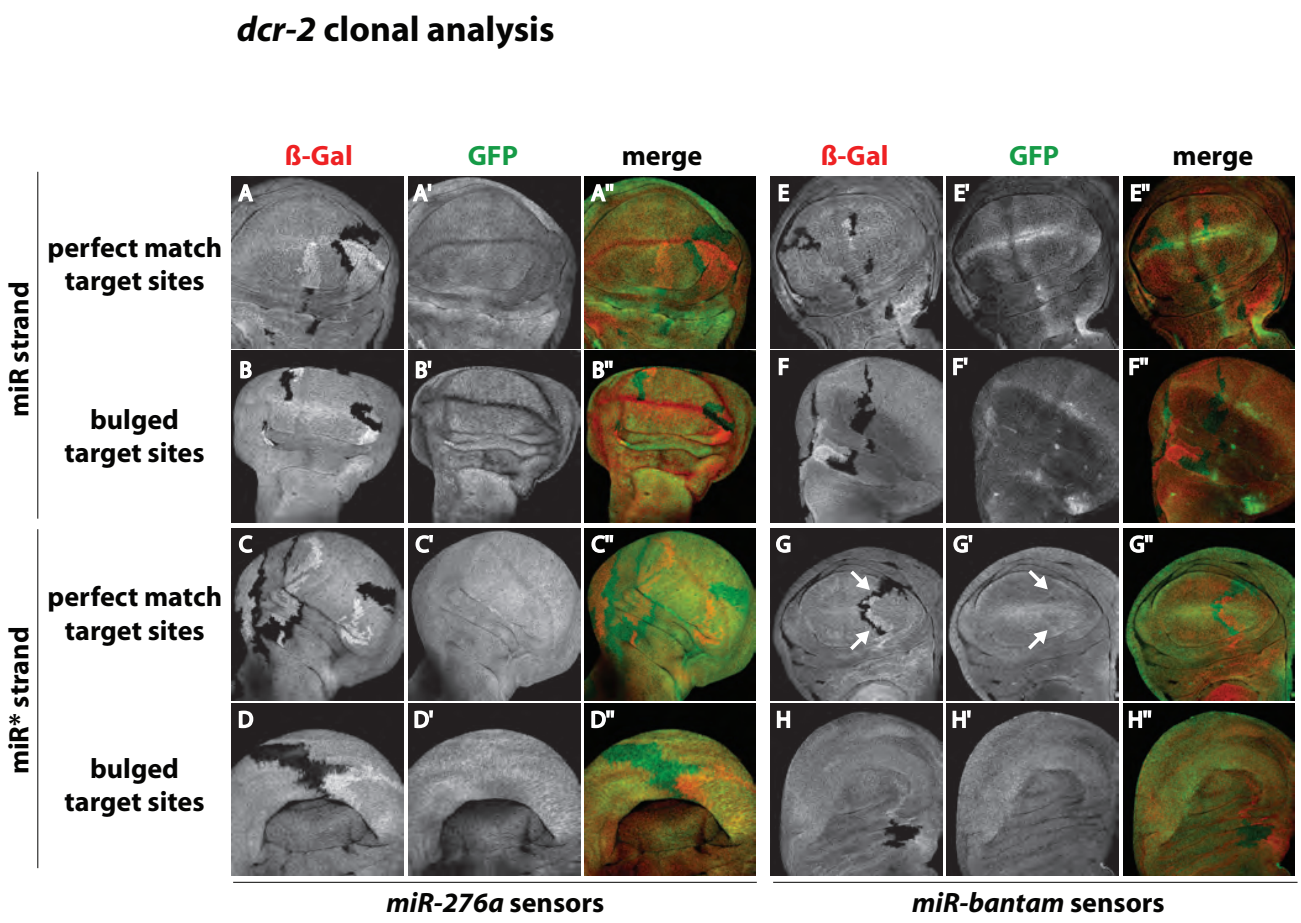

dcr-2 clonal analysis
ß-Gal
GFP
merge
ß-Gal
GFP
merge
miR strand
perfect match target sites
bulged target sites
miR* strand
perfect match target sites
bulged target sites
A
A'
A"
B
B'
B"
C
C'
C"
D
D'
D"
E
E'
E"
F
F'
F"
G
G'
G"
H
H'
H"
miR-276a sensors
miR-bantam sensors

Figure S8. Effects of R2D2 depletion on microRNA and microRNA* sensors in flies.

Shown are sensors for the miRNA strand or miRNA* strand of *miR-276a* **(A-D)** or *miR-bantam* (**E-H**) featuring perfectly matched or bulged target sites (as indicated to the left) in *r2d2* clones.

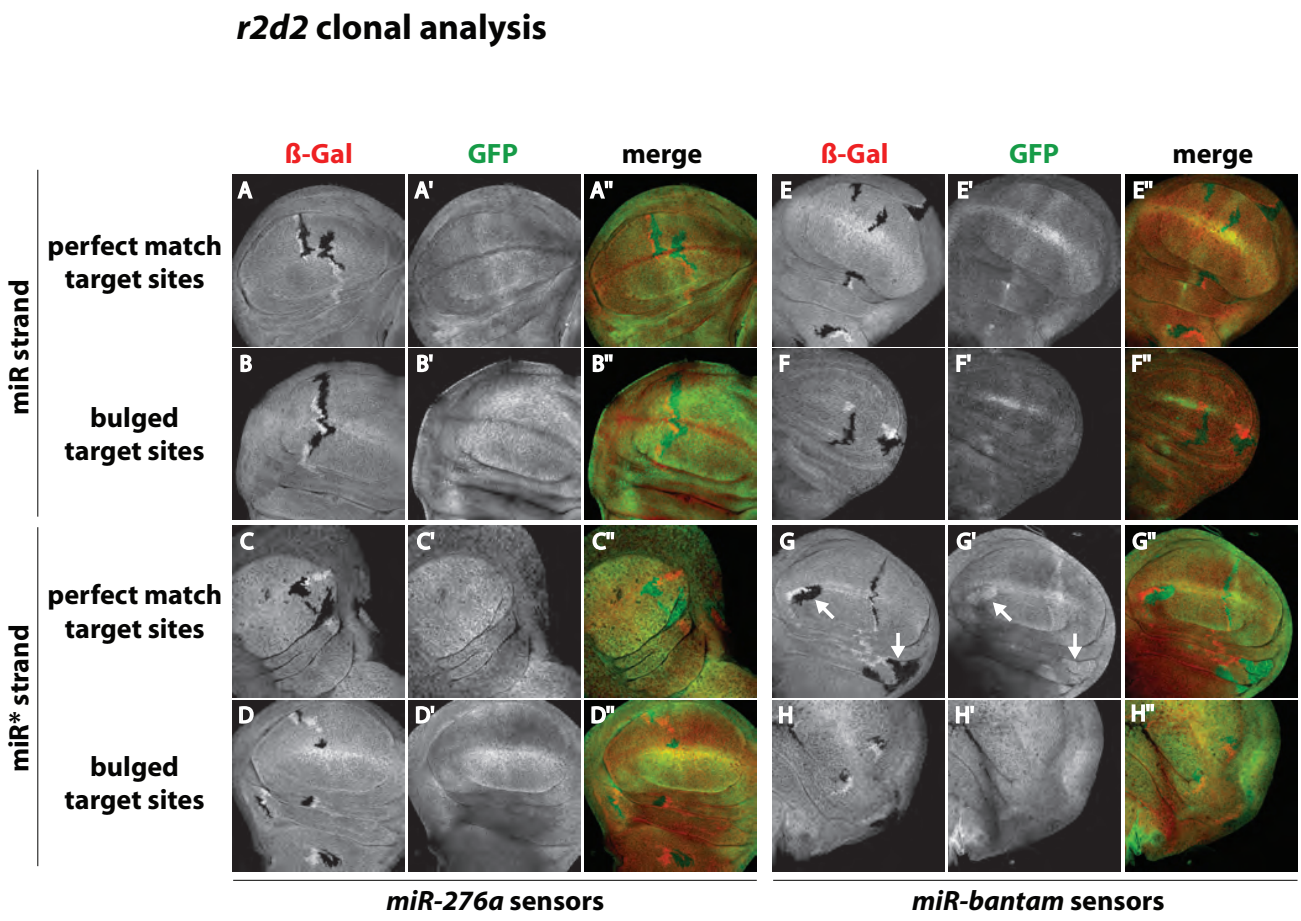

r2d2 clonal analysis
ß-Gal
GFP
merge
ß-Gal
GFP
merge
miR strand
perfect match target sites
bulged target sites
miR* strand
perfect match target sites
bulged target sites
A
A'
A"
B
B'
B"
C
C'
C"
D
D'
D"
E
E'
E"
F
F'
F"
G
G'
G"
H
H'
H"
miR-276a sensors
miR-bantam sensors

Figure S9. Effects of Loqs depletion on on microRNA and microRNA* sensors in flies.

Shown are sensors for the miRNA strand or miRNA* strand of *miR-276a* **(A-D)** or *miR-bantam* (**E-H**) featuring perfectly matched or bulged target sites (as indicated to the left) in *loqs* clones.

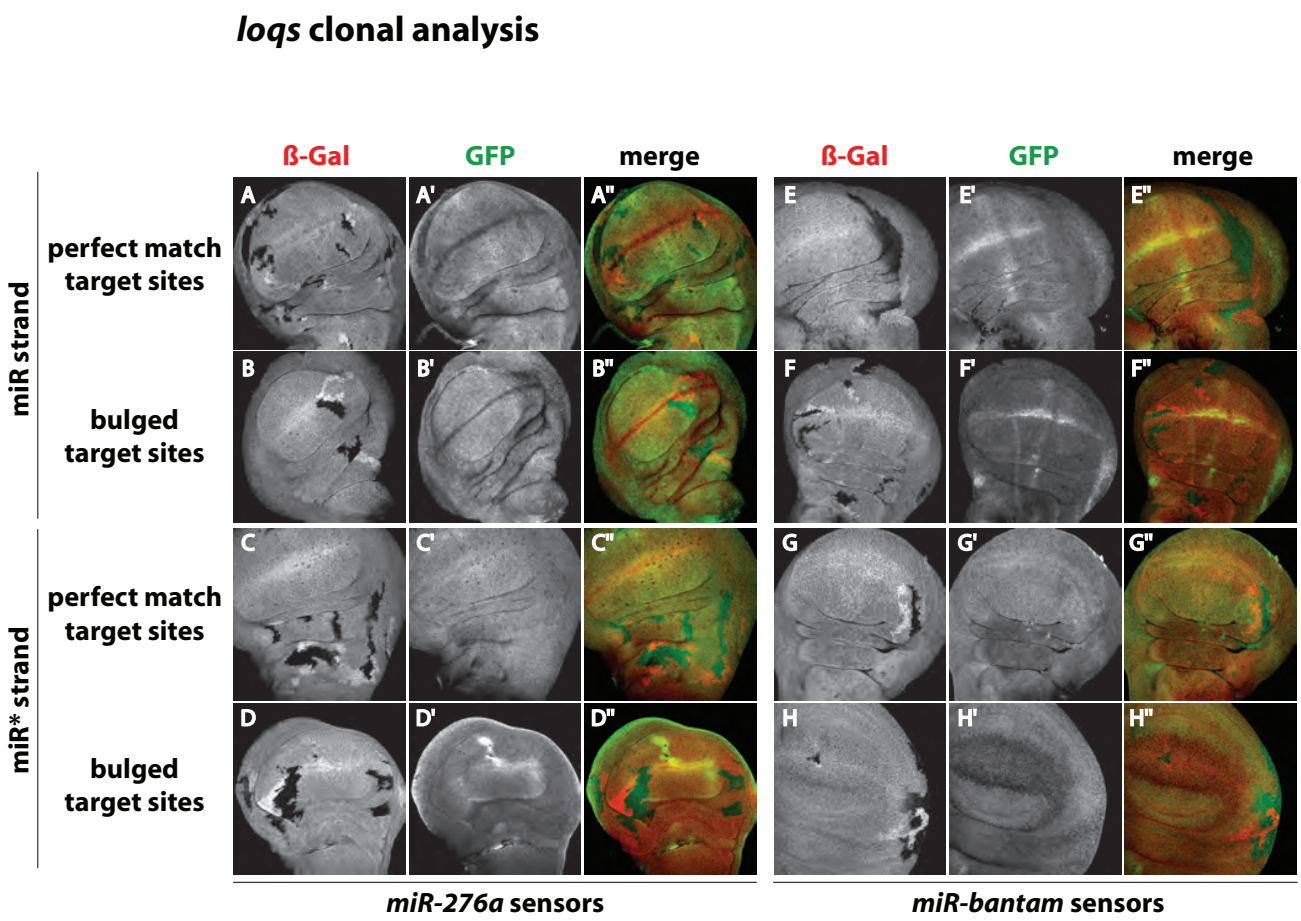
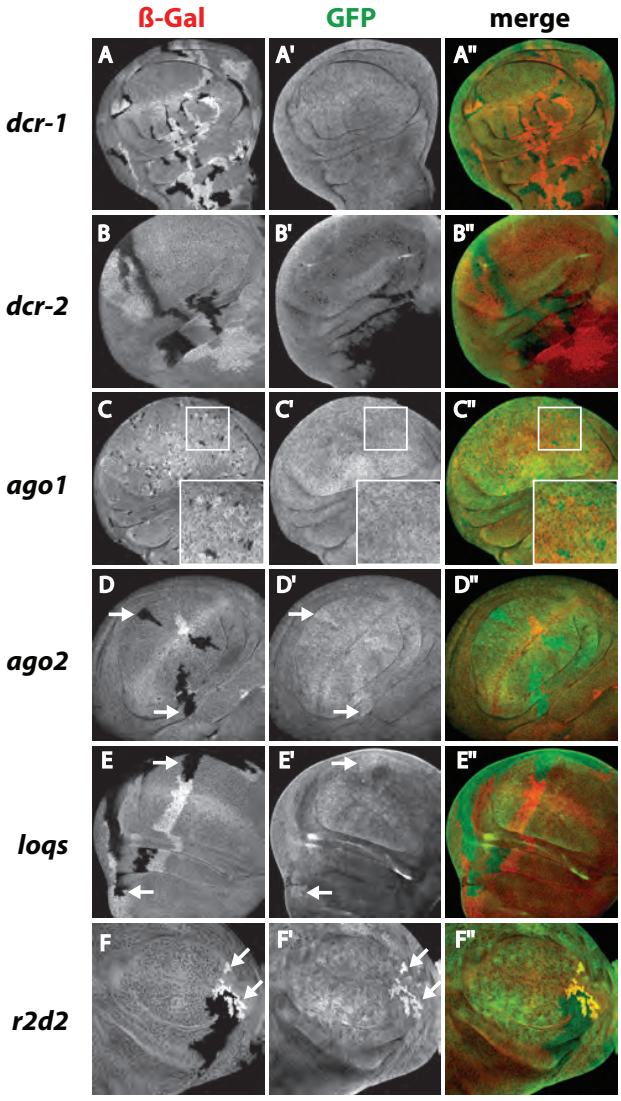

loqs clonal analysis
ß-Gal
GFP
merge
ß-Gal
GFP
merge
miR strand
perfect match target sites
bulged target sites
miR* strand
perfect match target sites
bulged target sites
A
A'
A"
B
B'
B"
C
C'
C"
D
D'
D"
E
E'
E"
F
F'
F"
G
G'
G"
H
H'
H"
miR-276a sensors
miR-bantam sensors

Figure S10. Effects of depletion of canonical miRNA and siRNA pathway components on *esi-2.1* sensor in flies.

Shown are sensors for *esi-2.1* featuring perfectly matched target sites in mitotic clones carrying mutations in genes encoding canonical miRNA and siRNA pathway components. **(A)** *dcr-1*; **(B)** *dcr-2*; **(C)** *ago1*; **(D)** *ago2*; **(E)** *loqs*; **(F)** *r2d2*.


ß-Gal
GFP
merge
dcr-1
dcr-2
ago1
ago2
loqs
r2d2
A
A'
A"
B
B'
B"
C
C'
C"
D
D'
D"
E
E'
E"
F
F'
F"
esi-2.1 sensor

Small RNA sorting: matchmaking for Argonautes

Benjamin Czech and Gregory J. Hannon

Nat Rev Genet. 2011 Jan;12(1):19-31. Epub 2010 Nov 30.

Small RNA sorting: matchmaking for Argonautes

Benjamin Czech and Gregory J. Hannon

Abstract | Small RNAs directly or indirectly impact nearly every biological process in eukaryotic cells. To perform their myriad roles, not only must precise small RNA species be generated, but they must also be loaded into specific effector complexes called RNA-induced silencing complexes (RISCs). Argonaute proteins form the core of RISCs and different members of this large family have specific expression patterns, protein binding partners and biochemical capabilities. In this Review, we explore the mechanisms that pair specific small RNA strands with their partner proteins, with an eye towards the substantial progress that has been recently made in understanding the sorting of the major small RNA classes — microRNAs (miRNAs) and small interfering RNAs (siRNAs) — in plants and animals.

RNase III protein
A member of a family of ribonucleases that process dsRNA, leaving 5′ monophosphates and 2-nt 3′ overhangs with hydroxyl ends. Drosha and Dicer are examples of such ribonucleases.

RNA-induced silencing complex
A regulatory multi-protein complex containing an Argonaute protein bound to a single-stranded small RNA that regulates gene expression through sequence complementarity between the guide RNA and the target transcript.

Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. Correspondence to G.J.H. e-mail: hannon@cshl.edu
doi:10.1038/nrg2916
Published online 30 November 2010

The discovery of RNA interference (RNAi) in the late 1990s sparked a renaissance in our understanding of RNAs as regulatory molecules. A growing number of small RNA classes has since emerged from studies of eukaryotic organisms, and these RNAs can be approximately divided into two groups: small RNAs that engage RNAi-related machinery and those that do not. As yet, we know very little about many newly discovered groups of small RNAs, but our understanding of the biogenesis and biological functions of RNAi-related small RNA classes is growing rapidly.

Small RNAs that engage RNAi-related pathways share several characteristic features. They are mainly ~20–30 nucleotides (nt) in length, have 5′ phosphate groups and 3′ hydroxyl (-OH) (although sometimes modified) termini, and they associate with specific members of a large protein family — the Argonautes. The precise combination of a small RNA with a particular Argonaute protein determines its biological function. Therefore, it is crucial that these very similar species are appropriately sorted among closely related partners. Only then can the target specificity conferred on Argonaute proteins by their small RNA guides enable their myriad important roles, which include the regulation of gene expression, modification of chromosome structure and protection from mobile elements. Conceptually, all small RNA-mediated regulatory events can be considered as the culmination of several consecutive steps: small RNA biogenesis, strand selection (in which dsRNA is the precursor), loading into Argonaute, target recognition and effector function.

The biogenesis of most small RNA classes, including microRNAs (miRNAs) and many small interfering RNAs (siRNAs), requires the action of RNase III family proteins (reviewed in REFS 1–3). Some small RNA classes, including Piwi-interacting RNAs (piRNAs) and secondary siRNAs in worms, however, are not derived from dsRNA precursors and are produced through alternative biogenesis mechanisms independently of RNase III enzymes[4–8].

Following their production, small RNAs are sorted to confer association with specific Argonaute family proteins, which function as the core of the RNA-induced silencing complex (RISC). Argonaute proteins can be classified into three subgroups according to their sequence relationships: the AGO subfamily, the Piwi subfamily and the worm-specific WAGO clade[9–11]. Piwi subfamily proteins load small RNAs derived from single-stranded precursors (piRNAs) and AGO clade proteins usually associate with small RNA duplexes processed by RNase III endonucleases (miRNAs and siRNAs; reviewed in REFS 1,2). Small RNAs that occupy WAGO clade proteins are usually direct products of RNA synthesis[6,7,9].

Mature RISC consists of a single-stranded small RNA bound to an Argonaute protein. As some small RNAs are generated as duplexes, only one strand (the guide strand) is retained and the other (passenger) strand is discarded during RISC assembly[12–14]. AGO clade proteins are generally loaded with small RNA duplexes before RISC maturation. Thus, it is of key importance to assemble RISC in a manner that ensures that the appropriate guide strand is selectively stabilized, as loading of the

passenger strand would obviously misdirect RISC towards inappropriate targets. Small RNAs guide mature RISC through complementary base pairing to its targets, with the most common outcome being target repression (reviewed in REFS 15–17).

The knowledge of the mechanisms that guide a particular small RNA strand into a specific Argonaute family member is crucial. It impacts our ability to predict the biological function of a small RNA and to effectively use small RNAs as experimental tools or therapeutics. This Review focuses on our understanding of small RNA sorting in plants and animals. We consider biogenesis as a starting point as this affects the nature of small RNAs and, in some cases, the complexes which the small RNAs join. Next, we discuss the small RNA-intrinsic determinants of sorting, followed by RISC loading and maturation. Finally, we briefly cover the implications of sorting for Argonaute function. We do not extensively discuss the effector mechanisms of mature RISC, but instead refer the reader to several excellent recent reviews on this topic[15–17].

Small RNA biogenesis

In effect, the first step of small RNA sorting is biogenesis, as this determines the small RNAs that are available for RISC loading. Moreover, the precise enzymes that liberate small RNAs from their precursor transcripts or generate them *de novo* seem to impact the choice of their ultimate Argonaute partner. Therefore, it is important to begin with an introduction to the varied mechanisms that can produce small RNAs.

Small RNA duplexes from partial or perfect dsRNA precursors are generated by RNase III family enzymes through sequential endonucleolytic cleavage events. These enzymes often partner with dsRNA binding domain (dsRBD) proteins, which serve to increase substrate specificity and affinity, leading to increased activity. The resulting products are duplex ~20–24-nt small RNAs consisting of two strands (the guide or miR and passenger or miR* strands). These small RNAs feature 5′ monophosphates and 2-nt overhangs that have hydroxyl groups at their 3′ termini.

Animal miRNA processing. miRNAs are ubiquitous in animal genomes and are often transcribed as separate coding units, many of which consist of polycistronic clusters containing multiple miRNAs. Some miRNAs are also present in introns and presumably arise from further processing of the excised introns of protein-coding genes[18]. Most miRNAs are transcribed by DNA-dependent RNA polymerase II (RNAPII) to generate a primary miRNA (pri-miRNA) containing a region of imperfect dsRNA, known as the stem–loop structure, that harbours the future mature miRNA[19,20] (FIG. 1). Primary miRNA transcripts seem largely like the transcripts of protein-coding genes. They have 5′ cap structures, polyA tails and may contain introns. The production of conventional miRNAs from these precursors proceeds through two site-specific cleavage events. Processing likely begins with a dsRBD protein, Pasha/DiGeorge syndrome critical region gene 8 (DGCR8), binding to the pri-miRNA and recruiting the RNase III enzyme Drosha to form a multiprotein complex called the Microprocessor[21–24]. This complex recognizes the duplex character of the pri-miRNA, although the precise RNA–protein interactions that select pri-miRNAs as Microprocessor substrates and how the cleavage site is determined by these interactions are matters of ongoing work. The pri-miRNA is cleaved by Drosha to liberate a ~60–70-nt precursor miRNA (pre-miRNA) from the primary transcript[25]. The nuclear export protein Exportin 5 recognizes the 2-nt single-stranded 3′ overhang of the pre-miRNA (characteristic of RNase III-mediated cleavage) and actively transports it in a Ran–GTP-dependent manner to the cytoplasm[26–28]. Additional factors, including the nuclear export receptor Exportin 1 (XPO1), the cap-binding complex (CBC) and the *Arabidopsis thaliana* SERRATE homologue, ARSENITE-RESISTANCE PROTEIN 2 (ARS2), were recently suggested to play a part in the transition from pri- to pre-miRNA[29–31].

Once in the cytoplasm, the pre-miRNA is cleaved into a ~22–23-nt miRNA:miRNA* duplex by Dicer[32–35]. For this purpose, the sole mammalian Dicer partners with the dsRBD protein TAR RNA-binding protein 2 (TARBP2, also known as TRBP)[36,37], whereas the *Drosophila melanogaster* miRNA-generating Dicer 1 (DCR1) similarly interacts with a specific isoform of its dsRBD protein partner Loquacious (LOQS-PB)[38–42]. Small RNA duplexes generated by Dicer (and its protein partner) exhibit 2-nt single-stranded 3′ overhangs at both ends, a signature of RNase III cleavage.

Several unconventional miRNAs that are defined by their use of alternative maturation strategies have now been noted. For example, mirtrons have been found in flies and mammals[43–45]. Mirtrons bypass the Drosha processing step and instead use the splicing machinery to generate pre-miRNAs. Mirtrons are very short introns and are excised, debranched and refolded into short stem–loop structures that mimic pre-miRNAs and are processed into mature miRNAs by Dicer. A few recently discovered mirtrons in flies are initially generated with extended 3′ tails that must be resected by the exosome to form a pre-miRNA suitable for Dicer processing[46].

miRNA biogenesis in plants. Plant miRNAs are transcribed by RNAPII to yield capped and polyadenylated pri-miRNAs with local stem–loop structures that are potentially stabilized by the RNA-binding protein DAWDLE (DDL)[47]. Plant pri-miRNAs typically display greater diversity in the size and structure of their stem–loops compared with their animal counterparts[48]. As plants lack a Drosha orthologue, pri-miRNAs are converted into mature miR:miR* duplexes by a single RNase III family enzyme, DICER-LIKE 1 (DCL1)[48–51], which fulfils the functions of both Drosha and Dicer (BOX 1). As in animals, Dicer is assisted by a dsRBD protein, in this case, HYPONASTIC LEAVES 1 (HYL1)[52–54]. HYL1 and the zinc finger protein SERRATE promote accurate miRNA processing[53,55–57]. miRNA maturation is also aided by the nuclear cap-binding complex[53,58,59], probably by facilitating the loading of miRNA-processing factors onto pri-miRNAs.

Guide strand
During RISC loading, one strand of an siRNA is selected and stabilized. This is termed the guide strand, and it confers target specificity. miRNA guide strands are termed miR strands.

Passenger strand
The non-incorporated strand of the siRNA duplex that is degraded during the assembly of RISC. Non-incorporated strands of miRNAs are called miR* strands.

Stem–loop structure
A region of dsRNA (stem) connected by an unpaired region (loop) in a single RNA molecule. This is a structure typical for miRNA precursors.

Mirtron
A miRNA that originates from a very short intron and is excised to form a pre-miRNA by the splicing machinery (and occasionally subsequent trimming), therefore bypassing the Drosha processing step.

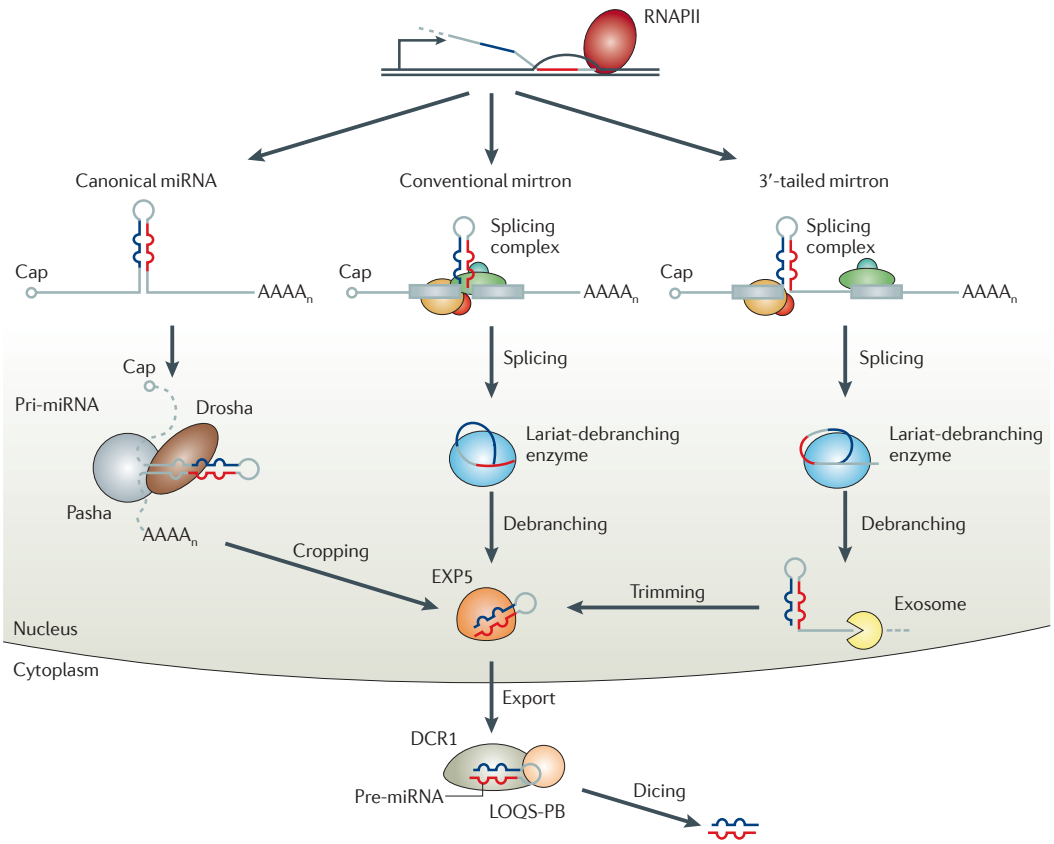


Figure 1 | **MicroRNA biogenesis in *Drosophila melanogaster*.** MicroRNAs (miRNAs) are generally transcribed by RNA polymerase II (RNAPII) to yield primary miRNAs (pri-miRNAs). pri-miRNAs are cropped in the nucleus by Drosha–Pasha complexes to release shorter precursor miRNAs (pre-miRNAs). miRNAs that reside within short introns of protein-coding genes are excised by the splicing machinery and are termed mirtrons. Following linearization of mirtron intermediates by the lariat-debranching enzyme, they fold into pre-miRNAs. Some 3′-tailed mirtrons undergo further trimming by the exosome. pre-miRNAs are transported to the cytoplasm by Exportin 5 (EXP5), where further processing takes place. Dicer 1 (DCR1), in collaboration with an isoform of its dsRNA binding domain protein partner Loquacious (LOQS-PB), liberates miR:miR* duplexes that dissociate from DCR1 for downstream sorting.

Maturation of plant miRNA duplexes often proceeds through several rounds of sequential Dicing from the base of a long stem–loop (BOX 1). Processed miRNA duplexes are modified by the methyltransferase HUA ENHANCER 1 (HEN1)[60–62]. In contrast to its *D. melanogaster* homologue, plant HEN1 is nuclear and adds methyl groups to the 3′ ends of both strands of the miR:miR* duplex. This 2′-*O*-methylation is thought to protect miRNAs from further modifications, such as 3′ uridylation[60,62], which mark single-stranded miRNAs for destruction by exonucleases of the SMALL RNA-DEGRADING NUCLEASE (SDN) family[63]. This adaptation may be necessitated by the fact that plant miRNAs pair extensively with target mRNAs and cleave them, a process which in animals provides a trigger for small RNA destruction[64]. Following methylation by HEN1, miR:miR* duplexes are thought to be transported by an Exportin 5 homologue, HASTY (HST), or through HST-independent mechanisms to the cytoplasm[65], where sorting and RISC assembly takes place. However, the exact form of the exported cargo and the subcellular localization of plant RISC loading and maturation remain subjects of current debate[3]. In this regard, a recent study proposed a model in which RISC is assembled in the nucleus and only mature AGO1–RISC containing a single-stranded miR can be exported to the cytoplasm[66].

siRNAs of endogenous or exogenous origin. The first siRNAs were discovered in plants[67]. The earliest identified examples were derived from viral replication intermediates or complex interactions between transgene copies. By considering the commonalities between these origins, dsRNAs were indicated as the source of small RNAs. It is now clear that plants and animals produce a wide range of siRNAs. These vary in their biogenesis mechanisms, but can be approximately divided into two classes, depending on whether they require RNA-dependent RNA polymerases (RdRPs) for their production.

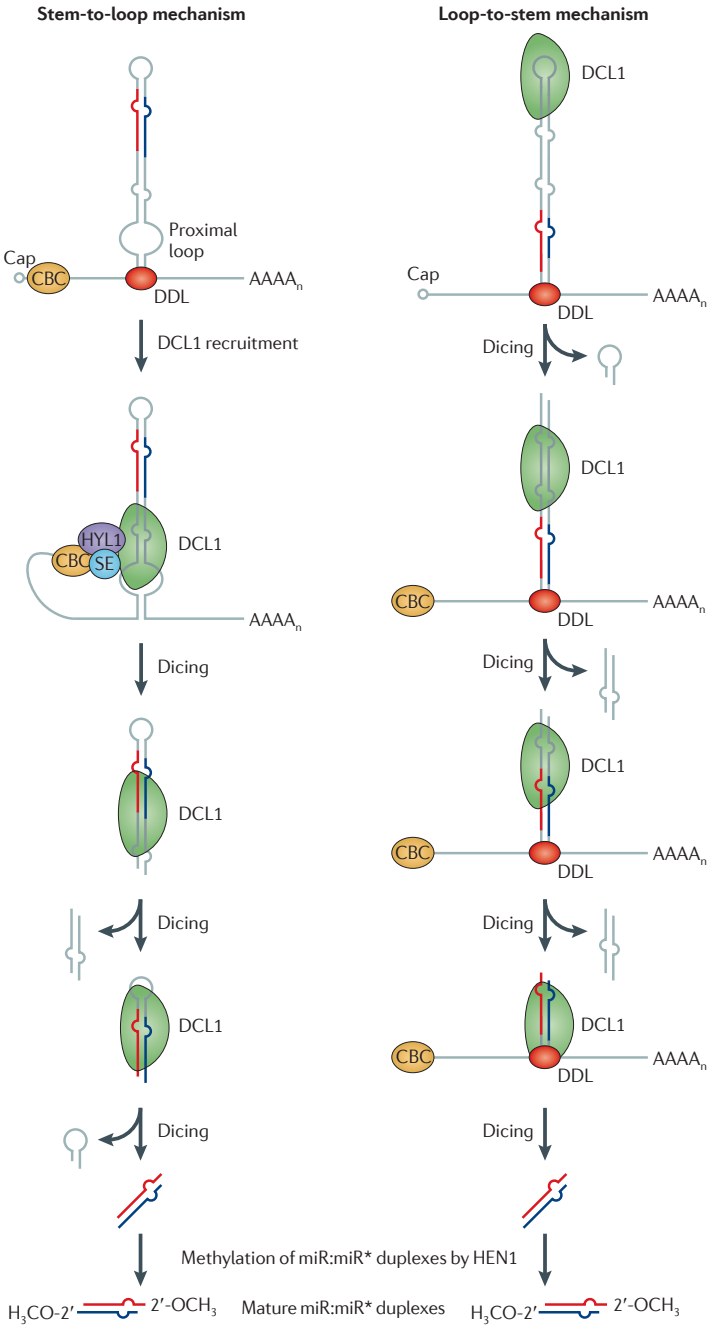
siRNAs derived from dsRNAs. The process of converting dsRNA into small RNAs is perhaps currently best understood in *D. melanogaster*. Here, the experimental

Dicing
Refers to the cleavage events carried out by the RNase III family nuclease Dicer.

RNA-dependent RNA polymerase
An RNA polymerase that uses ssRNA as a template to synthesize dsRNA.

Box 1 | **Plant microRNA-processing mechanisms**

Plant microRNAs (miRNAs) are generally produced by sequential rounds of Dicing. This is necessitated by the lack of a Drosha orthologue. The extensive nature of the hairpins that lead to many plant miRNAs also permits phased production of multiple small RNA duplexes through sequential Dicing events, conceptually the plant version of long hairpin endogenous small interfering RNA (siRNA) precursors or miRNA polycistrons in animals. **a** | Usually, consecutive Dicing proceeds from the base of the stem–loop. The secondary structure of the primary miRNA (pri-miRNA) flanking the mature miR:miR* duplex is important for proper and efficient processing, analogous to the proposed role of the 'basal stem' of animal pri-miRNAs[156–159]. Accurate processing depends on a region of imperfect pairing (junction between ssRNA and dsRNA) approximately 15 nucleotides (nt) from the miR:miR* duplex (towards the free end of the stem-loop), which localizes DICER-LIKE 1 (DCL1) to its initial cleavage site. This liberates an intermediate similar to animal precursor miRNAs (pre-miRNAs), which is further processed by DCL1 into the mature miR:miR* duplex. **b** | Variations in processing mechanisms are possible. For example, miR319 and miR159 (both with conserved long precursors) are produced by an unusual loop-to-stem mechanism. Following the first cleavage of the loop by DCL1, consecutive cuts by DCL1 are necessary to release the mature miRNA duplex[160]. CBC, cap-binding complex; DDL, DAWDLE; HEN1, HUA ENHANCER 1; HYL1, HYPONASTIC LEAVES 1; SE, SERRATE.



introduction of long dsRNAs results in the production of exo-siRNAs that are ~21 nt in size (FIG. 2a). Long dsRNAs are processed into siRNA duplexes through sequential cleavage events by the RNase III protein Dicer 2 (DCR2) (REFS 68,69) in collaboration with its dsRBD co-factor, a particular Loquacious isoform, LOQS-PD[42,70]. Dicer 2 also interacts with another dsRBD protein R2D2, but only LOQS-PD enhances siRNA production[69,71]. Recent studies indicate a role of R2D2 in loading siRNA duplexes into RISC (discussed below), suggesting that these two dsRBD proteins may have distinct and sequential functions[71,72].

In flies, siRNAs also originate from numerous endogenous loci and were termed endogenous siRNAs (endo-siRNAs)[73–77]. These can originate from RNA transcripts with extensive hairpin structures, from convergent transcription units (similar to plant nat-siRNAs, see below) or from the annealing of sense and antisense RNAs from

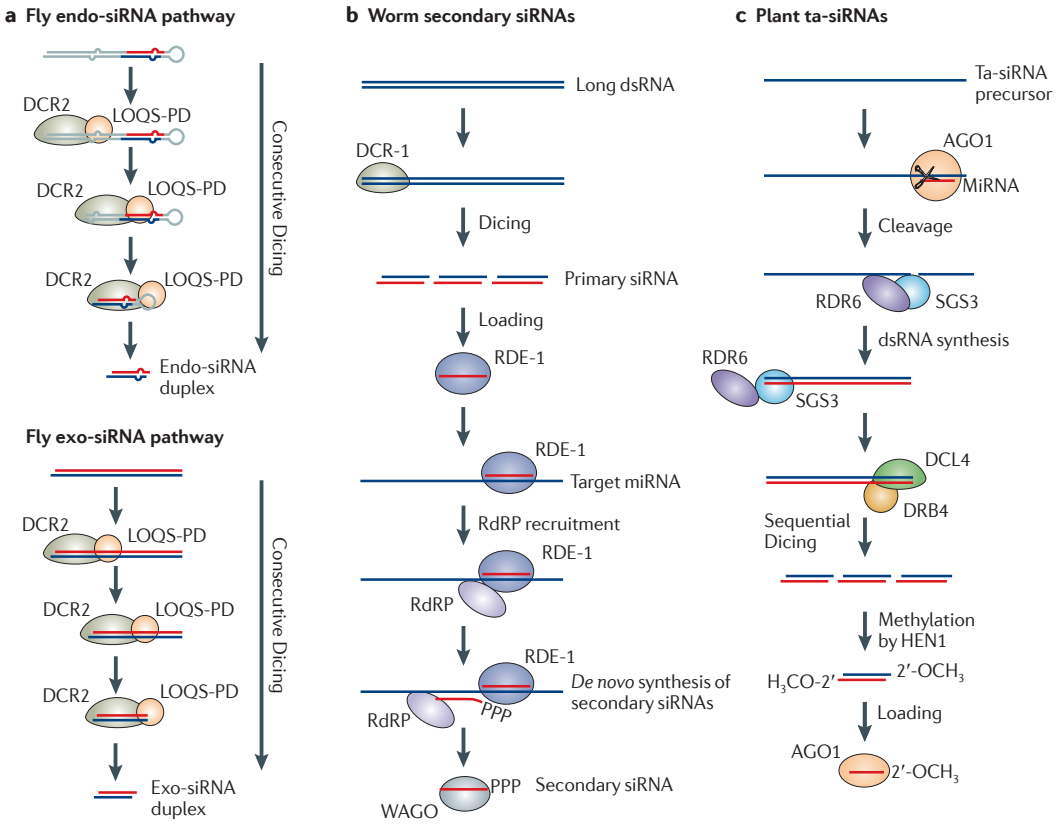


Figure 2 | **Production of small interfering RNAs. a** | In flies, perfect or nearly perfect dsRNA precursors of varying origin and structure are processed in the cytoplasm by the RNase III enzyme Dicer 2 (DCR2) and its co-factor, an isoform of Loquacious (LOQS-PD), to yield small interfering RNA (siRNA) duplexes that contain guide and passenger strands. **b** | *Caenorhabditis elegans* primary siRNAs are processed from long dsRNA triggers through the action of DCR-1. These primary siRNAs associate with the Argonaute family protein, RDE-1, and guide it to target transcripts. The RDE-1–target interaction recruits an RNA-dependent RNA polymerase (RdRP), which uses the target as template for the *de novo* synthesis of secondary siRNAs that feature 5′ triphosphate ends (see main text for further details). **c** | The production of *Arabidopsis thaliana trans*-acting siRNAs (ta-siRNAs) requires the interplay of canonical components of the microRNA (miRNA) and siRNA biogenesis machineries. The process is triggered by miRNA-mediated cleavage of non-coding TAS transcripts by miR390–AGO7 or miR173–AGO1, respectively. Slicing triggers the recruitment of SUPPRESSOR OF GENE SILENCING (SGS3) and RNA-DEPENDENT RNA POLYMERASE 6 (RDR6), which synthesize dsRNA using the cleavage site as the entry point. The resulting dsRNA is processed by DICER-LIKE 4 (DCL4) and its dsRBD protein partner DRB4 into a phased series of 21-nucleotide (nt) siRNA duplexes. ta-siRNAs are methylated by HUA ENHANCER 1 (HEN1) before AGO loading.

unlinked loci. One example of the latter type of siRNAs are endo-siRNAs that target transposons, which seem to arise at least in part from the hybridization of transposon mRNAs with piRNA cluster transcripts. Another possible source of dsRNA hybrids is the interaction of sense and antisense transcripts across individual transposon copies, and it has even been suggested that RdRPs may operate in animals to form dsRNAs[78]. As with exo-siRNAs, the biogenesis of endo-siRNAs depends on Dicer 2 assisted by LOQS-PD[42,73,75–77].

A similar situation has been described in mammals; however, the range of cell types in which dsRNAs are produced and converted into siRNAs seems to be limited. Thus far, endo-siRNAs have been detected in abundance only in mouse oocytes and embryonic stem (ES) cells[79–81]. The dsRNA triggers that give rise to murine endo-siRNAs are predicted to arise from *trans* interactions between gene and pseudogene transcripts, from overlapping transcription units and from transcripts that can form long hairpins. As in flies, endo-siRNA biogenesis is dependent on Dicer and, presumably, its dsRBD partners.

RdRP-dependent siRNAs. In contrast to mammals and flies, worms and plants produce numerous endo-siRNAs using biogenesis mechanisms that depend on the action of RdRPs. Plant RdRPs copy single-stranded precursors into long dsRNAs that are cleaved by Dicer, whereas worm RdRPs can directly synthesize siRNAs without Dicer processing.

Primary siRNAs in *Caenorhabditis elegans* are produced conventionally, from long dsRNA triggers through the action of DCR-1 (REFS 33,35,82)(FIG. 2b). The siRNAs

Trans-acting siRNA
A plant small RNA that primarily associates with AGO2. ta-siRNA biogenesis depends on miRNA-mediated cleavage of precursors that are further processed by DCL4 and other siRNA machinery factors.

Natural antisense transcript-derived siRNA
A stress-induced small RNA produced by DCL1 and DCL2 that originates from dsRNA formed by convergent transcription.

associate with the Argonaute family protein, RDE-1 and guide it to target transcripts. The RDE-1–target interaction recruits an RdRP, an outcome that is independent of RDE-1 catalytic activity[83]. The RdRP uses the target as template for the synthesis of secondary siRNAs of 22– 24 nt. Secondary siRNAs possess triphosphates at their 5′ ends, indicating that each small RNA is produced as a discrete moiety by *de novo* synthesis[6,7].

The production of most plant siRNAs requires the action of RdRPs to convert ssRNA precursors to dsRNA triggers. Three major subclasses of endogenous siRNAs can be distinguished in plants: *trans*-acting siRNAs (ta-siRNAs), natural antisense transcript-derived siRNAs (nat-siRNAs) and heterochromatic siRNAs (hc-siRNAs). Each of these small RNA subclasses is produced by a specific Dicer family member and preferentially loaded into a distinct AGO complex.

The biogenesis of ta-siRNAs requires the interplay of canonical components of miRNA and siRNA processing[84–90] (FIG. 2c). The process begins with miRNA-mediated cleavage of the *TAS1* or *TAS3* non-coding RNAs by miR390–AGO7 or miR173–AGO1, respectively. This triggers the recruitment of SUPPRESSOR OF GENE SILENCING 3 (SGS3) and RNA-DEPENDENT RNA POLYMERASE 6 (RDR6), which synthesizes dsRNA using the cleavage site as the entry point. The resulting dsRNA is processed by DCL4 and its dsRBD protein partner DRB4 into a phased series of 21-nt siRNA duplexes, which begins at the site of initial cleavage. ta-siRNAs are methylated by HEN1 before AGO loading. The subcellular localization of biogenesis factors and RNA intermediates, along with the recruitment of SDE5 (a putative export factor homologue), suggests that ta-siRNA biogenesis might involve specific nuclear–cytoplasmic shuttling[3,91].

Plant genomes often possess convergent transcription units that can give rise to dsRNA. Under certain conditions, often resulting from biotic and abiotic stress, bidirectional transcription is induced and the resulting dsRNA is processed into nat-siRNAs[92–94]. Production of nat-siRNAs requires DCL2 (which produces 24-nt siRNAs) or DCL1 (resulting in 22-nt siRNAs), depending on the genomic origin of the overlapping transcripts. Other essential biogenesis factors include RDR6, SGS3, HYL1, HEN1 and RNAPIV[92,93].

A highly abundant class of plant endo-siRNAs — hc-siRNAs — arises from repeats and transposable elements[95–101]. hc-siRNAs are predominantly 24 nt in size and their biogenesis, which is thought to occur in nucleolar bodies, depends on DCL3, its partner protein CLASSY1 (a SNF2 domain protein), the RdRP RDR2 and the plant-specific DNA-dependent RNA polymerases RNAPIV and RNAPV. Processed siRNA duplexes are methylated by HEN1 and primarily loaded into AGO4.

Small RNA sorting

Once produced, small RNAs and, in many cases, specific small RNA strands must be loaded into Argonaute proteins. Sorting is influenced by the Dicer that processes the precursor, the structure of the small RNA duplex, its terminal nucleotides, its thermodynamic properties and the destination AGO protein (see BOX 2 for structural properties of AGO proteins).

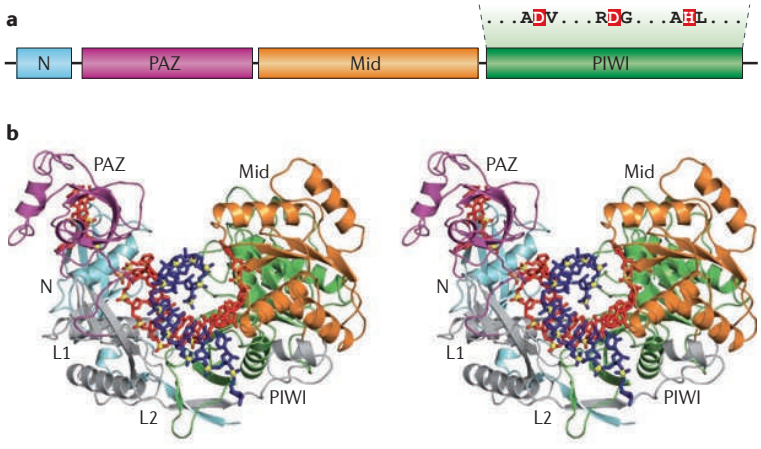
In part, sorting may be driven by specific protein–protein interactions between biogenesis and effector components. For example, in animals, Dicing and Argonaute loading have been proposed to occur as concerted processes[102,103]. This provides an opportunity for determining the fate of specific precursors to join certain effector complexes if a particular Dicer preferentially binds one Argonaute family member. However, Dicer and Argonaute cannot be the full story. Instead, it is clear that more complex-loading and strand-recognition pathways also influence the sorting of small RNAs. To exert its regulatory functions, mature RISC must be

Box 2 | **Structural determinants of Argonaute proteins for small RNA sorting**

Argonaute (AGO) proteins provide numerous possibilities for RNA–protein interactions that might underlie the proposed determinants of small RNA strand sorting. The interaction between AGOs and small RNAs occurs through several contact points in three characteristic domains of the protein: the PAZ, Mid and PIWI domains (**a** and **b**; part **b** shows a stereo view of the crystal structure of *Thermus thermophilus* AGO bound to a guide DNA–target RNA duplex[161]).

The PAZ domain hosts the 3′ end of the small RNA[162,163], whereas the Mid domain forms a binding pocket that anchors the 5′ phosphate of the terminal nucleotide of the small RNA[111,161,164–167]. These interactions provide opportunities for base-specific contacts that might provide preferences for 5′ nucleotides or might encourage the loading of duplexes with unstable 5′ ends. Whereas plant, fly and worm microRNAs (miRNAs) show a strong tendency to start with U, human miRNAs are biased towards U or A as 5′ terminal nucleotides[73,76,106–109,111,112]. Recent work provides structural evidence for nucleotide-specific interactions in the Mid domain of human AGO2 that ensure the preference for a 5′ terminal U (or A), while excluding G or C through a nucleotide specificity loop[111]. Interestingly, this structure is well conserved in all four human AGO proteins as well as in *Drosophila melanogaster* AGO1 or the worm miRNA acceptors ALG-1 and ALG-2. By contrast, AGO proteins that function in other small RNA pathways, such as *D. melanogaster* AGO2 or plant AGOs, lack this nucleotide specificity loop[111]. Whether the region corresponding to the nucleotide specificity loop in these distant proteins contributes to sorting of small RNAs, depending on the 5′ nucleotide or not, awaits further structural investigation.

The PIWI domain, which shows similarity to RNase H folds, harbours the residues required for catalytic activity (in AGO protein usually Asp–Asp–His). Thus, cleavage-competent AGO proteins carry out endonucleolytic cleavage of target transcripts through their PIWI domain[164,168–170]. Cleavage products of AGO enzymes feature 3′ hydroxyl and 5′ phosphate ends[171,172].



Panel **b** is reproduced from REF. 161 © (2008) Macmillan Publishers Ltd. All rights reserved.

programmed with a single-stranded RNA. Thus, for small RNAs that are initially produced as duplexes, one strand must be chosen and the other discarded — a process called RISC loading. Strand selection must not be random. For example, for most miRNAs, evolutionary pressure has honed one particular strand of the duplex as a crucial regulator and loading of the other strand, the miR*, would cause silencing of the wrong set of genes.

Even from the first mechanistic studies, it was clear that strand choice was partly encoded in the intrinsic structure of the small RNA duplex, and a major determinant resides in its thermodynamic properties[104,105]. For both miRNAs and siRNAs in flies and mammals, the strand with the least stable 5′ end is more often retained. There are also additional favourable sequence characteristics, such as a bias for a U at position 1 (see BOX 2 for further details)[73,76,106–110]. Recently, our understanding of small RNA-sorting determinants has expanded substantially, and Argonaute and RNA structural studies have begun to provide a mechanistic basis for observations from *in vitro* and *in vivo* analyses[90,106,107,110–112].

Small RNA sorting in animals. In mammals, a single Dicer assorts siRNAs and miRNAs among four Argonaute subfamily proteins, apparently without much discrimination. However, in *D. melanogaster*, two distinct Dicer proteins process small RNA duplexes that preferentially enter AGO1 or AGO2 complexes. Generally, AGO1 is occupied by miRNAs, whereas AGO2 associates with siRNAs. This parallels the processing of miRNAs by Dicer 1 and siRNAs by Dicer 2. However, there are exceptions to the rule. For example, there are Dicer 1-derived small RNAs that preferentially load AGO2, implying the existence of a post-processing sorting mechanism[107,113]. Although miRNA and siRNA processing intermediates are approximately 19–21-nt duplexes with 2-nt 3′ overhangs, the character of their duplexed portions substantially differs (FIG. 3a). siRNAs are derived from duplexes featuring perfect or nearly perfect dsRNA, whereas miRNAs originate from precursors that typically contain several mismatches or bulges. Other features that affect sorting include the terminal nucleotides and thermodynamic properties of the duplex ends.

The numerous inputs into the sorting decisions of small RNAs have posed a challenge to predicting their fates in *D. melanogaster*. However, recent studies have suggested the application of hierarchical rules to predict differential AGO loading[106,107,110]. At the top level is duplex structure, specifically its degree of base pairing. Small RNA strands with unpaired central regions (~nucleotides 9–10) tend to be directed into AGO1 and disfavoured for AGO2 loading. Although *D. melanogaster* AGO1 and AGO2 show different preferences for terminal nucleotides (AGO1 favours a terminal U, whereas AGO2 shows a bias towards a 5′ C)[76,106,107], the identity of the 5′ nucleotide only makes a minor contribution to sorting[107]. For perfect duplexes, thermodynamic asymmetry dominates strand choice, which is precisely as was originally proposed[104,105,107].

It should be noted that sorting is a strand-centric process. Once a duplex is made, it seems that one strand is assessed and its fate determined. Thus, for many miRNAs, miR strands are abundant in AGO1 complexes and miR* strands predominate in AGO2–RISC; however, these miR and miR* strands arise from independent precursor molecules rather than through the stabilization of both strands of a given duplex. Thus, for each processed duplex, the choice seems to be whether the miR strand becomes committed to AGO1 or the miR* is committed to AGO2, with the complementary strand of each miRNA duplex being discarded during RISC maturation. Thus, AGO1 and AGO2 may compete for the selection of strands from each duplex, with the strength of preferential loading signals determining the ultimate abundance of the miR and miR* in AGO1 and AGO2 complexes, respectively[106,107,110].

It was recently noted in *D. melanogaster* that some hairpin-derived endo-siRNAs accumulate in AGO2 even though they originate from mismatched duplexes and have a terminal U — features which are thought to direct them towards AGO1 (REF. 114). Interestingly, *in vitro*, these small RNAs are sorted into AGO1. *In vivo*, however, these AGO1-loaded endo-siRNAs silence targets with high sequence complementarity. This paradox can be resolved by invoking target-directed small RNA destruction; small RNAs of this sort may be loaded into AGO1 *in vivo*, but they are unstable owing to lack of the stabilizing 2′-*O*-methylation, which they acquire when loaded into AGO2.

As in *D. melanogaster*, worm miRNAs and siRNAs are partitioned among distinct AGO subfamily proteins. Although worm sorting rules have not been probed in detail, miRNAs show a tendency towards central mismatches and are sorted into ALG-1 or ALG-2, whereas siRNAs from perfect duplexes preferentially load RDE-1 (REFS 115,116). In contrast to flies and worms, individual mammalian AGO clade proteins show no specialized structural and 5′ nucleotide preferences for small RNA duplexes[117–119]. This raises the possibility that mammals lack a strict system for small RNA sorting, at least among their AGO subfamily members.

Sorting of small RNAs in plants. *A. thaliana* encodes ten Argonaute proteins, which vary in their degrees of specialization and expression patterns. As in animals, plant AGO proteins tend to show preferences for distinct small RNA classes, which are produced through somewhat compartmentalized biogenesis pathways. For example, AGO1 is manly occupied by miRNAs that arise through processing by DCL1. AGO4 prefers hc-RNAs that are processed by DCL3. AGO2 is the principal recipient for ta-siRNAs. An additional complexity is that different Dicers produce small RNAs of distinct sizes. Plant DCL1 and DCL4 produce 21-nt RNAs, DCL2 22-nt RNAs and DCL3 24-nt RNAs. Different Dicer proteins have also been proposed to reside in different subcellular compartments. Thus, a wide range of properties might be exploited to establish specificity in plant small RNA sorting. Surprisingly, although the terminal nucleotide of the siRNA had a minor effect on sorting in flies and mammals, it strongly impacts sorting in plants.

Heterochromatic siRNA
A highly abundant plant small RNA that arises from transposons and repeats. hc-siRNAs depend on DCL3 and mainly load into AGO4.

RNase H
A conserved family of endonucleases that cleave the RNA strand of RNA:DNA hybrid duplexes. AGO proteins contain RNase H-like domains.

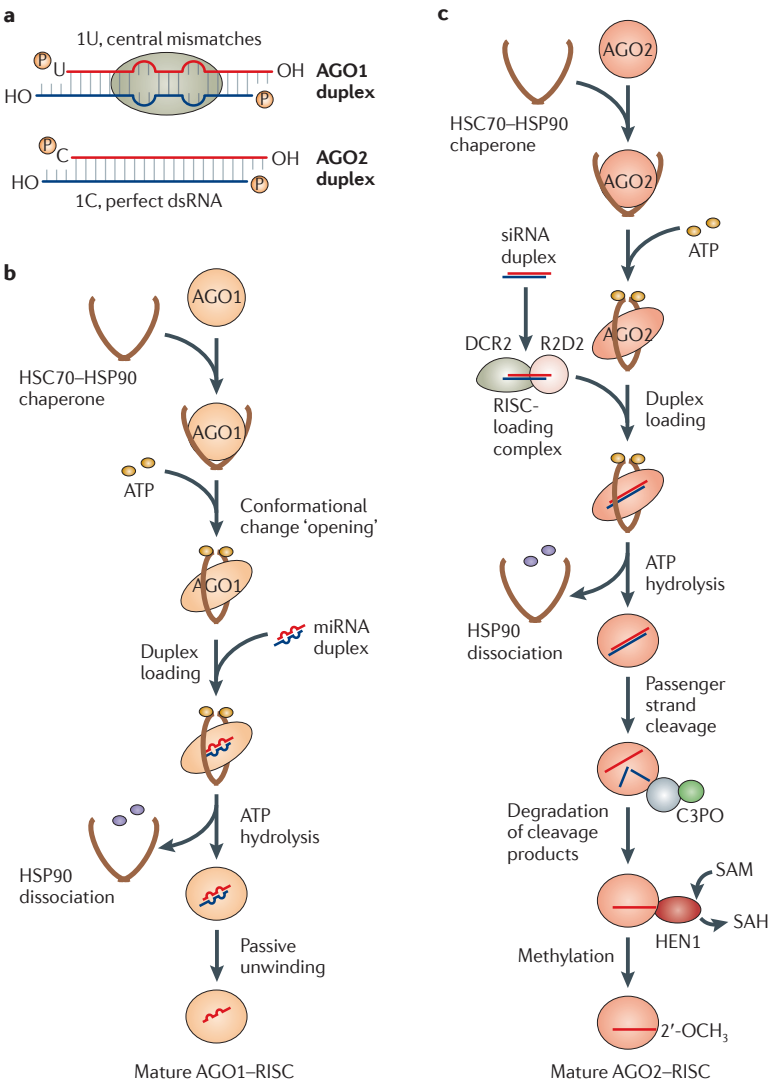


Figure 3 | **Small RNA sorting and RNA-induced silencing complex assembly in flies.** **a** | Structural determinants dominate the decision to sort small RNAs into fly Argonaute 1 (AGO1) or AGO2. AGO1-biased (usually microRNA (miRNA)) duplexes contain several bulges and mismatches, especially in the central region of the duplex. Mature miR strands show a strong bias for a terminal U. By contrast, AGO2-biased (usually siRNA) duplexes show extensive base pairing. Loaded guide strands often start with C. **b** | Unloaded AGO1 is recognized and bound by the heat shock cognate 70 (HSC70)–heat shock protein 90 (HSP90) chaperone complex and, following binding of ATP, adopts an 'open' conformational state. Loading-competent AGO1 receives miRNA duplexes containing several mismatches. The incorporation of duplexes into AGO1 is likely aided by as yet unidentified loading factors. ATP hydrolysis results in dissociation of the chaperone complex from AGO1, followed by passive unwinding of the duplex, a process promoted by mismatches. The miR* strand is degraded following unwinding. **c** | The HSC70–HSP90 chaperone complex associates with unloaded AGO2. Binding of ATP to the chaperone complex leads to conformational changes that allow AGO2 to receive small duplexes from the AGO2-loading machinery. Small RNA duplexes with perfect or near-perfect base-pairing (especially those with good pairing in the central region) are recognized by Dicer 2 (DCR2) and its co-factor R2D2 (AGO2–RISC-loading machinery) and inserted into AGO2. The chaperone complex dissociates following ATP hydrolysis, causing a change in the conformation of AGO2. Following passenger strand slicing by AGO2, component 3 promoter of RISC (C3PO) degrades the cleavage products. Subsequently, the 3′ terminus of the guide strand is methylated by HUA ENHANCER 1 (HEN1) to yield mature AGO2–RISC. SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine.

Deep sequencing of small RNAs associated with AGO family members clearly indicated that distinct AGO proteins preferentially load small RNAs with specific 5′ nucleotides[90,112,120,121]. AGO1 showed a strong bias towards a terminal U. AGO2 and AGO4 selected sequences that begin with an A, and AGO5 mainly bound RNAs starting with a 5′ C. Simply changing the terminal nucleotides could redirect small RNAs into different complexes in a predictable manner, strongly supporting the dominance of this sorting signal.

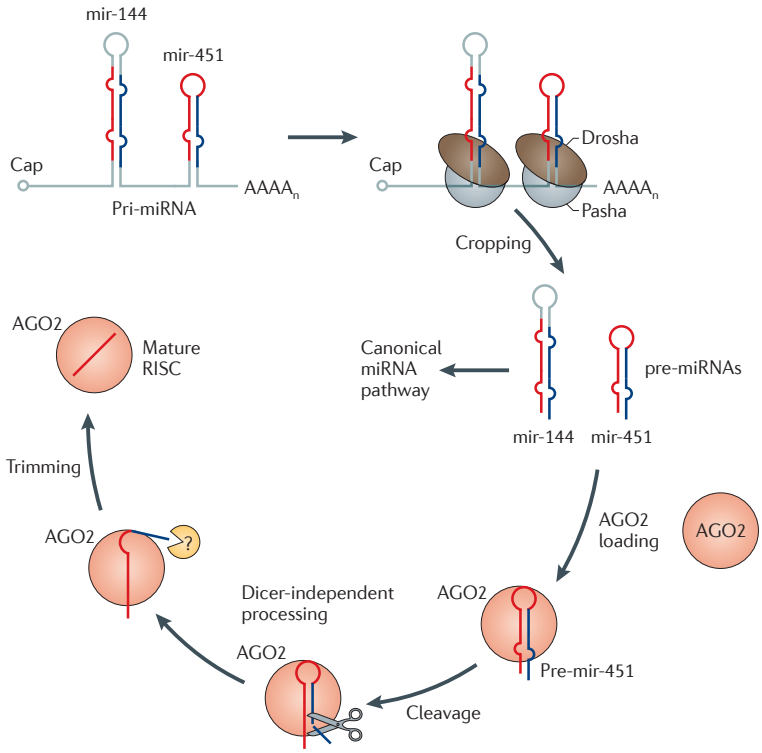
There were exceptions to the simple rule proposed above. MiR390, which begins with an A, would be predicted to load AGO2 but, instead, exclusively occupied AGO7 (REF. 90). Moreover, miR390 could not be redirected by altering its terminal base. Thus, although base recognition contributes strongly to sorting, other characteristics of small RNAs must also be taken into account. These could include duplex properties, such as thermodynamic asymmetry or degree of base pairing, although this hypothesis has yet to be examined. Overall, the data support a model in which plant small RNAs dissociate following Dicer cleavage and are subject to a sorting process, which surveys their terminal base. Other considerations, their size and the Dicer that produced them may contribute to specificity in a manner that varies with the small RNA species, but which becomes the dominant determinant of sorting in a few instances.

Sorting of other small RNA classes. To date, we know far more about the loading determinants of miRNAs and siRNAs than of any other small RNA class. Even within these well-studied groups, there are exceptions to the rules outlined above. For example, several reports now support the idea that pre-miRNA hairpins can be successfully loaded into RISC[118,122–126] (BOX 3). Mirtrons bypass the Drosha step but are presumably loaded using the normal miRNA strand determinants following Dicer cleavage.

Several small RNA classes are formed without a double-stranded precursor. Even though this should pose a simpler sorting problem, with no need to discriminate guide versus passenger strands, we know little about how these species are selectively loaded into specific Argonautes. Among good examples are the secondary siRNAs in worms, which are generated as direct RdRP products, presumably without the need for further processing[6,7]. These are specifically loaded into WAGO clade Argonautes through a still mysterious mechanism[9]. One could easily imagine that biogenesis and loading could be tightly coupled, or that the 5′ triphosphate termini on these small RNAs could contribute to binding specificity through interactions with the mid-domain of the Argonaute, but these ideas remain to be tested.

piRNAs, including worm 21U RNAs, do not depend on Dicer processing and are thought to originate from single-stranded precursors[4,5,8,127,128]. The loading of these small RNAs into Piwi subfamily proteins and the requirements of associated partner proteins for proper Piwi–RISC assembly are unknown. Whether the striking bias for a terminal U seen in many piRNAs reflects

Box 3 | **Non-canonical biogenesis and loading of small RNAs**



It had been reported in the literature that precursor-microRNA (pre-miRNA) hairpins are sometimes directly loaded into RNA-induced silencing complex (RISC) instead of being funnelled into the canonical Dicer-dependent biogenesis pathway[118,125,126]. Recently, it was shown that this strategy is actually used as a biogenesis mechanism by a conserved vertebrate miRNA, miR-451 (REFS 122–124). Like other endogenous miRNAs, mir-451 is synthesized by RNA polymerase II (RNAPII) as a polycistronic transcript together with mir-144 (see the figure above). This primary miRNA (pri-miRNA) is initially processed by the Microprocessor (Drosha–Pasha complex) through the canonical biogenesis pathway. However, following export to the cytoplasm, the two pre-miRNAs adopt distinct fates. Although pre-mir-144 continues along the canonical miRNA path and is processed by Dicer, pre-mir-451 is not a Dicer substrate, perhaps because its 17-nucleotide (nt)-duplexed region is too short. Instead, the pre-mir-451 hairpin is directly loaded into Argonaute 2 (AGO2). There, the duplexed portion of the hairpin is cleaved by the Argonaute RNase H-like motif and the cleaved product is resected by an unknown activity to form mature miR-451. Although it is unclear whether pre-mir-451 is actively sorted into AGO2, only those species which occupy this catalytically competent AGO family member can mature.

As a second example, the pre-miRNA equivalents for mirtrons are formed by the splicing machinery rather than by Drosha. Their biogenesis is outlined in FIG. 1.

upstream processing activities or is a consequence of the nucleotide-binding preferences of these Piwi proteins (as is seen in plants) remains unclear.

The RISC-loading machinery

Small RNA duplexes cannot be efficiently incorporated into AGO proteins without assistance from additional proteins[118,119]. These factors are also known as the RISC-loading machinery (or pre-RISC) and their precise nature differs for distinct AGO proteins. RISC loading is an active process that requires ATP[118,129–132], probably owing to the necessity to drive conformational changes so that AGO proteins accept small RNA duplexes. This concept, which was originally suggested based on structural analyses of AGO proteins, has gained recent support from studies that characterized interactions between Argonautes and the heat shock cognate 70 (HSC70)–heat shock protein 90 (HSP90) chaperone complex[133–136]. These studies support a model in which the interaction between Argonautes and the chaperone complex creates an 'open' conformation that is suitable for the loading of duplexed small RNAs. ATP hydrolysis and dissociation of the chaperone results in a structure that can discard or cleave the miR* or passenger strand to form an active RISC.

In flies, the loading machinery for AGO2–RISC also involves Dicer 2 and its dsRBD partner R2D2 (REFS 68,69,130,131,137,138) (FIG. 3c). In fact, these factors have been proposed to be the biochemical sensors for thermodynamic asymmetry. In this regard, R2D2 has been shown to bind the more stable end of the dsRNA duplex, whereas Dicer 2 is positioned at the less-stable end of the duplex, providing a mechanism for orientated AGO2 loading[139]. Although a minimal pre-RISC could be constituted with only Dicer 2, R2D2 and AGO2 (REF. 140), the bona fide AGO2–RISC-loading machinery *in vivo* undoubtedly contains additional components, including the chaperone complexes described above. Roles for Dicers have also been suggested for AGO1 loading. One report suggests that AGO1–Dicer 1 complexes correspond to the AGO1–RISC-loading complex[141], whereas a second report indicated that Dicer 1 was dispensable for AGO1–RISC assembly[132].

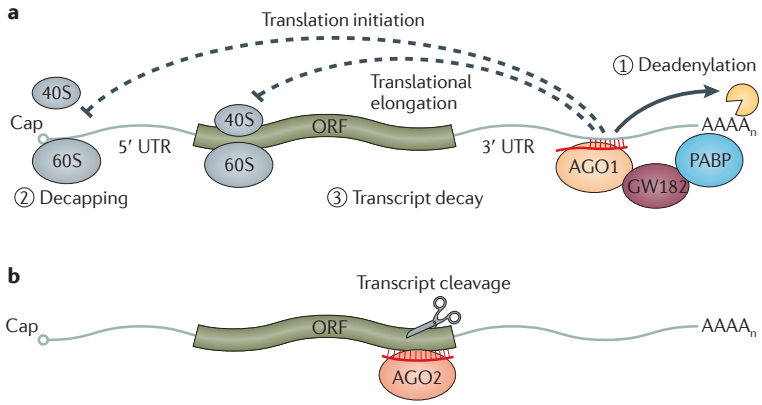
Although little is known about the loading machinery in plants, a recent study proposed that the thermodynamic properties of duplex ends (instead of terminal nucleotides) are the dominant determinant for strand selection of some DCL1-processed miRNAs and that HYL1, like fly R2D2, functioned as a component of the asymmetry sensor[66].

RISC maturation

For RISC to exert its function, pre-RISC needs to mature (FIG. 3c). Although the orientation of the miRNA duplex was determined during RISC assembly and loading, the crucial maturation step is discarding of the passenger or miR* strand. In flies and mammals, distinct AGO proteins seem to achieve this by different mechanisms, which depend on the nature of the AGO protein and the degree of base pairing in the loaded duplex. Using their 'slicer' activity, fly or mammalian AGO2 can cleave the passenger strand of perfect or nearly perfect duplexes[12–14]. The cleaved strand dissociates from RISC and, in flies, is degraded by a multimeric endonuclease complex (consisting of Translin and Trax), termed C3PO (component 3 promoter of RISC)[140]. Following passenger strand removal, AGO2-bound single-stranded small RNAs are methylated at their 3′ termini by the methyltransferase HEN1 (also known as Pimet) to yield mature AGO2–RISC[142,143].

Maturation of miRNA RISC is less well understood (FIG. 3b, bottom). Human AGO1, AGO3 and AGO4 all lack slicer activity and fly AGO1 is a poor slicer.

Box 4 | **Mechanisms of target regulation in *Drosophila melanogaster***



Individual Argonaute (AGO) proteins differ in their expression patterns, subcellular localization and enzymatic properties. Thus, distinct AGOs can function through many different effector modes that may involve slicing of target transcripts, cleavage-independent regulation and chromatin modification (reviewed in REFS 15–17). Another layer of complexity is added by the degree of sequence complementarity between the AGO-bound small RNA and target transcripts, which determines the mechanism of regulation. **a** | In flies, AGO1-associated microRNAs (miRNAs) typically target mRNAs in their 3′ UTRs to reduce protein synthesis. Owing to limited sequence complementarity between the small RNA (seed region) and the mRNA, such interactions usually do not result in direct cleavage of the targeted transcript. Instead, AGO1 and its partner protein GW182 are likely to disrupt crucial interactions between the polyA tail and the cap of the transcript, leading to a reduction in translational initiation and an induction of mRNA decay[173]. In mammals, it was recently shown that reduced protein output is predominantly owing to destabilization of the target transcript[174]. **b** | Small RNAs bound to *Drosophila melanogaster* AGO2 do not exhibit a bias towards binding their targets in the 3′ UTR. AGO2 primed with a small RNA sharing extensive complementarity with its target typically directs endonucleolytic cleavage of the mRNAs through AGO2 slicer activity. The 2-*O*-methyl modification of AGO2-bound small RNAs prevents their degradation when targeting perfectly complementary transcripts[64,142,143]. However, other modes are possible: AGO2 can also regulate targets with limited sequence complementarity through a block in translation initiation (not shown)[173]. PABP, poly(A)-binding protein

Moreover, miRNA duplexes often contain sufficient bulges to prevent slicing of miR* strands even by competent enzymes. Therefore, it has been proposed that miR* strands dissociate in a cleavage-independent manner by unwinding — a process that is facilitated by the presence of mismatches in the loaded duplexes[113,118,132]. Biochemical evidence supports unwinding as a passive, ATP-independent process, with degradation of the miR* strand on its release. It is unclear how plant Argonautes remove the miR* or passenger strand during RISC maturation. MiR* and passenger strands could be cleaved through the slicer activity of AGOs (similar to fly AGO2) or unwound passively (like fly AGO1)[12–14,132].

The impact of sorting on target regulation

The ultimate result of accurate strand selection and sorting is that an active RISC is formed, which is imbued with the ability to regulate a target gene or process. Argonaute family members differ in their biochemical properties, subcellular localization and expression patterns, and matching the right small RNA with the correct partner is key to proper biological function.

Seed region
A region consisting of nucleotides 2–8 counted from the 5′ end of miRNAs that participates in the interaction between a small RNA and target transcript.

Although AGO proteins evolved as ribonucleases, animal miRNAs affect their targets without the need for this activity. miRNAs generally interact with their targets through limited base-pairing interactions that are insufficient to place the scissile phosphate of the target in the enzyme active site where cleavage can occur. The prevalence of cleavage-independent repression modes is also reflected in the diversity of the Argonaute family. In mammals, three of the four AGO proteins have lost catalytic potential, and AGO1, the *D. melanogaster* AGO protein into which most miRNAs are sorted, is a poor enzyme compared with its siRNA-binding cousin[113].

miRNA-directed target cleavage has only been reported in a few cases[144,145]. However, this is assumed to be the principal regulatory mode for endo-siRNAs and for piRNA-mediated repression of transposons. Here again, the choice of a particular AGO partner is crucial. Piwi family members all retain catalytic competence and *D. melanogaster* AGO2, the main partner for endogenous and viral siRNAs, is tuned for highly active slicing (BOX 4).

AGO1-associated plant miRNAs usually share extensive sequence complementary with their mRNA targets and these interactions often result in target cleavage[146]. However, recent studies have indicated that cleavage-independent translational repression is widespread in plants, even for highly complementary target sites[147]. Nevertheless, miRNA-mediated cleavage is of key importance for some processes like the biogenesis of ta-siRNAs, for which the initial slicing event is key to RdRP recruitment and dsRNA synthesis[88].

Notably, small RNAs that direct cleavage, for example, plant miRNAs, piRNAs and fly endo-siRNAs, often have a 2′-*O*-methyl modification on their 3′ termini. Although the purpose of this modification was initially mysterious, it is now clear that this functions as a protective group to prevent small RNA destruction[64,142,143]. In flies and mammals, small RNAs that have extensive complementarity to their targets can be recognized by terminal uridyl transferases, which mark small RNAs for degradation[64]. The uridylation event is blocked by the 2′-*O*-methyl modification, preserving these small RNAs, which have evolved to function through cleavage[64]. The balance between protection and targeted destruction has been proposed as a quality control on small RNA sorting and as an evolutionary mechanism to drive animal miRNAs toward a cleavage-independent repression mode[64].

hc-siRNAs are thought to function by different mechanisms[148,149]. They must be sorted into a particular Argonaute, AGO4, which they guide to target DNA loci by base pairing with nascent non-coding transcripts synthesized by RNAPV. Effector proteins, such as the chromatin-remodelling factor DRD1, the *de novo* methyltransferase DRM2 and other factors, are then recruited, resulting in DNA methylation at cytosine residues[150,151]. As this regulation functions by repressing RNA synthesis, it was termed transcriptional gene silencing to distinguish it from post-transcriptional gene-silencing modes. Some piRNAs in flies and mammals must associate with particular Piwi-family proteins — that is, PIWI and

MIWI2, respectively — which enable these small RNAs to enter the nucleus, where they are thought to induce transcriptional repression through changes in chromatin structure or DNA methylation, respectively[152–154]. Similarly, worm NRDE-3, an Argonaute of the WAGO clade, transports siRNAs to the nucleus and functions through co-transcriptional gene silencing[155].

Thus, the final effects of small RNA sorting are felt in the modes of repression that become available as they join specific AGO proteins. The consequences of improper sorting may range from a loss of target regulation to inappropriate regulatory modes.

Conclusions

An understanding of the mechanisms by which small RNAs are selected and sorted among different potential effector complexes is crucial. In part, this knowledge guides hypotheses concerning the cellular roles of an ever-growing roster of small RNA species. However, the ability to predict the fate of small RNAs based on their sequence and structural characteristics is also essential to their effective use as experimental tools and potential therapeutics. We have begun to piece together the properties that determine small RNA fates and, in some instances, these properties can even predict with reasonable accuracy which small RNAs will efficiently join a particular effector complex. Yet, we still have a relatively poor ability to design effective small RNAs *ab initio* for experimental or therapeutic use. This capacity will rest on advances in both our understanding of RISC as an enzyme, including its mechanisms of target recognition, silencing and product release, and a detailed knowledge of how specific RNA strands are efficiently loaded into RISC as guides.

1. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nature Rev. Genet.* **10**, 94–108 (2009).
2. Chapman, E. J. & Carrington, J. C. Specialization and evolution of endogenous small RNA pathways. *Nature Rev. Genet.* **8**, 884–896 (2007).
3. Voinnet, O. Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**, 669–687 (2009).
4. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
5. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
6. Pak, J. & Fire, A. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* **315**, 241–244 (2007).
7. Sijen, T., Steiner, F. A., Thijssen, K. L. & Plasterk, R. H. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* **315**, 244–247 (2007).
8. Houwing, S. *et al.* A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell* **129**, 69–82 (2007).
9. Yigit, E. *et al.* Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**, 747–757 (2006).
10. Tolia, N. H. & Joshua-Tor, L. Slicer and the Argonautes. *Nature Chem. Biol.* **3**, 36–43 (2007).
11. Hutvagner, G. & Simard, M. J. Argonaute proteins: key players in RNA silencing. *Nature Rev. Mol. Cell Biol.* **9**, 22–32 (2008).
12. Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. & Siomi, M. C. Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev.* **19**, 2837–2848 (2005).
13. Rand, T. A., Petersen, S., Du, F. & Wang, X. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**, 621–629 (2005).
14. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P. & Zamore, P. D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620 (2005).
References 12–14 described passenger strand cleavage and showed that Argonaute proteins themselves function in RISC maturation.
15. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* **79**, 351–379 (2010).
16. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
17. Brodersen, P. & Voinnet, O. Revisiting the principles of microRNA target recognition and mode of action. *Nature Rev. Mol. Cell Biol.* **10**, 141–148 (2009).
18. Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L. & Bradley, A. Identification of mammalian microRNA host genes and transcription units. *Genome Res.* **14**, 1902–1910 (2004).
19. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).
20. Cai, X., Hagedorn, C. H. & Cullen, B. R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957–1966 (2004).
21. Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235 (2004).
22. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).
23. Han, J. *et al.* The Drosha–DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–3027 (2004).
24. Landthaler, M., Yalcin, A. & Tuschl, T. The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr. Biol.* **14**, 2162–2167 (2004).
25. Han, J. *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha–DGCR8 complex. *Cell* **125**, 887–901 (2006).
26. Bohnsack, M. T., Czaplinski, K. & Gorlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191 (2004).
27. Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95–98 (2004).
28. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).
29. Bussing, I., Yang, J. S., Lai, E. C. & Grosshans, H. The nuclear export receptor XPO-1 supports primary miRNA processing in *C. elegans* and *Drosophila*. *EMBO J.* **29**, 1830–1839 (2010).
30. Gruber, J. J. *et al.* Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell* **138**, 328–339 (2009).
31. Sabin, L. R. *et al.* Ars2 regulates both miRNA- and siRNA-dependent silencing and suppresses RNA virus infection in *Drosophila*. *Cell* **138**, 340–351 (2009).
32. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
This paper identified the ribonuclease Dicer as the dsRNA processing enzyme.
33. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
34. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838 (2001).
35. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
36. Haase, A. D. *et al.* TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.* **6**, 961–967 (2005).
37. Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744 (2005).
38. Saito, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Processing of pre-microRNAs by the Dicer-1–Loquacious complex in *Drosophila* cells. *PLoS Biol.* **3**, e235 (2005).
39. Jiang, F. *et al.* Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* **19**, 1674–1679 (2005).
40. Forstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
41. Park, J. K., Liu, X., Strauss, T. J., McKearin, D. M. & Liu, Q. The miRNA pathway intrinsically controls self-renewal of *Drosophila* germline stem cells. *Curr. Biol.* **17**, 533–538 (2007).
42. Zhou, R. *et al.* Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA* **15**, 1886–1895 (2009).
43. Berezikov, E., Chung, W. J., Willis, J., Cuppen, E. & Lai, E. C. Mammalian mirtron genes. *Mol. Cell* **28**, 328–336 (2007).
44. Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. & Lai, E. C. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130**, 89–100 (2007).
45. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83–86 (2007).
46. Flynt, A. S., Greimann, J. C., Chung, W. J., Lima, C. D. & Lai, E. C. MicroRNA biogenesis via splicing and exosome-mediated trimming in *Drosophila*. *Mol. Cell* **38**, 900–907 (2010).
47. Yu, B. *et al.* The FHA domain proteins DAWDLE in *Arabidopsis* and SNIP1 in humans act in small RNA biogenesis. *Proc. Natl Acad. Sci. USA* **105**, 10073–10078 (2008).
48. Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. & Bartel, D. P. MicroRNAs in plants. *Genes Dev.* **16**, 1616–1626 (2002).
49. Park, W., Li, J., Song, R., Messing, J. & Chen, X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484–1495 (2002).
50. Henderson, I. R. *et al.* Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nature Genet.* **38**, 721–725 (2006).
51. Kurihara, Y. & Watanabe, Y. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl Acad. Sci. USA* **101**, 12753–12758 (2004).
52. Han, M. H., Goud, S., Song, L. & Fedoroff, N. The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl Acad. Sci. USA* **101**, 1093–1098 (2004).
53. Kurihara, Y., Takashi, Y. & Watanabe, Y. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* **12**, 206–212 (2006).

54. Vazquez, F., Gasciolli, V., Crete, P. & Vaucheret, H. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr. Biol.* **14**, 346–351 (2004).
55. Dong, Z., Han, M. H. & Fedoroff, N. The RNA-binding proteins HYL1 and SE promote accurate *in vitro* processing of pri-miRNA by DCL1. *Proc. Natl Acad. Sci. USA* **105**, 9970–9975 (2008).
56. Lobbes, D., Rallapalli, G., Schmidt, D. D., Martin, C. & Clarke, J. SERRATE: a new player on the plant microRNA scene. *EMBO Rep.* **7**, 1052–1058 (2006).
57. Yang, L., Liu, Z., Lu, F., Dong, A. & Huang, H. SERRATE is a novel nuclear regulator in primary microRNA processing in *Arabidopsis*. *Plant J.* **47**, 841–850 (2006).
58. Laubinger, S. *et al.* Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **105**, 8795–8800 (2008).
59. Gregory, B. D. *et al.* A link between RNA metabolism and silencing affecting *Arabidopsis* development. *Dev. Cell* **14**, 854–866 (2008).
60. Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in *Arabidopsis*. *Curr. Biol.* **15**, 1501–1507 (2005).
61. Yang, Z., Ebright, Y. W., Yu, B. & Chen, X. HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 2′ OH of the 3′ terminal nucleotide. *Nucleic Acids Res.* **34**, 667–675 (2006).
62. Yu, B. *et al.* Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932–935 (2005).
63. Ramachandran, V. & Chen, X. Degradation of microRNAs by a family of exoribonucleases in *Arabidopsis*. *Science* **321**, 1490–1492 (2008).
64. Ameres, S. L. *et al.* Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**, 1534–1539 (2010).
65. Park, M. Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H. & Poethig, R. S. Nuclear processing and export of microRNAs in *Arabidopsis*. *Proc. Natl Acad. Sci.* USA **102**, 3691–3696 (2005).
66. Eamens, A. L., Smith, N. A., Curtin, S. J., Wang, M. B. & Waterhouse, P. M. The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA* **15**, 2219–2235 (2009).
67. Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
This seminal report was the first to link small RNAs to post-transcriptional gene silencing.
68. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
69. Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003).
70. Hartig, J. V., Esslinger, S., Bottcher, R., Saito, K. & Forstemann, K. Endo-siRNAs depend on a new isoform of Loquacious and target artificially introduced, high-copy sequences. *EMBO J.* **28**, 2932–2944 (2009).
71. Miyoshi, K., Miyoshi, T., Hartig, J. V., Siomi, H. & Siomi, M. C. Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in *Drosophila*. *RNA* **16**, 506–515 (2010).
72. Marques, J. T. *et al.* Loqs and R2D2 act sequentially in the siRNA pathway in *Drosophila*. *Nature Struct. Mol. Biol.* **17**, 24–30 (2010).
73. Ghildiyal, M. *et al.* Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320**, 1077–1081 (2008).
74. Okamura, K. *et al.* The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453**, 803–806 (2008).
75. Kawamura, Y. *et al. Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**, 793–797 (2008).
76. Czech, B. *et al.* An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**, 798–802 (2008).
77. Chung, W. J., Okamura, K., Martin, R. & Lai, E. C. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr. Biol.* **18**, 795–802 (2008).
78. Lipardi, C. & Paterson, B. M. Identification of an RNA-dependent RNA polymerase in *Drosophila* involved in RNAi and transposon suppression. *Proc. Natl Acad. Sci. USA* **106**, 15645–15650 (2009).
79. Tam, O. H. *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534–538 (2008).
80. Watanabe, T. *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**, 539–543 (2008).
81. Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* **22**, 2773–2785 (2008).
82. Knight, S. W. & Bass, B. L. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**, 2269–2271 (2001).
83. Steiner, F. A., Okihara, K. L., Hoogstrate, S. W., Sijen, T. & Ketting, R. F. RDE-1 slicer activity is required only for passenger-strand cleavage during RNAi in *Caenorhabditis elegans*. *Nature Struct. Mol. Biol.* **16**, 207–211 (2009).
84. Vazquez, F. *et al.* Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* **16**, 69–79 (2004).
85. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L. & Poethig, R. S. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* **18**, 2368–2379 (2004).
86. Yoshikawa, M., Peragine, A., Park, M. Y. & Poethig, R. S. A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* **19**, 2164–2175 (2005).
87. Williams, L., Carles, C. C., Osmont, K. S. & Fletcher, J. C. A database analysis method identifies an endogenous *trans*-acting short-interfering RNA that targets the *Arabidopsis ARF2, ARF3*, and *ARF4* genes. *Proc. Natl Acad. Sci. USA* **102**, 9703–9708 (2005).
88. Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207–221 (2005).
89. Axtell, M. J., Jan, C., Rajagopalan, R. & Bartel, D. P. A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**, 565–577 (2006).
90. Montgomery, T. A. *et al.* Specificity of ARGONAUTE7–miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell* **133**, 128–141 (2008).
91. Hernandez-Pinzon, I. *et al.* SDE5, the putative homologue of a human mRNA export factor, is required for transgene silencing and accumulation of *trans*-acting endogenous siRNA. *Plant J.* **50**, 140–148 (2007).
92. Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R. & Zhu, J. K. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123**, 1279–1291 (2005).
93. Katiyar-Agarwal, S. *et al.* A pathogen-inducible endogenous siRNA in plant immunity. *Proc. Natl Acad. Sci. USA* **103**, 18002–18007 (2006).
94. Henz, S. R. *et al.* Distinct expression patterns of natural antisense transcripts in *Arabidopsis*. *Plant Physiol.* **144**, 1247–1255 (2007).
95. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, e104 (2004).
96. Chan, S. W. *et al.* RNA silencing genes control *de novo* DNA methylation. *Science* **303**, 1336 (2004).
97. Onodera, Y. *et al.* Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613–622 (2005).
98. Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118–120 (2005).
99. Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* **19**, 2030–2040 (2005).
100. Kasschau, K. D. *et al.* Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol.* **5**, e57 (2007).
101. Smith, L. M. *et al.* An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in *Arabidopsis*. *Plant Cell* **19**, 1507–1521 (2007).
102. Gregory, R. I., Chendrimada, T. P., Cooch, N. & Shiekhattar, R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123**, 631–640 (2005).
103. Maniataki, E. & Mourelatos, Z. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.* **19**, 2979–2990 (2005).
104. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
105. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
References 104 and 105 identified intrinsic determinants in small RNA duplexes that affect their sorting into Argonaute complexes.
106. Ghildiyal, M., Xu, J., Seitz, H., Weng, Z. & Zamore, P. D. Sorting of *Drosophila* small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA* **16**, 43–56 (2010).
107. Czech, B. *et al.* Hierarchical rules for Argonaute loading in *Drosophila*. *Mol. Cell* **36**, 445–456 (2009).
108. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
109. Hu, H. Y. *et al.* Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* **10**, 413 (2009).
110. Okamura, K., Liu, N. & Lai, E. C. Distinct mechanisms for microRNA strand selection by *Drosophila* Argonautes. *Mol. Cell* **36**, 431–444 (2009).
111. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822 (2010).
This recent study found structural evidence in Argonaute proteins that explains 5′ terminal nucleotide biases.
112. Mi, S. *et al.* Sorting of small RNAs into *Arabidopsis* Argonaute complexes is directed by the 5′ terminal nucleotide. *Cell* **133**, 116–127 (2008).
Together with reference 90, this paper described the identification of 5′ terminal nucleotides as sorting determinants in plants.
113. Forstemann, K., Horwich, M. D., Wee, L., Tomari, Y. & Zamore, P. D. *Drosophila* microRNAs are sorted into functionally distinct Argonaute complexes after production by Dicer-1. *Cell* **130**, 287–297 (2007).
114. Ameres, S. L., Hung, J.-H., Xu, J., Weng, Z. & Zamore, P. D. Target RNA-directed tailing and trimming purifies the sorting of endo-siRNAs between the two *Drosophila* Argonaute proteins. *RNA* (in the press).
115. Steiner, F. A. *et al.* Structural features of small RNA precursors determine Argonaute loading in *Caenorhabditis elegans*. *Nature Struct. Mol. Biol.* 14, 927–933 (2007).
116. Jannot, G., Boisvert, M. E., Banville, I. H. & Simard, M. J. Two molecular features contribute to the Argonaute specificity for the microRNA and RNAi pathways in *C. elegans*. *RNA* **14**, 829–835 (2008).
117. Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
118. Yoda, M. *et al.* ATP-dependent human RISC assembly pathways. *Nature Struct. Mol. Biol.* **17**, 17–23 (2010).
119. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
120. Wu, L. *et al.* Rice microRNA effector complexes and targets. *Plant Cell* **21**, 3421–3435 (2009).
121. Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M. & Watanabe, Y. The mechanism selecting the guide strand from small RNA duplexes is different among Argonaute proteins. *Plant Cell Physiol.* **49**, 493–500 (2008).
122. Yang, J. S. *et al.* Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proc. Natl Acad. Sci. USA* **107**, 15163–15168 (2010).
123. Cheloufi, S., Dos Santos, C. O., Chong, M. M. & Hannon, G. J. A Dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**, 584–589 (2010).
124. Cifuentes, D. *et al.* A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328**, 1694–1698 (2010).
125. Diederichs, S. & Haber, D. A. Dual role for Argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell* **131**, 1097–1108 (2007).
126. Tan, G. S. *et al.* Expanded RNA-binding activities of mammalian Argonaute 2. *Nucleic Acids Res.* **37**, 7533–7545 (2009).
127. Das, P. P. *et al.* Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol. Cell* **31**, 79–90 (2008).

128. Batista, P. J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol. Cell* **31**, 67–78 (2008).
129. Nykanen, A., Haley, B. & Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309–321 (2001).
130. Tomari, Y. *et al.* RISC assembly defects in the *Drosophila* RNAi mutant *armitage*. *Cell* **116**, 831–841 (2004).
131. Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W. & Sontheimer, E. J. A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell* **117**, 83–94 (2004).
132. Kawamata, T., Seitz, H. & Tomari, Y. Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nature Struct. Mol. Biol.* **16**, 953–960 (2009).
133. Johnston, M., Geoffroy, M. C., Sobala, A., Hay, R. & Hutvagner, G. HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. *Mol. Biol. Cell* **21**, 1462–1469 (2010).
134. Iki, T. *et al. In vitro* assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol. Cell* **39**, 282–291 (2010).
135. Iwasaki, S. *et al.* Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* **39**, 292–299 (2010).
136. Miyoshi, T., Takeuchi, A., Siomi, H. & Siomi, M. C. A direct role for Hsp90 in pre-RISC formation in *Drosophila*. *Nature Struct. Mol. Biol.* **17**, 1024–1026 (2010).
137. Tomari, Y., Du, T. & Zamore, P. D. Sorting of *Drosophila* small silencing RNAs. *Cell* **130**, 299–308 (2007). **References 113, 118, 132 and 137 made important contributions to understanding small RNA sorting and showed that miRNA maturation is independent of cleavage activity.**
138. Liu, X., Jiang, F., Kalidas, S., Smith, D. & Liu, Q. Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. *RNA* **12**, 1514–1520 (2006).
139. Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377–1380 (2004).
140. Liu, Y. *et al.* C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science* **325**, 750–753 (2009).
141. Miyoshi, K., Okada, T. N., Siomi, H. & Siomi, M. C. Characterization of the miRNA–RISC loading complex and miRNA-RISC formed in the *Drosophila* miRNA pathway. *RNA* **15**, 1282–1291 (2009).
142. Horwich, M. D. *et al.* The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272 (2007).
143. Saito, K., Sakaguchi, Y., Suzuki, T., Siomi, H. & Siomi, M. C. Pimet, the *Drosophila* homolog of HEN1, mediates 2′-O-methylation of Piwi- interacting RNAs at their 3′ ends. *Genes Dev.* **21**, 1603–1608 (2007).
144. Davis, E. *et al.* RNAi-mediated allelic *trans*-interaction at the imprinted *Rtl1/Peg11* locus. *Curr. Biol.* **15**, 743–749 (2005).
145. Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of *HOXB8* mRNA. *Science* **304**, 594–596 (2004).
146. Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–520 (2002).
147. Brodersen, P. *et al.* Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**, 1185–1190 (2008).
148. Wierzbicki, A. T., Ream, T. S., Haag, J. R. & Pikaard, C. S. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nature Genet.* **41**, 630–634 (2009).
149. Matzke, M., Kanno, T., Daxinger, L., Huettel, B. & Matzke, A. J. RNA-mediated chromatin-based silencing in plants. *Curr. Opin. Cell Biol.* **21**, 367–376 (2009).
150. Chan, S. W., Henderson, I. R. & Jacobsen, S. E. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Rev. Genet.* **6**, 351–360 (2005).
151. Mosher, R. A., Schwach, F., Studholme, D. & Baulcombe, D. C. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl Acad. Sci. USA* **105**, 3145–3150 (2008).
152. Yin, H. & Lin, H. An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*. *Nature* **450**, 304–308 (2007).
153. Carmell, M. A. *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* **12**, 503–514 (2007).
154. Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol. Cell* **31**, 785–799 (2008).
155. Guang, S. *et al.* Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* **465**, 1097–1101 (2010).
156. Werner, S., Wollmann, H., Schneeberger, K. & Weigel, D. Structure determinants for accurate processing of miR172a in *Arabidopsis thaliana*. *Curr. Biol.* **20**, 42–48 (2010).
157. Mateos, J. L., Bologna, N. G., Chorostecki, U. & Palatnik, J. F. Identification of microRNA processing determinants by random mutagenesis of *Arabidopsis* MIR172a precursor. *Curr. Biol.* **20**, 49–54 (2010).
158. Song, L., Axtell, M. J. & Fedoroff, N. V. RNA secondary structural determinants of miRNA precursor processing in *Arabidopsis*. *Curr. Biol.* **20**, 37–41 (2010).
159. Cuperus, J. T. *et al.* Identification of MIR390a precursor processing-defective mutants in *Arabidopsis* by direct genome sequencing. *Proc. Natl Acad. Sci. USA* **107**, 466–471 (2010).
160. Bologna, N. G., Mateos, J. L., Bresso, E. G. & Palatnik, J. F. A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J.* **28**, 3646–3656 (2009).
161. Wang, Y. *et al.* Structure of an Argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008).
162. Song, J. J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Struct. Mol. Biol.* **10**, 1026–1032 (2003).
163. Ma, J. B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).
164. Ma, J. B. *et al.* Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**, 666–670 (2005).
165. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain–siRNA guide complex. *Nature* **434**, 663–666 (2005).
166. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing Argonaute silencing complex. *Nature* **456**, 209–213 (2008).
167. Wang, Y. *et al.* Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**, 754–761 (2009).
168. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).
169. Parker, J. S., Roe, S. M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23**, 4727–4737 (2004).
170. Yuan, Y. R. *et al.* Crystal structure of *A. aeolicus* Argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell* **19**, 405–419 (2005).
171. Schwarz, D. S., Tomari, Y. & Zamore, P. D. The RNA-induced silencing complex is a Mg^{2+}-dependent endonuclease. *Curr. Biol.* **14**, 787–791 (2004).
172. Martinez, J. & Tuschl, T. RISC is a 5′ phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980 (2004).
173. Iwasaki, S., Kawamata, T. & Tomari, Y. *Drosophila* Argonaute1 and Argonaute2 employ distinct mechanisms for translational repression. *Mol. Cell* **34**, 58–67 (2009).
174. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).

Acknowledgements
The authors thank O. Tam, F. Muerdter, J.-W. Wang and R. Zhou for comments on the manuscript. The authors are greatly indebted to J. Duffy for assistance with figures. B.C. is supported by a Ph.D fellowship from the Boehringer Ingelheim Fonds. This work was supported by grants from the National Institutes of Health and a kind gift from K. W. Davis. G.J.H. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
miRBase: http://www.mirbase.org/index.shtml
mir-144 | miR173 | miR390 | miR-451
UniProtKB: http://www.uniprot.org
AGO1 | AGO2 | DCR1 | DCR2 | Drosha | HEN1 | LOQS-PB | R2D2

FURTHER INFORMATION
Gregory J. Hannon's homepage: http://hannonlab.cshl.edu/index.html
Gregory J. Hannon's Cold Spring Harbor Laboratory homepage: http://www.cshl.edu/Faculty/hannon-gregory.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

A genome-scale shRNA resource for transgenic RNAi in *Drosophila*

Jian-Quan Ni*, Rui Zhou*, **Benjamin Czech***, Lu-Ping Liu, Laura Holderbaum, Donghui Yang-Zhou, Hye-Seok Shim, Rong Tao, Dominik Handler, Phillip Karpowicz, Richard Binari, Matthew Booker, Julius Brennecke, Lizabeth A. Perkins, Gregory J. Hannon and Norbert Perrimon

* authors contributed equally to this work

Nat Methods. 2011 May;8(5):405-7. Epub 2011 Apr 3.

A genome-scale shRNA resource for transgenic RNAi in *Drosophila*

Jian-Quan Ni[1,5,6], Rui Zhou[1,5,6], Benjamin Czech[2,6], Lu-Ping Liu[1,5], Laura Holderbaum[1], Donghui Yang-Zhou[1], Hye-Seok Shim[1], Rong Tao[1], Dominik Handler[3], Phillip Karpowicz[1], Richard Binari[1], Matthew Booker[1], Julius Brennecke[3], Lizabeth A Perkins[4], Gregory J Hannon[2] & Norbert Perrimon[1]

Existing transgenic RNAi resources in *Drosophila melanogaster* based on long double-stranded hairpin RNAs are powerful tools for functional studies, but they are ineffective in gene knockdown during oogenesis, an important model system for the study of many biological questions. We show that shRNAs, modeled on an endogenous microRNA, are extremely effective at silencing gene expression during oogenesis. We also describe our progress toward building a genome-wide shRNA resource.

Current *Drosophila* transgenic RNAi resources use long hairpins as silencing triggers[1]. However, for reasons unknown, long hairpins are ineffective for gene silencing in the female germline, a conclusion that we reached after extensive testing of various construct designs (**Supplementary Fig. 1** and **Supplementary Note 1**). In *Drosophila*, RNAi can be triggered via distinct routes, each of which generates small silencing RNAs via discrete processing and loading machineries[2]. In particular, artificial microRNAs, referred to as shRNAs, have been shown to trigger effective silencing in somatic cells[3] and in the female germline in one case[4]. As shRNAs have not been used extensively and compared to long hairpins for their efficacies, we systematically evaluated their use as a transgenic trigger of RNAi in *Drosophila*.

We constructed Valium20, a vector that combines the optimized expression features of the previously reported Valium10 for somatic RNAi[5] with a modified scaffold of the microRNA miR-1 (**Fig. 1a** and **Supplementary Fig. 2**). Unique cloning sites allow the generation of shRNAs that accommodate the desired sequences, leading to a hairpin with perfect duplex structure, which favors shRNA loading into AGO2, the principal effector of RNAi in flies (**Fig. 1a**). To test the effectiveness of Valium20, we generated several fly lines containing shRNA constructs that target genes associated with either distinctive germline or maternal effect phenotypes. We induced shRNA expression specifically in the germline with *MTD-Gal4* (ref. 6), a line that carries three *Gal4* drivers expressed at various stages during oogenesis (Online Methods). For all examined lines, we recovered the expected oogenesis and maternal effect mutant phenotypes (**Supplementary Note 1**), indicating that shRNAs triggered potent gene knockdown during oogenesis (**Fig. 1b**,**c**). To determine whether maternal expression of shRNAs can also block expression of zygotically expressed genes, we generated shRNAs to a few zygotic genes that result in embryonic lethality when mutated. In all cases, we observed the expected phenotypes as shown for decapentaplegic (*dpp*; **Fig. 1c**). Finally, we tested the effectiveness of shRNAs expressed from Valium20 in somatic tissues. In general, the obtained phenotypes were stronger than those obtained with the long-hairpin–based vector Valium10 and resembled genetic null mutations for the respective genes (**Fig. 1d**,**e** and **Supplementary Fig. 3**).

Whereas shRNAs expressed from Valium20 generated effective knockdown phenotypes in germline and soma, the phenotypic penetrance in the germline was influenced by temperature and maternal age, indicating room for improvement (**Supplementary Note 1**). We therefore generated Valium22 (**Supplementary Fig. 2**) based on the UASp vector[7], which is optimized for transgene expression in the female germline. Indeed, Valium22-mediated knockdowns in the germline were overall stronger than those generated using Valium20. We note, however, that Valium22 did not allow robust transgene expression in the soma, leading to incomplete somatic knockdowns (data not shown).

We chose the ovarian Piwi-interacting RNA (piRNA) pathway as a system to compare the efficacies and specificities of Valium20 and Valium22. Both somatic and germline cells of the *Drosophila* ovary produce piRNAs to silence transposable elements, but the pathway architecture differs in both cell types[8]. We generated multiple shRNA lines targeting proteins with a role in the piRNA pathway. Consistent with the strong knockdown observed for each target (**Fig. 2a**), RNAi phenotypes generated using the maternal *MTD-Gal4* line and Valium22 were highly reminiscent of each published null mutant. Depletion of the proteins Piwi, Aub, Spn-E or Armi resulted in complete sterility (**Fig. 2b**), and we observed strong derepression of three transposable elements known to be targets of the germline piRNA pathway by quantitative reverse transcription–PCR (RT-PCR) or by antibody staining (**Fig. 2c**,**d**).

[1]Department of Genetics, Harvard Medical School, Howard Hughes Medical Institute, Boston, Massachusetts, USA. [2]Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. [3]Institute of Molecular Biotechnology, Vienna, Austria. [4]Pediatric Surgical Research Labs, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [5]Present addresses: Gene Regulation Laboratory and Tsinghua Fly Center, School of Life Sciences and School of Medicine, Tsinghua University, Beijing, China (J.-Q.N. and L.-P.L.) and Sanford-Burnham Medical Research Institute, La Jolla, California, USA (R.Z.). [6]These authors contributed equally to this work. Correspondence should be addressed to N.P. (perrimon@receptor.med.harvard.edu).

RECEIVED 22 NOVEMBER 2010; ACCEPTED 14 MARCH 2011; PUBLISHED ONLINE 3 APRIL 2011; DOI:10.1038/NMETH.1592

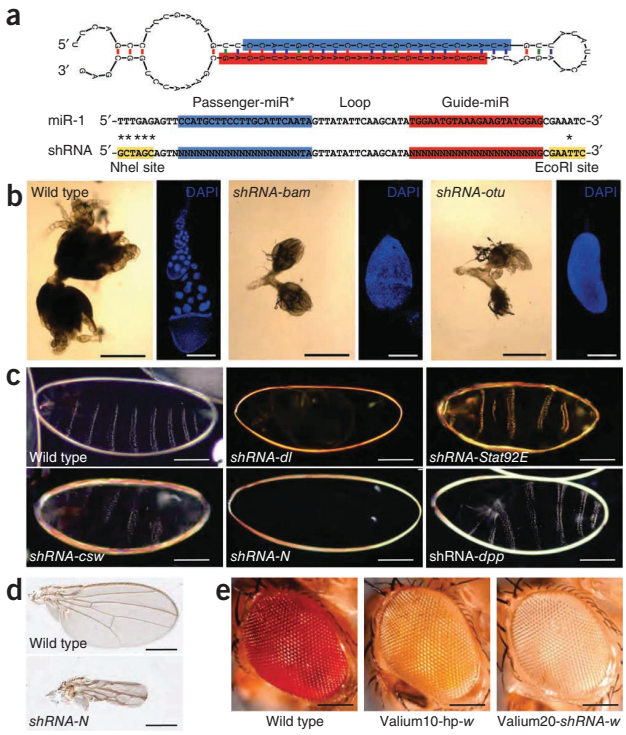


Figure 1 | Design of shRNA constructs and phenotypes of shRNA-mediated gene silencing. (**a**) Structure of the *Drosophila* miR-1 and shRNA hairpins (miR-1 nucleotides replaced by the sequence of interest are indicated by N). (**b**) Phase contrast images showing ovary phenotypes associated with knockdown of bag-of-marbles (shRNA to *bam*, labeled *shRNA-bam*) and ovarian tumor (*shRNA-otu*) in *MTD-Gal4/UAS-shRNA* females (using Valium20). DAPI images show single tumorous egg chamber and wild-type egg chamber. Scale bars, 500 µm (phase contrast) and 200 µm (DAPI). (**c**) Dark field images of the cuticle of wild-type embryo and embryos derived from *MTD-Gal4/UAS-shRNA* females. Scale bars, 100 µm. (**d**) Knockdown of Notch in the wing using *C96-Gal4/UAS-shRNA-N*. Scale bars, 400 µm. (**e**) Knockdown of white using *GMR-Gal4*. In the labels, hp stands for hairpin. Scale bars, 100 µm.

Loss-of-function mutations in *tudor* are known to have no impact on transposon silencing in the germline or on fertility, but these mutants are defective in Aub localization to the posterior end of the growing oocyte. This leads to a failure in the specification of primordial germ cells and results in a 'grandchild-less' phenotype. We observed all of these characteristics in the *shRNA-tud* (shRNA to the tudor gene) transgenic flies (data not shown). Typically, we observed similar phenotypes when we tested the same shRNAs with Valium20. However, Valium20-mediated knockdown of Piwi or Armi led to more severe phenotypes. This is in full agreement with data for their respective genetic null mutants, as it has been shown that the function of Piwi and Armi in the somatic support cells of the ovary is critical for germline development and therefore ovarian morphology. This indicates leaky shRNA expression from Valium20 transgenes in somatic ovarian cells (**Supplementary Table 1**, **Supplementary Fig. 4** and **Supplementary Note 1**). We suspect that a temperature-sensitive element present in the *hsp70* minimal promoter of Valium20 causes low-level expression that, in combination with basal Gal4 levels, leads to shRNA expression sufficient for gene knockdowns in the soma. Altogether, our results indicate that Valium22 is not only more effective for knockdowns in the germline but is also superior in terms of tissue specificity as the effects appear to be strictly restricted to the female germline.

It was important to understand the biogenesis requirements, processing accuracy and loading characteristics of shRNAs, which are designed to take advantage of different aspects of the mostly distinct fly miRNA and siRNA pathways to favor loading into AGO2, the major RNAi machinery in *Drosophila*. Our analysis indicated that shRNAs are processed by the sequential action of Drosha-Pasha and Dicer-1-LoqsPB complexes followed by loading into AGO1 and AGO2 proteins (**Supplementary Figs. 5a**,**b** and **6** and **Supplementary Note 1**). Loading into AGO2 required Dcr-2 and presumably also R2D2 and perfect base-pairing at positions 9 and 10 of the siRNA, as has been shown for endogenous siRNAs. Immunopurification experiments indicated that shRNAs are efficiently loaded into AGO2 as intended but that a fraction is also associated with AGO1. Small RNA sequencing efforts from cultured *Drosophila* cells transfected with three different Valium20 constructs showed that shRNAs are accurately processed at the intended 5′ ends. Furthermore, the amount of mature shRNA strands accumulated at levels comparable to those of the most abundant cellular microRNA (miR) strands and exceed miR complementary strand (miR*)-passenger strand strand levels several fold as predicted from the differences in thermodynamic stability of guide and passenger strand 5′ ends.

RNAi approaches can suffer from unspecific targeting of genes exhibiting sufficient sequence complementarity to the experimental siRNA (referred to as off-target effects)[9]. In comparison to long-hairpin constructs, shRNA constructs are expected to be advantageous as they give rise to only two siRNA species, the guide and passenger strands (**Supplementary Fig. 6** and **Supplementary Note 1**). We minimized off-target effects caused by extended complementarity by filtering out shRNAs whose guide or passenger strands have complementary matches of 16 nucleotides or more to the fly transcriptome. As shRNAs are partially loaded into AGO1, they could also produce off-target effects through fortuitous recognition of mRNAs via the 'seed' region leading to miRNA-like repression[10]. To address the problem of off-target effects experimentally, we generated several shRNA lines targeting genes that are not required for viability. Only one line out of eight tested was associated with unexpected lethality (**Supplementary Table 2**). This is similar to what has been observed with long hairpins[5,11]. Thus, off-target effects with shRNAs are still a matter of concern, and we suggest verification of obtained phenotypes using independent shRNA constructs to the same gene.

Encouraged by the remarkable silencing potency of shRNAs, we started to generate a large-scale community resource for transgenic RNAi in *Drosophila* (http://www.flyrnai.org/TRiP-HOME.html). As of March 2011, we have constructed over 2,900 fly stocks (an updated list is available at the Transgenic RNAi Project (TRiP) homepage), and they are available from the Bloomington *Drosophila* Stock Center. We aim to construct shRNAs to all 14,208 annotated *Drosophila* protein-coding genes (genome release 5). Toward this end, we predicted shRNA sequences for all genes using designer of small interfering RNA (DSIR)[12], an algorithm trained on effective siRNAs (http://biodev.extra.cea.fr/dsir/dsir.html). DSIR has proven quite reliable for the prediction of

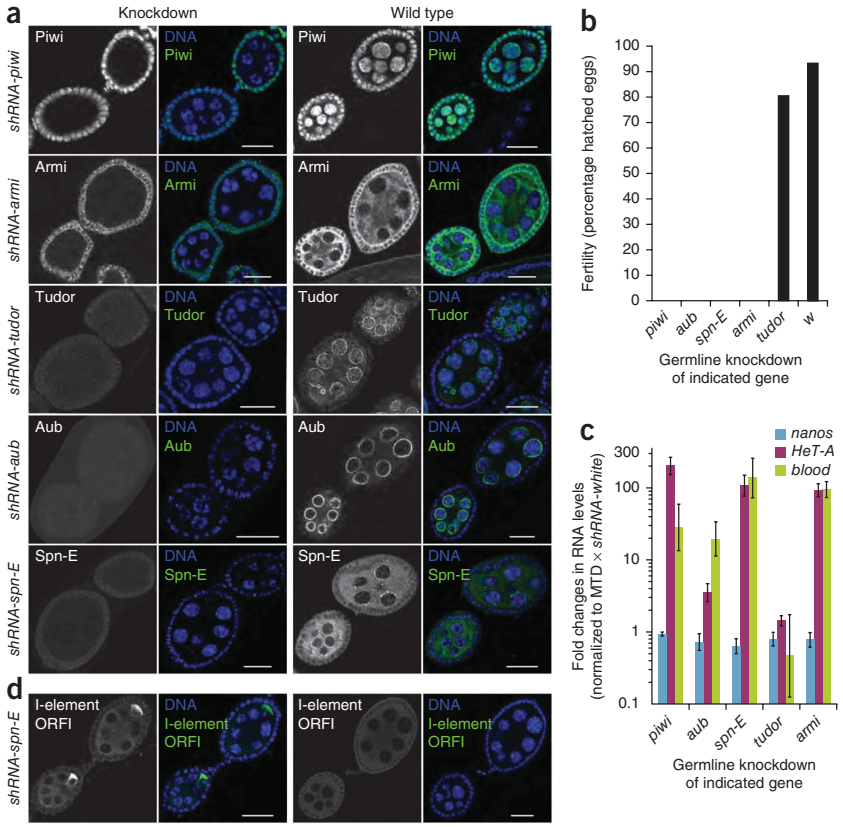


Figure 2 | Analysis of the piRNA pathway during oogenesis. (**a**) Immunofluorescence staining of early egg chambers showing depletion of the indicated piRNA pathway components using specific antibodies (green) upon shRNA expression via *MTD-Gal4* (using Valium22). DNA was visualized with DAPI (blue). Black and white images are of the antibody staining only. Scale bars, 20 µm. (**b**) Fertility rates of females in which the indicated genes were knocked down with shRNAs in the germline via *MTD-Gal4* (using Valium22). For each knockdown 300–500 eggs were counted. (**c**) Fold changes in steady-state RNA levels of the transposable elements *HeT-A* and *blood* in comparison to the germline-specific *nanos* transcript upon knockdown of the indicated genes via shRNAs. The data were compared to a control sample in which the white gene was knocked down (*rp49* transcript levels were used for normalization). Data are averages of three independent biological replicates; error bars, s.d. (**d**) Immunofluorescence staining of early egg chambers with an antibody to the I-element ORF1p. Left two images are of flies expressing *shRNA-spn-E* with *MTD-Gal4* (using Valium22); right two images are of wild-type flies. DNA was visualized with DAPI (blue).

shRNAs for effective knockdown in transgenic flies, and the vast majority of the shRNAs described in this paper were designed using the DSIR algorithm. We synthesized 83,256 unique shRNA oligonucleotides *in situ* on four custom glass-slide microarrays[13]. We amplified these as pools, and inserted them into Valium20 and Valium22. We analyzed ~160,000 individual clones per vector, and identified accurate clones through either conventional sequencing or a two-step process involving DNA Sudoku[14] compression followed by Illumina sequencing. We anticipate that at least 8,000 constructs per year will become available from the TRiP for distribution to the community.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Accession codes. Gene Expression Omnibus: GSE27039 (small RNA sequences).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

The design and construction of the first shRNAs were supported in part by the Janelia Farm Visitor Program. We thank G. Rubin, C. Zuker and T. Laverty for their interest and support; R. Hardy and C. Zuker for the data presented in **Supplementary Table 2**; B. Haley for helpful discussion on shRNAs; L. Cooley (Yale University) for the gift of the MTD-Gal4 line; and Z. Xuan for help with library design. S. Zusman and M. Tworoger of Genetic Services, Inc. generated the transgenic lines. R.Z. is supported by the Leukemia and Lymphoma Society. B.C. is supported by a PhD fellowship from the Boehringer Ingelheim Fonds. This work was supported by two US National Institute of General Medical Sciences R01 grants (GM067761 and GM084947) to N.P., an EU FP7 European Research Council starting grant to J.B. and contributions from the US National Institute of Neurological Disorders and Stroke.

AUTHOR CONTRIBUTIONS

J.-Q.N., R.Z. and B.C. carried out major experiments; L.-P.L., L.H., D.Y.-Z., H.-S.S., R.B., M.B. and L.A.P. produced the TRiP lines; P.K. performed the luciferase experiments in ovaries; D.H. and J.B. analyzed the piRNA pathway during oogenesis; and G.J.H. and N.P. supervised the project. R.Z., B.C., J.-Q.N., D.H., J.B., G.J.H. and N.P. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Perrimon, N., Ni, J.Q. & Perkins, L. *Cold Spring Harb. Perspect. Biol.* **2**, a003640 (2010).
2. Czech, B. & Hannon, G.J. *Nat. Rev. Genet.* **12**, 19–31 (2011).
3. Haley, B., Hendrix, D., Trang, V. & Levine, M. *Dev. Biol.* **321**, 482–490 (2008).
4. Chen, C.H. *et al. Science* **316**, 597–600 (2007).
5. Ni, J.Q. *et al. Genetics* **182**, 1089–1100 (2009).
6. Petrella, L.N., Smith-Leiker, T. & Cooley, L. *Development* **134**, 703–712 (2007).
7. Rorth, P. *Mech. Dev.* **78**, 113–118 (1998).
8. Malone, C.D. & Hannon, G.J. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 225–234 (2009).
9. Kulkarni, M.M. *et al. Nat. Methods* **3**, 833–838 (2006).
10. Birmingham, A. *et al. Nat. Methods* **3**, 199–204 (2006).
11. Dietzl, G. *et al. Nature* **448**, 151–156 (2007).
12. Vert, J.P., Foveau, N., Lajaunie, C. & Vandenbrouck, Y. *BMC Bioinformatics* **7**, 520 (2006).
13. Cleary, M.A. *et al. Nat. Methods* **1**, 241–248 (2004).
14. Erlich, Y. *et al. Genome Res.* **19**, 1243–1253 (2009).

ONLINE METHODS

***Drosophila* strains.** The maternal triple driver (*MTD*)-*Gal4* stock[6] was a gift from L. Cooley (Yale University). The stock contained homozygous insertions of three *Gal4* constructs, which together provide robust germline and maternal *Gal4* expression. The genotype was *P{COG-Gal4:VP16}*; *P{Gal4-nos.NGT}40*; *P{nos-Gal4-VP16}* (Bloomington stock 31777). *P{COG-Gal4:VP16}*[7] contained a promoter from the *otu* gene and the 3′ untranslated region (UTR) from the *K10* gene. *Gal4:VP16* expression from this transgene was weak or absent in the germarium and robust beginning in stage-1 egg chambers. *P{nos-Gal4-VP16}* contained both the promoter and 3′ UTR from the *nanos* gene[15] and was expressed throughout the germarium and in all stages of egg chambers, with lower expression in young egg chambers (~stages 2–6)[7]. *P{Gal4-nos.NGT}40* contained the nanos promoter and *αTub84E* 3′ UTR[16], and was made for maternal loading of *Gal4* to drive expression during embryogenesis.

GMR-Gal4 and *C96-Gal4* were used to drive expression in the eye and wing, respectively, as described previously[5]. Their descriptions are available from FlyBase (http://flybase.org/). Details on the full genotype of all the lines used in this study are available on the TRiP website (http://www.flyrnai.org/TRiP-HOME.html).

Phenotypic analyses. For DAPI staining, ovaries were dissected in PBS and fixed in 4% electron microscopy (EM)-grade paraformaldehyde (Electron Microscopy Sciences) diluted in PBS for 30 min. Ovaries were counterstained with DAPI (Invitrogen) for 10 min. Embryonic cuticles and wings were prepared as described previously[17,18]. For immunofluorescence, ovaries were dissected from 3–5-day-old flies into ice-cold PBS and subsequently fixed in 4% formaldehyde (Thermo Scientific) containing 0.15% Triton X 100 (Sigma-Aldrich), diluted in PBS, for 25 min. After three rinses with PBT (PBS with 0.3% Triton X 100) ovaries were blocked in BBX (PBS containing 0.3% Triton X 100 and 0.1% BSA) for 30 min at room temperature (20–22 °C). Ovaries were incubated with primary antibodies over night at 4 °C diluted in BBX (antibodies to Piwi, Aub and Ago3, 1:500; antibodies to Armi and I element, 1:1,000; antibodies to Tudor, 1:10; antibodies to Spn-E, 1:50). After four PBT washes secondary antibodies were incubated 5 h at room temperature diluted in BBX (1:500; Molecular Probes). Ovaries were stained with DAPI for 10 min in the second of four PBT washes. Antibodies used were: antibody to Piwi, antibody to Aub and antibody to AGO3 (ref. 19); antibody to Tudor, antibody to Spn-E[20]; antibody to Armi[21] and antibody to I element (gift from D. Finnegan; University of Edinburgh). For the sterility test, ten 3–5-day-old female flies were pre-mated with wild-type males overnight in small cages on apple juice plates with yeast paste. Apple juice plate was changed without anesthetizing flies. After 18 h at 25 °C, the flies were removed and the number of laid eggs was counted (typically ~200 eggs). Forty-eight hours later hatched and non-hatched eggs were determined.

Additional information on the phenotypic analyses of RNAi reagents is available in **Supplementary Figures 7** and **8** as well as in **Supplementary Table 3**.

Vector construction. For descriptions of vector construction, see **Supplementary Note 2**.

β-elimination. The chemical structure of the 3′ termini of small RNAs was analyzed as described previously[22]. In brief, RNA from immunoprecipitates or 25 µg of total RNA from S2 cells treated with the indicated dsRNAs (17.5 µl total volume for each sample) was incubated at room temperature for 30 min with 5 µl 5× borate buffer (148 mM borax and 148 mM boric acid; pH 8.6) supplemented with 3.125 µl freshly prepared 200 mM $NaIO_4$. We added 5 µl of 50% glycerol to quench nonreacted sodium periodate by incubating for an additional 15 min at room temperature. Samples were then vacuum-dried and dissolved in 60 µl 1× borax buffer (30 mM borax, 30 mM boric acid and 50 mM NaOH; pH 9.5). β-elimination was carried out by incubation for 2 h at 45 °C. RNAs were ethanol-precipitated and resolved in 1× gel loading buffer.

Northern blotting. Northern blotting was carried out as described previously[23,24]. In brief, total RNAs from knockdown cells were isolated using TRIzol (Invitrogen). We separated 30 µg total RNAs from cultured cells (with or without β-elimination) or RNAs from immunoprecipitations on 15% denaturing polyacrylamide gels and transferred to Hybond-N+ membranes (Amersham Biosciences) in 1× TBE buffer. Small RNAs were UV-light cross-linked to the membrane and prehybridized in ULTRAhyb-Oligo buffer (Ambion) for 1 h. DNA probes complementary to the indicated strands were 5′ radio-labeled and added to the hybridization buffer (hybridization for 6 h at 30 °C). Membranes were washed 4 times in 1× SSC with 0.1% SDS at 30 °C and exposed to PhosphorImager screens (GE Healthcare) for 12–48 h. Membranes were stripped by heating in 0.2× SSC containing 0.1% SDS in a microwave twice. Sequences of the oligonucleotide probes are listed in **Supplementary Note 3**.

Immunoprecipitation. Cell extracts were prepared, evenly split and immunoprecipitated using antibodies to AGO1 (Abcam) or the Flag epitope (Sigma), respectively, as previously described[23]. RNAs were recovered from the immunoprecipitated samples using TRIzol and used for northern blotting.

Transposon qPCR analysis. Total RNA was extracted from ovaries of 3–5-day-old flies using TRIzol. cDNA was prepared with random primers. qPCR was performed using Maxima SYBR Green/ROX qPCR Master mix (Fermentas). Calculation of steady-state RNA levels was calculated applying the $2^{-\Delta\Delta Ct}$ method[25]. *Rp49* was used for normalization of all samples, and fold enrichments were calculated in comparison to an shRNA knockdown targeting the white gene. Fold changes in steady-state transcript levels and s.d. were calculated from three biological replicates. Primer sequences are available in **Supplementary Note 3**.

Small RNA libraries. Small RNAs were cloned as described previously[19]. For this study, the following small RNA libraries from total RNAs were prepared: 19-nucleotide (nt) to 24-nt from S2 cells transfected with shRNA to *dlg1* (*shRNA-dlg1*); 19-nt to 24-nt from S2 cells transfected with *shRNA-N*; and 19-nt to 24-nt from S2 cells transfected with *shRNA-dpp*. For each construct, ~4×10^6 S2-NP cells were transfected with 2 µg of Valium20-shRNA construct and 1 µg of pMT-Gal4 plasmid. ShRNA expression in cells was induced by adding

doi:10.1038/nmeth.1592

500 μM $CuSO_4$ 2 d after transfection. Total RNA was isolated using TRIzol 24 h after induction. Libraries were sequenced in-house using the Illumina GA-II sequencing platform.

Bioinformatic analysis of small RNA libraries. The analysis of small RNA libraries was performed as previously described[26]. Illumina reads were stripped of the 3′ linker and collapsed, and the resulting small RNA sequences were matched without mismatches to the *Drosophila* release 5 genome and to the genomes of *Drosophila* C virus, Flock house virus and Cricket paralysis virus with up to three mismatches. Only reads that met these conditions were analyzed further. For annotations we used a combination of University of California Santa Cruz genome browser, miRBase and Flybase tracks for protein-coding genes, repeats or transposons, noncoding RNAs and microRNAs as well as custom tracks (for shRNAs, synthetic markers, endo-siRNAs from structured loci, miR and miR* strands) with different priorities (annotation priority list is available upon request). For comparison of small RNA counts between libraries, reads were normalized to the same total number after bioinformatic removal of sequences matching to synthetic cloning markers or assumed degradation products of abundant cellular RNAs (rRNAs, snoRNAs and tRNAs). Heatmaps were computed by plotting the abundance and ratio of individual miR, miR* and shRNA strands in each library.

Construction of the shRNA library. An shRNA library representing 83,256 unique synthetic hairpins was synthesized on four custom 22K Agilent microarrays[13]. The library covered all 14,208 annotated genes (excluding small RNA and noncoding RNA genes) of the *Drosophila* release 5 genome with up to six shRNAs per gene (14,138 genes were covered by six hairpins, and 14,147 genes were covered by five hairpins). Hairpin constructs were based on the miR-1 backbone and essentially resembled those described above with perfect complementarity between guide and passenger strands. Additional sequence was attached on both ends for PCR amplification. In addition, to eliminate off-target effects only shRNAs that lacked sequence complementarity to annotated microRNA 'seed' sequences were considered. DNA pools from microarray chips were amplified[13] and cloned into Valium20 and Valium22 destination vectors. Plasmid DNA was transformed, clones were picked (160,000 individual clones per destination vector), and resulting transformants were multiplexed using DNA Sudoku[14] at Open Biosystems. Pools were barcoded via PCR, and amplicons were sequenced in-house using the Illumina GA-II sequencing platform. Positive clones were picked into 96-well plates and Sanger sequencing was carried out to validate correct shRNA sequences. Once shRNA clones in Valium20 and Valium22 are available, they will be openly available to the *Drosophila* community.

15. Van Doren, M., Williamson, A.L. & Lehmann, R. *Curr. Biol.* **8**, 243–246 (1998).
16. Tracey, W.D. Jr. *et al. Genetics* **154**, 273–284 (2000).
17. Ni, J.Q. *et al. Nat. Methods* **5**, 49–51 (2008).
18. Perrimon, N., Engstrom, L. & Mahowald, A.P. *Genetics* **121**, 333–352 (1989).
19. Brennecke, J. *et al. Cell* **128**, 1089–1103 (2007).
20. Nishida, K.M. *et al. EMBO J.* **28**, 3820–3831 (2009).
21. Cook, H.A., Koppetsch, B.S., Wu, J. & Theurkauf, W.E. *Cell* **116**, 817–829 (2004).
22. Vagin, V.V. *et al. Science* **313**, 320–324 (2006).
23. Czech, B. *et al. Nature* **453**, 798–802 (2008).
24. Zhou, R. *et al. RNA* **15**, 1886–1895 (2009).
25. Livak, K.J. & Schmittgen, T.D. *Methods* **25**, 402–408 (2001).
26. Czech, B. *et al. Mol. Cell* **36**, 445–456 (2009).

Nature Methods

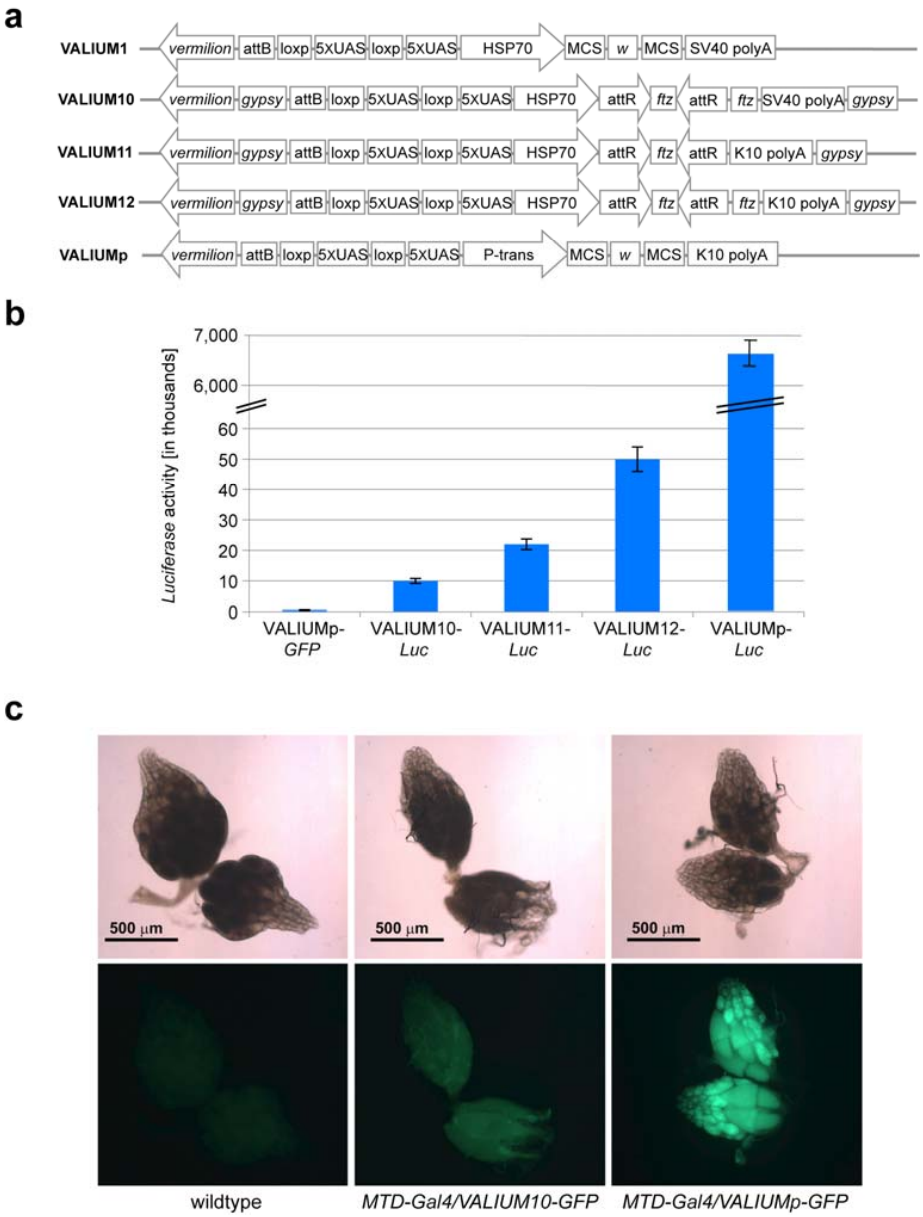
A genome-scale shRNA resource for transgenic RNAi in *Drosophila*

Jian-Quan Ni, Rui Zhou, Benjamin Czech, Lu-Ping Liu, Laura Holderbaum, Donghui Yang-Zhou, Hye-Seok Shim, Dominik Handler, Phillip Karpowicz, Richard Binari, Matthew Booker, Julius Brennecke, Lizabeth A Perkins, Gregory J Hannon & Norbert Perrimon

Supplementary Figure 1	Vectors tested for efficient expression in the female germline.
Supplementary Figure 2	VALIUM20 and VALIUM22 are *miR-1* based shRNA vectors for transgenic RNAi.
Supplementary Figure 3	VALIUM20 is a very effective vector for somatic RNAi.
Supplementary Figure 4	Leaky expression by VALIUM20-shRNA transgenes.
Supplementary Figure 5	Biogenesis of shRNAs and their loading into effector complexes.
Supplementary Figure 6	Abundance and processing accuracy of shRNAs.
Supplementary Figure 7	VALIUM11 and VALIUM12 vectors are not as effective as VALIUM10 to generate somatic phenotypes.
Supplementary Figure 8	Dependence of shRNA-mediated knockdowns on mother age.
Supplementary Table 1	Analysis of oogenesis phenotypes using shRNA lines in VALIUM20.
Supplementary Table 2	ShRNA lines targeting genes that are not required for viability.
Supplementary Table 3	Transgenic long-hairpin RNAi lines analyzed for oogenesis phenotypes.
Supplementary Note 1	Additional text
Supplementary Note 2	Vectors construction
Supplementary Note 3	Primer sequences

TABLES AND FIGURES

Supplementary Figure 1: Vectors tested for efficient expression in the female germline.



a. Structure of the VALIUM vectors tested for germline expression and their abilities to generate oogenesis and maternal effect phenotypes.

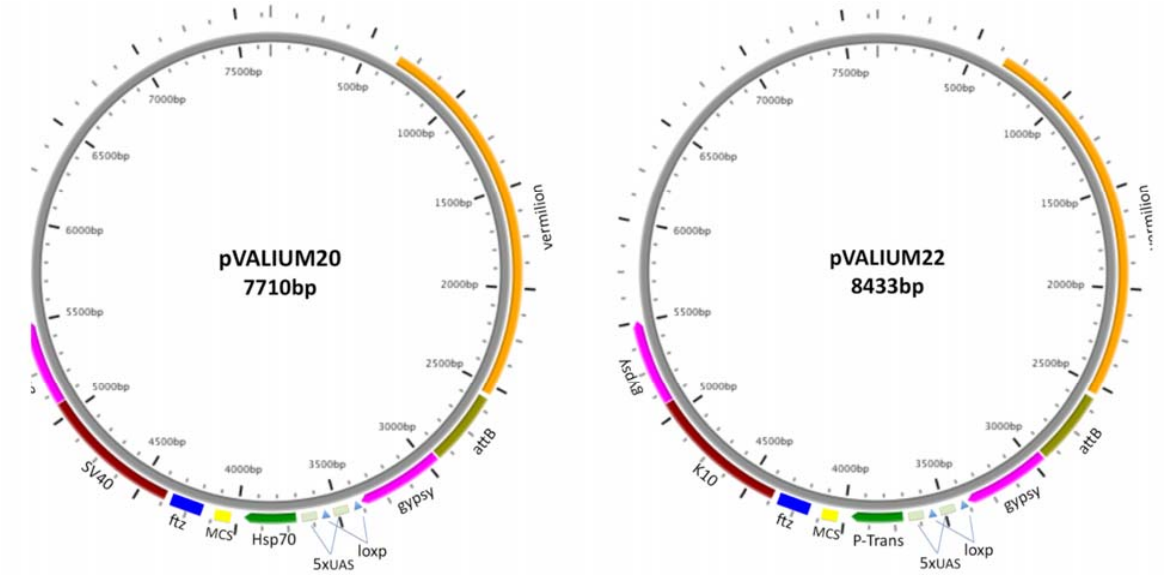
b. *Luciferase* expression levels in the indicated *MTD-Gal4* driven *VALIUM-Luciferase* flies.

c. Expression of *GFP* in *MTD-Gal4/VALIUM10-GFP* and *MTD-Gal4/VALIUMp-GFP*. None of these vectors generated phenotypes during oogenesis and embryogenesis with the

Nature Methods: doi.10.1038/nmeth.1592

lines tested (Supplementary Table 3).

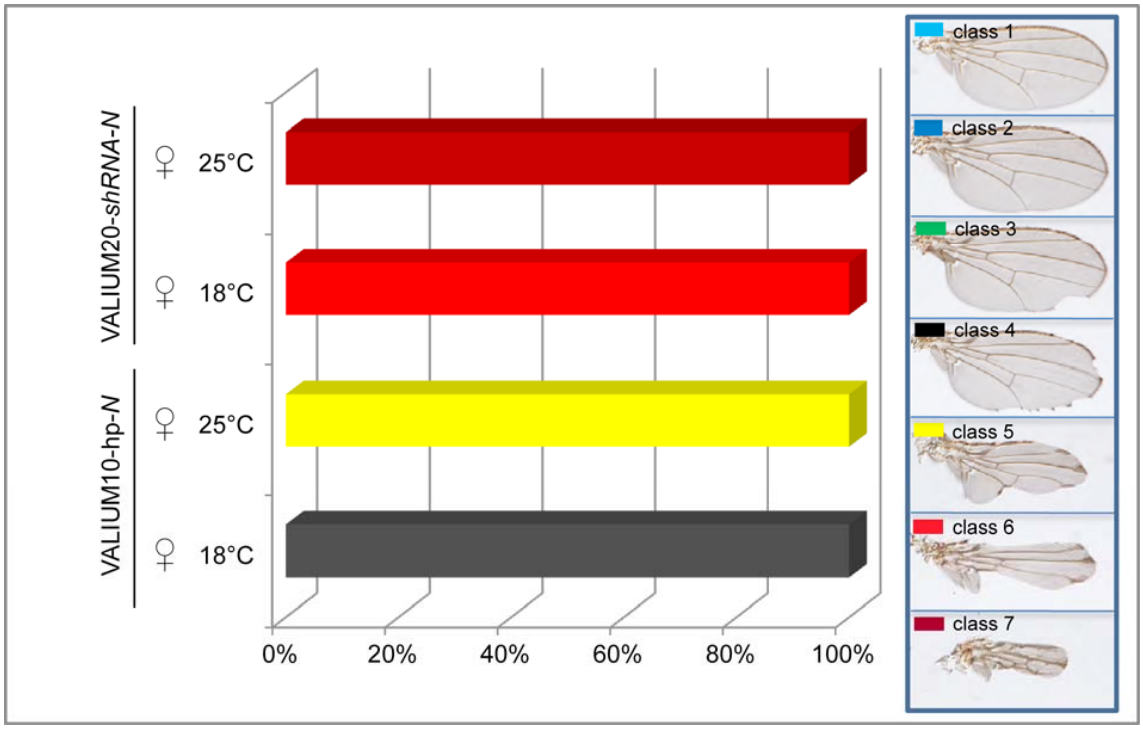
Supplementary Figure 2: VALIUM20 and VALIUM22 are *miR-1* based shRNA vectors for transgenic RNAi.



VALIUM20 contains *vermilion* as a selectable marker, an attB sequence to allow for targeted, phiC31-mediated integration at genomic attP landing sites, two pentamers of UAS (one can be excised using the Cre/loxP system to generate a 5XUAS derivative), the *hsp70* core promoter; the an SV40 polyadenylation signal, and an intronic sequence to facilitate RNA nuclear export. The relevant sequences are flanked by two gypsy insulators to ensure stable transgene expression. VALIUM22 comprises essentially the same features except that it contains the K10 3'UTR and the P-element transposase minimal promoter.

Nature Methods: doi.10.1038/nmeth.1592

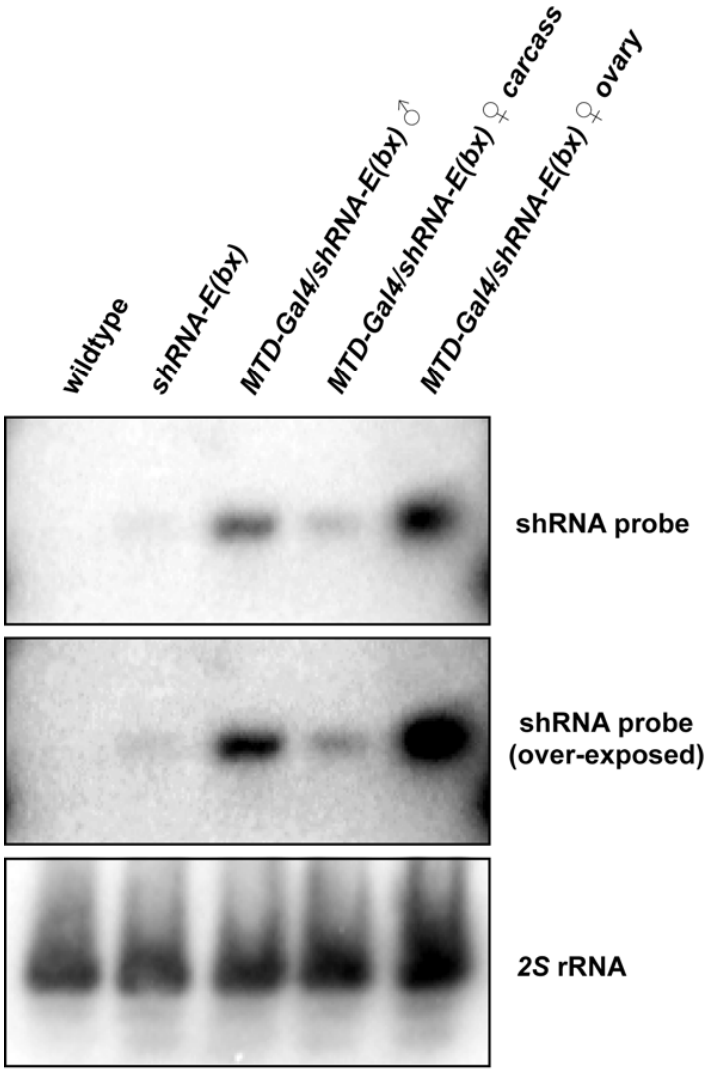
Supplementary Figure 3: VALIUM20 is a very effective vector for somatic RNAi.



The phenotypes of N knockdown using long-hairpins (VALIUM10) and shRNAs (VALIUM20) are compared in the wing. *C96-Gal4* was used to express *VALIUM10-hp-N* or *C96-Gal4/VALIUM20-shRNA-N***.** Phenotypes were classified by the severity of their wing defects[1] (class 1: wildtype or a few bristles missing; class 2: margin bristles missing but no notches; class 3: Moderate wing notching; class 4: extensive wing notching; class 5: most of the wing margin is missing; class 6: complete lack of wing margin and the wing blade is greatly reduced in size; class 7: Wings are almost completely missing. Note that wing images labeled as “class 1” and “class 7” are also presented in **Fig. 1d**).

Nature Methods: doi.10.1038/nmeth.1592

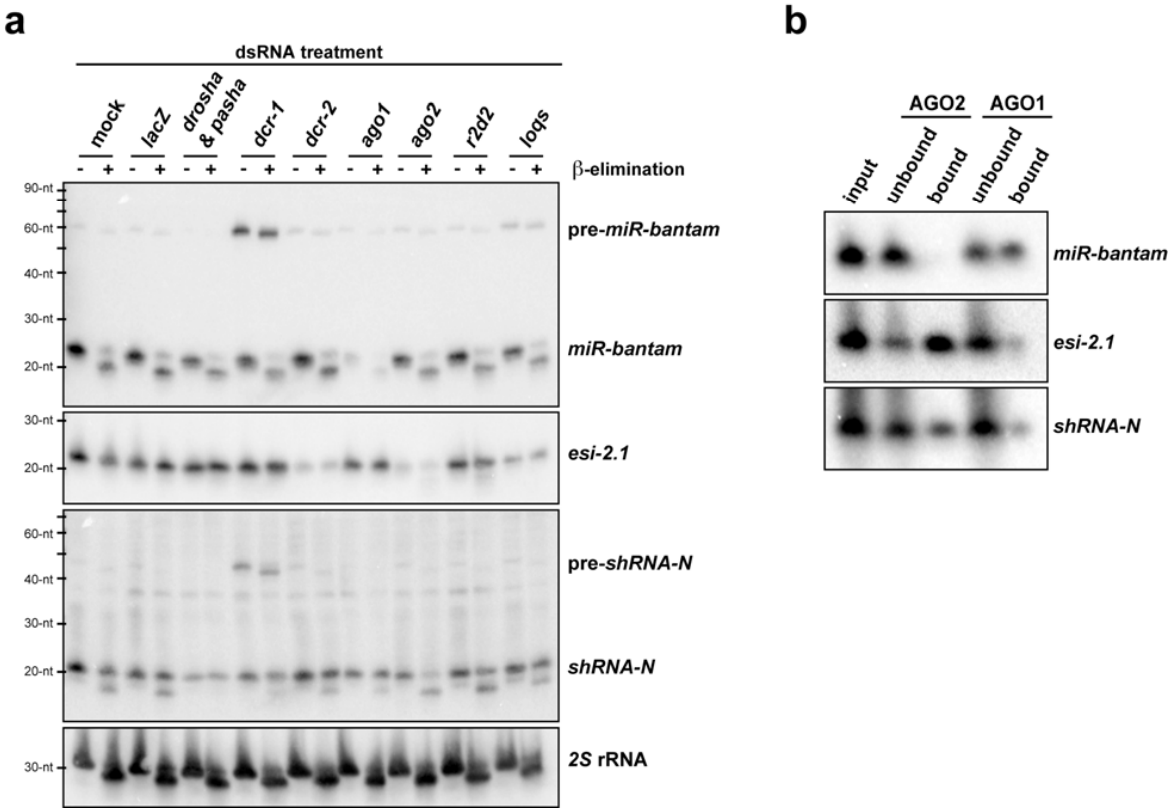
Supplementary Figure 4: Leaky expression by VALIUM20-shRNA transgenes.



Northern blotting showing the expression levels of *MTD-Gal4/UAS-shRNA-E(bx)* in total RNA preparation from (left to right) wildtype flies, *MTD-Gal4/UAS-shRNA-E(bx)* parental stocks, *MTD-Gal4/UAS-shRNA-E(bx)* males, carcasses of *MTD-Gal4/UAS-shRNA-E(bx)* females in which ovaries have been removed, or ovaries of *MTD-Gal4/UAS-shRNA-E(bx)* females. An over-exposed image is shown in the middle panel. *2S* rRNA serves as loading control.

Nature Methods: doi.10.1038/nmeth.1592

Supplementary Figure 5: Biogenesis of shRNAs and their loading into effector complexes.

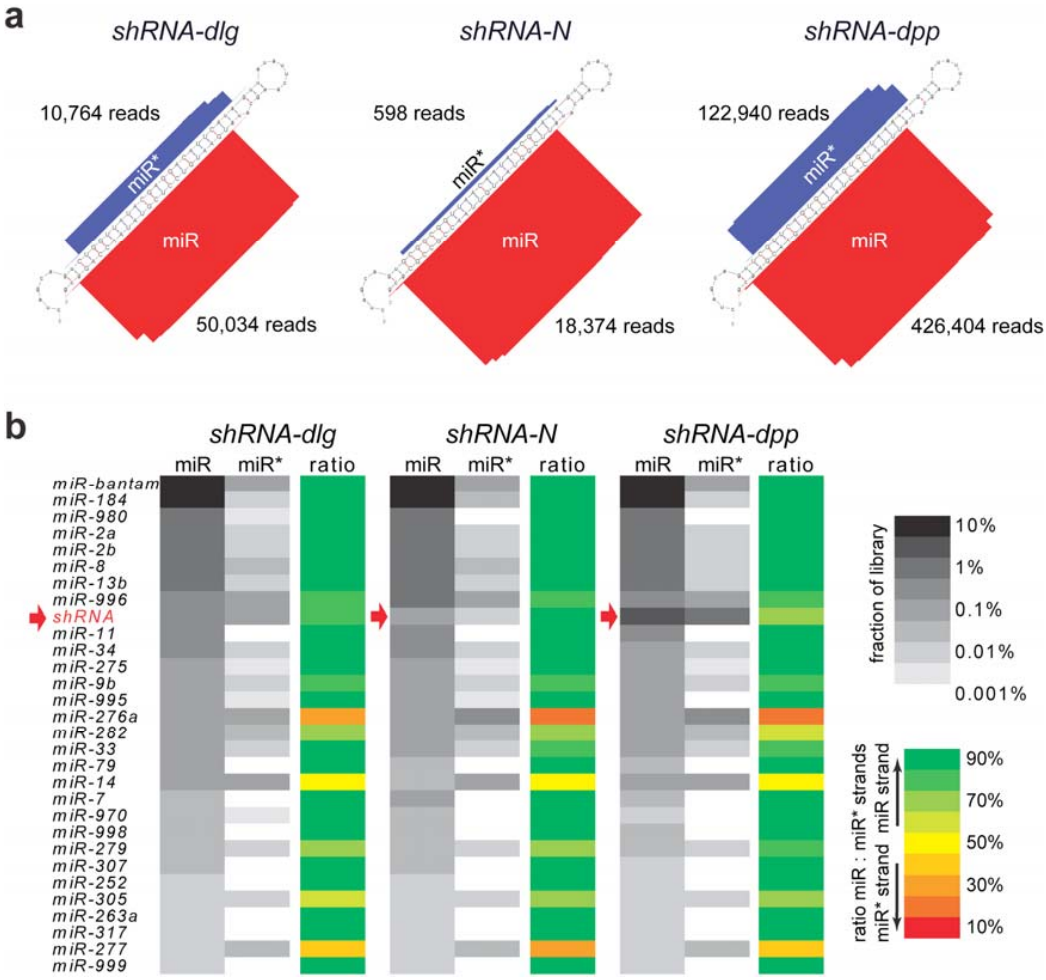


a. Northern blotting showing the steady-state levels of a representative shRNA (*shRNA-N* in VALIUM20) as well as those of an endogenous siRNA (*esi-2.1*) and miRNA (*miR-bantam*) in cultured *Drosophila* cells upon dsRNA-mediated depletion of canonical components of the siRNA and miRNA pathways (knockdowns indicated). In addition, a fraction of the RNA samples was subjected to periodate treatment followed by β-elimination (indicated by +/-). *2S* rRNA serves as the loading control.

b. Cytoplasmic extracts from cells expressing the same shRNA and Flag-tagged-AGO2 were evenly split and subjected to immunoprecipitation using anti-AGO1 and anti-Flag antibodies, respectively. Total RNAs recovered from the immunoprecipitates, as well as those recovered from cell extracts prior to and after immunoprecipitation, were subjected to sequential Northern blotting using probes against *esi-2.1*, *miR-bantam* and *shRNA-N*.

Nature Methods: doi.10.1038/nmeth.1592

Supplementary Figure 6: Abundance and processing accuracy of shRNAs.

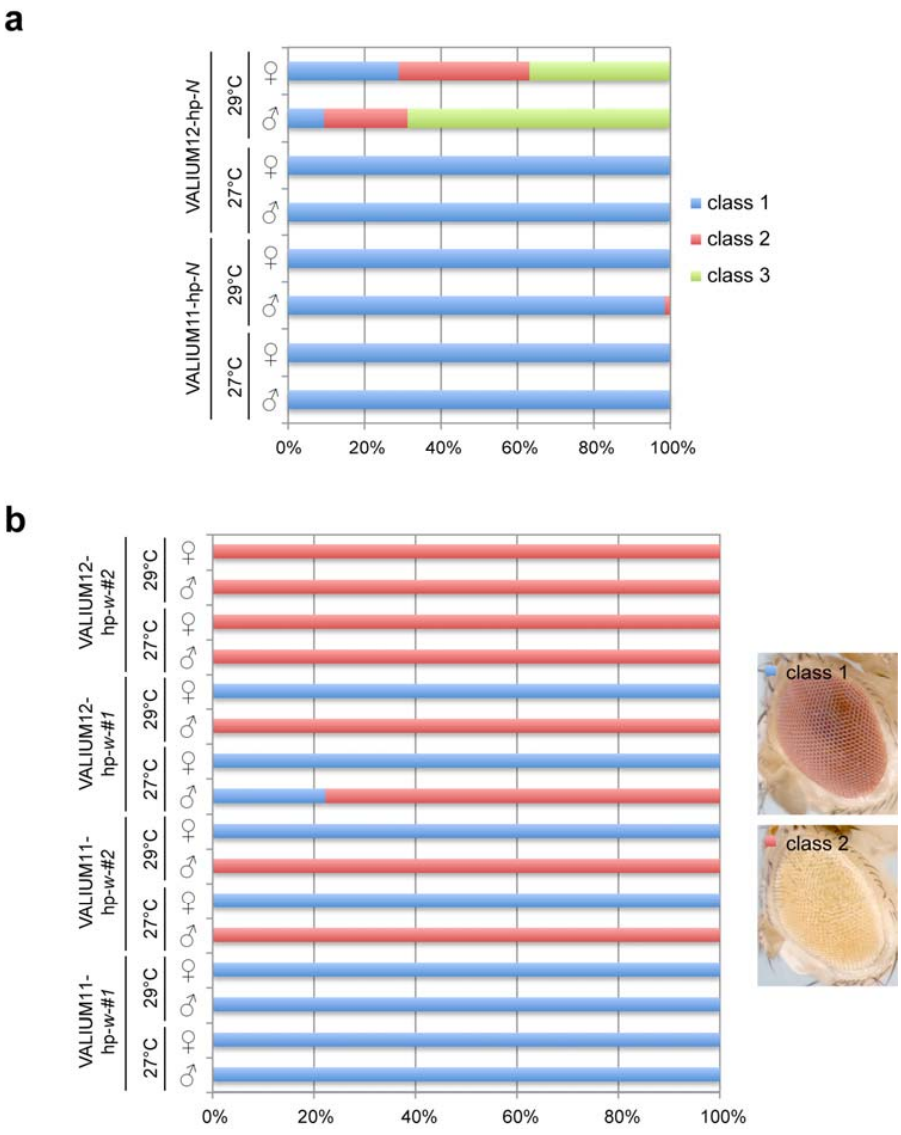


a. Cultured *Drosophila* cells were independently transfected with three different shRNAs and small RNAs were sequenced. Display of the shRNA precursors together with normalized cloning counts in each of the indicated small RNA libraries. Abundances of guide/miR strands (shown in red) and passenger/miR* strands (shown in blue) are indicated by bars. The accuracy of 5' end processing for the shown shRNAs is represented by sharp peaks at the intended sites.

b. Heat maps showing relative levels of the 30 most abundant microRNAs and the indicated shRNA (highlighted by the arrow) in each small RNA library. Both, miR and miR* strands are indicated separately in grey scale, while the ratio between miR and miR* strands is shown in green-red scale.

Nature Methods: doi.10.1038/nmeth.1592

Supplementary Figure 7: **VALIUM11 and VALIUM12 vectors are not as effective as VALIUM10 to generate somatic phenotypes.**



a. *C96-Gal4/VALIUM11-hp-N* and *C96-Gal4/VALIUM12-hp-N* were tested for their wing phenotypes. Phenotypes were classified by the severity of their wing defects (class 1: wildtype or a few bristles missing; class 2: margin bristles missing but no notches; and class 3: moderate wing notching. More details on classification can be found in Supplementary Fig. 3). The phenotypes were analyzed in males and females and at different temperatures.

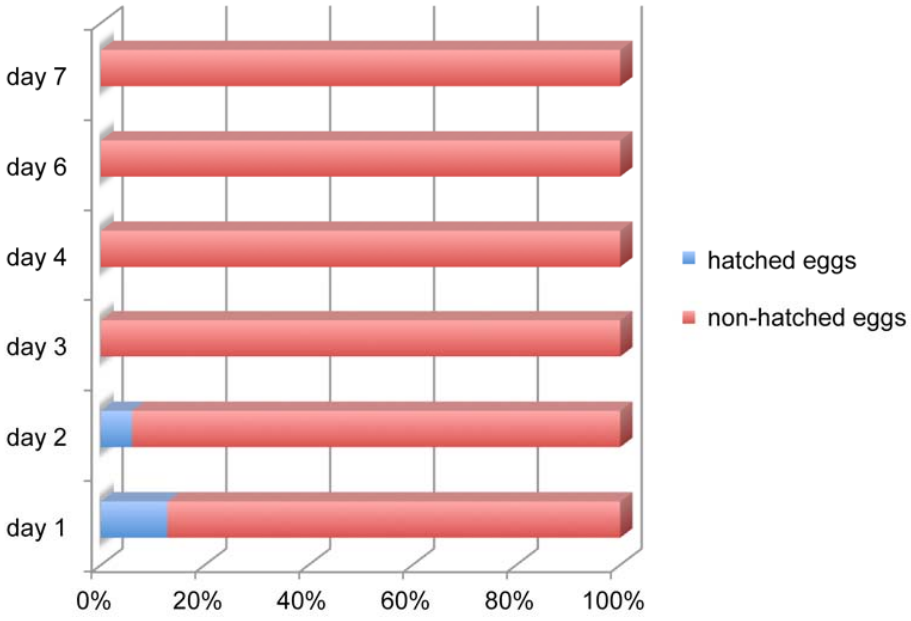
b. Two different shRNAs against *white* were tested in either VALIUM11 or VALIUM12. For expression, eye-specific *GMR-Gal4* was used.

Nature Methods: doi.10.1038/nmeth.1592

Supplementary Figure 8: Dependence of shRNA-mediated knockdowns on mother age.

a

b



Age dependence of the neurogenic phenotypes in eggs derived from *MTD-Gal4/VALIUM20-shRNA-N* mothers crossed to siblings. The number of eggs collected at different days (**a**) and the percentage of eggs hatching (**b**) are indicated.

Nature Methods: doi.10.1038/nmeth.1592

Supplementary Table 1: Analysis of oogenesis phenotypes using shRNA lines in VALIUM20.

Genes	Temperature	Hairpin ID	Fertility Phenotype	Ovaries	*HeT-A*	*Blood*
w		SH00017.N	lots of eggs and larvae	normal	1	1
spn-E	29C	SH00574.N	lots of eggs, no larvae	normal	69.13	160.35
	29C				39.40	79.59
	29C				107.79	177.13
	29C				114.19	155.21
	25C		lots of eggs, no larvae	normal	53.19	199.99
	25C				53.61	186.12
	25C				56.43	239.14
AGO3	29C	SH00108.N	lots of eggs, few larvae	normal	23.95	160.88
	29C				15.47	127.26
	29C				19.47	81.40
	25C		lots of eggs, few larvae	normal	3.39	100.35
	25C				14.47	163.18
	25C				3.31	68.40
armi	29C	SH00116.N	no egg	intermediate	23.18	82.35
	29C				19.42	85.69
	29C				17.84	77.66
	25C		no egg	intermediate	63.41	206.39
	25C				41.90	170.65
piwi	29C	SH00477.N	no egg	tiny	5.33	10.61
	29C				9.15	35.90
	29C				3.63	6.85
	25C		no egg	tiny	7.47	23.47
	25C				8.23	20.80
	29C	SH00951.N	no egg	tiny	6.26	8.94
	29C				6.29	9.71
	25C		no egg	tiny	11.34	17.84
	25C				10.41	21.73
tudor	29C	SH00601.N	lots of eggs and larvae	normal	1.24	2.42
	29C				1.34	1.35
	25C		lots of eggs and larvae	normal	1.36	1.40
	25C				1.15	1.41
	25C				1.11	1.51

ShRNA females were crossed to *MTD-Gal4* males at the indicated temperature, and the resulting F1 females were crossed to wildtype (Ore^R) males to determine fertility. Total RNAs from ovaries were prepared and subjected to qRT-PCR analysis to determine the knockdown efficiency and degree of transposon de-repression. Results were first normalized against *rp49* levels and then to control samples from the *shRNA-w* cross.

Supplementary Table 2: ShRNA lines targeting genes that are not required for viability.

Nature Methods: doi.10.1038/nmeth.1592

Gene Name	Line Name	Crossed elav-Gal4		Crossed elav-Gal4/Cyo		Crossed Scratch-Gal4		Crossed RH1-Gal4		Crossed GMR-Gal4	
		25℃	29℃	25℃	29℃	25℃	29℃	25℃	29℃	25℃	29℃
w	SH000017	viable,white eye	viable,white eye	viable	viable	viable	viable	viable	viable		
TrpA1	SH000019	viable	viable	viable	viable	viable	viable	viable	viable	viable	viable
channel	SH000022	viable,male lethal	viable,male lethal	viable,male lethal	viable,male lethal	viable,male lethal	viable,male lethal	viable	viable	viable	viable
eye	SH000029	viable,eye-like O.N.	viable,eye-like O.N.	viable,eye-like O.N.	viable,eye-like O.N.	viable,eye-like O.N.	viable,eye-like O.N.	viable	viable	viable	viable
Galpha498	SH000032	lethal	lethal	lethal	lethal	lethal	lethal	viable	viable	viable	viable
InaC	SH000036	viable	viable	viable	viable	viable	viable	viable	viable	viable	viable
Arr2	SH000037	viable	viable	viable	viable	viable	viable	viable	viable	viable	viable
sev	SH000039	viable,sev-like O.N.	viable,sev-like O.N.	viable	viable	viable	viable	viable	viable	viable	viable

O.N. = optical neutralization

We generated shRNA lines against a number of genes for which null alleles are homozygous viable. Details on these genes and their associated phenotypes, as well as the used Gal4 lines, can be found in Ni et al.[1]. Data are courtesy of Robert Hardy and Charles Zuker.

Nature Methods: doi.10.1038/nmeth.1592

Supplementary Table 3: Transgenic long-hairpin RNAi lines analyzed for oogenesis phenotypes.

CG#	Gene name	Vector	TRiP#	Others
CG15793	Dsor1	VALIUM1	JF01697	
CG32443	Pc	VALIUM1	JF01581	
CG3954	csw	VALIUM1	HM04072	
CG4006	Akt1	VALIUM1	HM04007	
CG4257	Stat92E	VALIUM1	JF01265	
CG10079	Egfr	VALIUM10	JF01368	
CG12072	wts	VALIUM10	JF02741	
CG31666	chinmo	VALIUM10	JF02341	
CG3936	N	VALIUM10	JF01637	
CG4059	ftz-f1	VALIUM10	JF02738	
CG5671	Pten	VALIUM10	JF01987	
CG6667	dl	VALIUM10	JF02825	
CG7935	msk	VALIUM10	JF02727	
CG9885	dpp	VALIUM10	JF01371	
CG10023	Fak56D	VALIUM12	JF01646	
CG10595	d	VALIUM12	JF01668	
CG10917	fj	VALIUM12	JF01667	
CG11326	Tsp	VALIUM12	JF01644	
CG13345	tum	VALIUM12	JF01753	
CG13852	mats	VALIUM12	JF01690	
CG1389	tor	VALIUM12	JF01548	
CG14228	Mer	VALIUM12	JF01688	
CG14992	Ack	VALIUM12	JF01718	
CG1560	mys	VALIUM12	JF01645	
CG1730	dlg1	VALIUM12	JF01638	
CG17593	CG17593	VALIUM12	JF01903	
CG2019	disp	VALIUM12	JF01662	
CG2047	ftz	VALIUM12	JF01550	
CG2747	CG2747	VALIUM12	JF01643	
CG2759	w	VALIUM12	JF01543	
CG2759	w	VALIUM12	JF01546	
CG31094	LpR1	VALIUM12	JF01670	
CG3218	fs(1)K10	VALIUM12	JF01676	
CG32702	CG32702	VALIUM12	JF01663	
CG33193	sav	VALIUM12	JF01778	
CG3340	Kr	VALIUM12	JF01789	
CG3352	ft	VALIUM12	JF01689	
CG33991	nuf	VALIUM12	JF01642	
CG3936	N	VALIUM12	JF01575	
CG3969	PR2	VALIUM12	JF01647	
CG4005	yki	VALIUM12	JF01666	
CG4114	ex	VALIUM12	JF01687	
CG4656	Rassf	VALIUM12	JF01665	
CG6667	dl	VALIUM12	JF01862	
CG6716	prd	VALIUM12	JF01791	
CG6757	SH3PX1	VALIUM12	JF01760	
CG7793	Sos	VALIUM12	JF01555	
CG7935	msk	VALIUM12	JF02079	
CG8425	Jhe	VALIUM12	JF01669	
CG9885	dpp	VALIUM12	JF01372	
CG10260	CG10260	VALIUMp	JF01177	
CG10574	l-2	VALIUMp	JF01168	
CG14895	Pak3	VALIUMp	JF01175	
CG15305	flw	VALIUMp	JF01170	
CG31049	Doa	VALIUMp	JF01174	
CG4032	Abl	VALIUMp	JF01178	
CG5483	Lrrk	VALIUMp	JF01180	
CG6593	Pp1a-96A	VALIUMp	JF01171	
CG9115	mtm	VALIUMp	JF01115	
CG9156	Pp1-13C	VALIUMp	JF01169	
CG4006	Akt	pGD267		VDRC2902
CG1389	torso	pGD267		VDRC4298
CG3936	Notch	pGD267		VDRC27229
CG3218	K10	pGD267		VDRC102101
CG4257	Stat92E	pGD267		VDRC43866
CG3936	Notch	pUAST-R57		NIG3936R-2
CG4006	Akt	pUAST-R57		NIG4006R-1

Experiments were performed at both 25°C and 29°C. None of the lines tested showed oogenesis or embryonic phenotypes. VDRC and NIG lines correspond to lines obtained from the Vienna and National Institute of Genetics stock centers (http://stockcenter.vdrc.at/; http://www.shigen.nig.ac.jp/fly/nigfly), respectively.

Nature Methods: doi.10.1038/nmeth.1592

SUPPLEMENTARY NOTE 1:

RNAi via expression of long hairpins is not effective in the female germline

Previously, we reported the construction of long hairpin-based vectors, the "VALIUM series", and described in particular VALIUM1 and VALIUM10, that proved effective for RNAi in the soma[1,2] (**Supplementary Fig. 1a**). Both vectors contain *vermilion* as a selectable marker[3], an attB sequence to allow for targeted phiC31-mediated integration at genomic attP landing sites[4,5], two pentamers of UAS (one of which can be excised using the Cre/loxP system)[6] to generate a 5XUAS derivative for reduced expression levels[2]), the *hsp70* core promoter and a SV40 polyadenylation signal. An intronic sequence was placed between the two arms of the hairpin to facilitate hairpin RNA processing and nuclear export. A major difference between VALIUM1 and VALIUM10 is that the latter contains two gypsy insulator sequences to enhance transgene transcription[1]. Further differences concern the hairpin cloning strategy (multiple cloning site (MCS)-based for VALIUM1; recombination-based for VALIUM10) and an additional *ftz* intron upstream of the SV40 polyadenylation signal in VALIUM10. To test whether these vectors can drive transgene expression in the female germline, we used the *MTD-Gal4* line containing three different Gal4 insertions that drive expression at all stages of oogenesis[7] (see **Online Methods**). Although *MTD-Gal4* was able to drive expression of either *Luciferase* or *GFP* in a VALIUM1 and VALIUM10 vector (**Supplementary Figs. 1b,c**; data not shown for VALIUM1), we did not detect any germline or embryonic phenotypes when a number of long-hairpin transgenes were tested against genes associated with either oogenesis or maternal effect mutant phenotypes (data not shown; **Supplementary Table 3**). Our results extend a previous report showing that exogenously introduced long dsRNAs are ineffective in silencing target genes at certain stages of oogenesis[8].

To address the possibility that long-hairpin precursor transcripts were unstable in the germline, we added the 3'UTR of the maternally expressed *fs(1)K10* gene to VALIUM1 and VALIUM10[9] to generate VALIUM11 and VALIUM12. Both vectors carry the K10 3'UTR but differ in the presence or absence of the *ftz* 3'UTR intron (**Supplementary Fig. 1a**). Although this led to an overall increase in transgene expression, as determined by assaying a *Luciferase* marker (**Supplementary Fig. 1b**), it did not result in detectable germline or maternal effect phenotypes (data not shown; **Supplementary Table 3**). VALIUM11 and VALIUM12 were also much less effective at generating somatic RNAi phenotypes than VALIUM10, most likely because of the K10 3'UTR sequence (**Supplementary Fig. 7a,b**; data not shown for VALIUM11).

As VALIUM vectors possess the *hsp70* basal promoter that may not be optimal for germline expression[9], we tested a modified vector (VALIUMp) that contains both, the K10 3'UTR and the P-element transposase minimal promoter, which has been previously shown to efficiently drive expression in the female germline[9] (**Supplementary Fig. 1a**). VALIUMp was considerably more effective at driving either *Luciferase* or *GFP* in the germline (**Supplementary Fig. 1b,c**); however, the increased expression capacity did still not result in detectable phenotypes when long-hairpin constructs were expressed with *MTD-Gal4* (data not shown; **Supplementary Table 3**). As a final test of whether long dsRNAs could generate phenotypes in the female germline, we tested other commonly used vectors for somatic RNAi[10] (http://stockcenter.vdrc.at/control/main) (data not shown; **Supplementary Table 3**). As was observed with the VALIUM vectors, no phenotypes resulted from the expression of these long-hairpin constructs. Considered together, these results indicate that long dsRNAs are ineffective silencing triggers in female germ cells, even when delivered from vectors that can drive efficient expression of protein coding mRNAs.

Nature Methods: doi.10.1038/nmeth.1592

VALIUM20 is an effective vector for RNAi in both the female germline and the soma

Ovaries of *MTD-GAL4/UAS-shRNA-otu* or *MTD-GAL4/UAS-shRNA-bam* females showed ovarian tumor phenotypes that were morphologically identical to those associated with mutations in these genes (**Fig. 1b**). These phenotypes were fully penetrant at both 25°C and 29°C. The function of *otu* and *bam* is required in the germarium, indicating that VALIUM20 can effectively trigger RNAi during early oogenesis stages. VALIUM20 also proved effective at inducing RNAi at later stages. ShRNAs targeting *dl, tor* or *csw* led to the expected embryonic cuticle phenotypes (**Fig. 1c**). Depending on the shRNAs tested, the embryonic phenotypes were more severe when females with *MTD-Gal4* driven shRNAs were grown at 29°C rather than 25°C. For example, while *shRNA-dl* was fully penetrant both at 25°C and 29°C, approximately 5% of the embryos derived from *MTD-GAL4/UAS-shRNA-tor* hatched at 25°C but none hatched at 29°C. In the case of *shRNA-csw,* ~20% of the embryos hatched at 25°C while only ~2% hatched at 29°C. We also noticed that maternal age influenced phenotypic penetrance; eggs laid in the first 2-3 days following eclosion usually showed less penetrant phenotypes (data not shown).

To determine whether maternally loaded shRNAs can effectively knock down genes that are expressed zygotically, we generated shRNAs against a number of genes that result in embryonic lethality when mutated. For example, *decapentaplegic* (*dpp*) is expressed zygotically soon after fertilization and *dpp* mutant embryos show an almost complete replacement of the dorsal abdominal cuticle by ventral abdominal epidermal pattern elements[11]. This phenotype is solely dependent of the lack of zygotic *dpp* expression since the gene is not expressed in the female germline. Embryos derived from *MTD-GAL4/UAS-shRNA-dpp* mothers showed the characteristic *dpp* ventralization phenotype (**Fig. 1c**) demonstrating that maternally loaded shRNAs are effective at silencing mRNAs that are being transcribed following fertilization. As observed for other shRNAs, the phenotype was fully penetrant at 29°C but slightly weaker at 25°C.

Notch (*N*) is required for neurogenesis in early embryos. In the absence of zygotic *N*, embryos lack ventral cuticle due to hyperplasia of the nervous system[12]. In addition, *N* also has a maternal effect phenotype, as *N/+* embryos derived from *N* homozygous germline clones have a weak neurogenic phenotype. The neurogenic phenotype was observed in embryos derived from *MTD-GAL4/UAS-shRNA-N* females (**Fig. 1c**). Interestingly, *MTD-Gal4/UAS-shRNA-N* females could only be generated when *MTD-Gal4* males were crossed to *UAS-shRNA-N* females. Almost all *MTD-Gal4/UAS-shRNA-N* animals derived from *MTD-Gal4* females crossed to *UAS-shRNA-N* males died as embryos and showed a neurogenic phenotype or early larval stage lethality (data not shown). This strongly suggests an ability of maternally produced Gal4 to drive robust expression in the embryo of *UAS-shRNA-N* to levels sufficient to silence *N* transcripts. The neurogenic phenotype observed in embryos derived from *MTD-Gal4/UAS-shRNA-N* females was influenced by both temperature and maternal age. At 29°C the phenotype was 100% penetrant with all the embryos exhibiting a strong neurogenic phenotype. At 25°C, we observed a small fraction of hatching embryos during the first two days of egg laying (**Supplementary Fig. 8**). This age dependence may be caused by the first eggs made by females being produced faster than in older females[13], leading to a lower overall production and loading of maternal shRNAs.

We used *shRNA-N* and an shRNA against *white* (*shRNA-w)* to test the efficacy of VALIUM20 as a vector for RNAi in the soma. *UAS-shRNA-N*, in combination with the wing

Nature Methods: doi.10.1038/nmeth.1592

specific *C96-Gal4* driver, gave the wing phenotype that had been previously described[1,2] (**Fig. 1d**; data not shown). Notably, the phenotype was stronger than was previously achieved using a long hairpin in the optimized VALIUM10 vector (**Supplementary Fig. 3**). Similarly, when the *UAS-shRNA-w* line was tested with the eye specific *GMR-Gal4* driver, it generated an eye color phenotype similar to a complete null *white* mutation. This phenotype was again more severe than those generated previously using long dsRNA hairpins (**Fig. 1e**). To date, hundreds of shRNAs have been tested in the soma, and more than 90% generated the expected phenotypes (data not shown). Our combined results indicate that the expression of shRNAs from the VALIUM20 vector generates effective knockdown phenotypes in both the germline and the soma.

VALIUM22 is a superior vector for the female germline

In addition to the VALIUM22 experiments on silencing of piRNA pathway genes presented in **Fig. 2**, we tested the identical shRNAs expressed from the VALIUM20 vector. Very similar results were obtained upon depletion of Spn-E, Armi or Piwi using *MTD-Gal4* (**Supplementary Table 1**). Importantly, however, ovaries from flies expressing the *armi* or *piwi* shRNAs from VALIUM20 were rudimentary and resembled the phenotype of ovaries mutant for *piwi* or *armi* in germline and soma. This phenotype is highly suggestive of a significant depletion of Piwi or Armi in somatic support cells as *piwi* or *armi* are required in these cells for proper germline development[14,15]. We note that the identical shRNA sequences were used to generate the VALIUM20 and VALIUM22 constructs and that the identical genomic landing site was employed for the transgenes. Further, an independent shRNA expressed from VALIUM20 targeting a different region in *piwi* gave results identical to the previous *piwi* shRNA transgene. As homozygous *shRNA-piwi* or *shRNA-armi* lines are fertile, we speculate that low levels of Gal4 expression in the soma from *MTD-Gal4* trigger sufficient shRNA expression from VALIUM20 but not from VALIUM22 constructs. In support of this, shRNAs could be detected in carcasses (flies where the gonads were manually removed) when expressed from VALIUM20 using *MTD-Gal4* (**Supplementary Fig. 4**). Very low levels of processed shRNAs could even be detected in RNA samples prepared from the parental VALIUM20 shRNA fly stocks (**Supplementary Fig. 4**). We suspect that a temperature-sensitive element in the *hsp70* minimal promoter utilized in VALIUM20 allows for low-level expression that in combination with basal Gal4 levels leads to shRNA expression sufficient for gene knockdowns, at least in some cases. In summary, our results indicate that VALIUM22 is optimized for gene knockdowns and tissue specificity in the female germline, whereas VALIUM20 is favorable for silencing in somatic tissues.

Biogenesis and loading of shRNAs into effector complexes

To understand the genetic requirements of shRNA processing and loading, we depleted cultured *Drosophila* cells (Schneider/S2 cells) of components of the miRNA (Drosha, Pasha, Dcr-1, Loqs and AGO1) and siRNA (Dcr-2, Loqs, R2D2 and AGO2) pathways[16]. We compared the effects on shRNAs with those on a endogenous microRNA, *miR-bantam*, and the endogenous siRNA, *esi-2.1*. Knockdown of Drosha and Pasha simultaneously or depletion of Dcr-1 and AGO1 individually caused a significant reduction in levels of mature *miR-bantam* (**Supplementary Fig. 5a**). Depletion of Dcr-1 or Loqs caused a concomitant accumulation of the precursor miRNAs, whereas depletion of Dcr-2, R2D2 or AGO2 had no effect on levels of either precursor or mature *miR-bantam*. In contrast, depletion of Dcr-2, Loqs or AGO2 led to a

Nature Methods: doi.10.1038/nmeth.1592

substantial decrease of *esi-2.1* levels[16]. As expected due to their modeling onto an endogenous miRNA backbone, shRNAs behaved similar to *miR-bantam* with respect to knockdown of Drosha, Pasha, and Dcr-1 (effects of Loqs were too weak to reach a meaningful conclusion).

AGO1 and AGO2 both accept small RNAs from dsRNA precursors. However, they differ in their biochemical properties and in their bound populations of endogenous RNAs[17,18,19,20,21]. In addition, small RNAs that join AGO2 are modified at their 3' ends by the methyltransferase Hen1, making them resistant to β-elimination[22,23], which can be illustrated by the differential sensitivity of *esi-2.1*, which is AGO2-bound, and *miR-bantam*, which occupies AGO1. A substantial fraction of mature shRNAs resists β-elimination but becomes susceptible after depletion of AGO2 (**Supplementary Fig. 5a**). This suggests that AGO2 serves as the destination for the majority of shRNA strands. Support for this conclusion came from examination of AGO1 and AGO2 complexes (**Supplementary Fig. 5b**), although a substantial portion of shRNAs were also detected in AGO1 immunoprecipitates. Thus, although shRNAs are produced by the microRNA biogenesis machinery, they are efficiently loaded into AGO2, presumably by the canonical siRNA loading machinery consisting of Dcr-2 and R2D2. This is consistent with several recent reports of hierarchical loading rules for small RNAs in *Drosophila*, which predict that many of the shRNAs analyzed should show a preference for AGO2[17,18,19,20,21].

Abundance of shRNAs and processing accuracy

To reliably suppress their intended targets, shRNAs must be precisely and efficiently processed from their artificial precursor transcripts. In particular, the 5' end, a major determinant of target recognition via small RNAs must be predictable, so that design algorithms can aid in choosing potent shRNAs. While modeling on the *miR-1* backbone created some expectations of specific processing sites, this had to be tested explicitly in the remodeled constructs. We therefore transfected a number of different shRNA constructs into cultured S2 cells and sequenced the small RNA populations from these cells. The vast majority of shRNAs in these libraries generated from total RNAs (19- to 24-nt) were 22-nt in size (**Supplementary Fig. 6a**). Importantly, the guide shRNA strands derived from the 3p arm of the hairpin and their respective 5' ends precisely corresponded to the expected products, indicating accurate Dcr-1 cleavage. The 3' ends of shRNAs show slight variation, similar in extent to endogenous miRNAs. Furthermore, guide shRNA strands derived from the 3p arm were invariably higher in abundance than their passenger counterparts from the 5p arm (**Supplementary Fig. 6a,b**). In order to evaluate cellular shRNA levels, we compared their abundances and strand biases with those of the 30 most abundant endogenous miRNAs from three independent transfections. We found shRNA guide strand levels comparable to those of highly abundant microRNAs (**Supplementary Fig. 6b**), while their passenger strands ranked similarly to miR* strands. Also, strand selection of shRNAs parallels that of most miRNAs with strong biases towards the guide/miR strands (**Supplementary Fig. 6b**). Considered together, these data indicate that placing a sequence perfectly complementary to the target of interest into the 3p arm of the *miR-1* backbone (together with a suitable 5p arm sequence) leads to efficient and accurate production of the intended small RNA.

Nature Methods: doi.10.1038/nmeth.1592

SUPPLEMENTARY NOTE 2: Vector construction

VALIUM11 and VALIUM12: To construct VALIUM11, the *SV40* polyA signal of *VALIUM10*[1] was replaced with fill-in nucleotides (fwd: 5'-AATTGAACCGCGGAATCGATTCTGCAGTTGAGCT-3', rev: 5'-CAACTGCAGAATCGATTCCGCGGTTC-3') using SacI and SacII. The vector was then cut with MfeI and PstI, and filled with a 1.7kb *K10* 3'UTR from *UASp*[9]. To construct VALIUM12, the *ftz* intron was amplified with specific primers (fwd: 5'-CCTCTAGAGAATTGTTGGCATCAGGTAGG-3', rev: 5'-TTCAATTGCCGCGGCTCTAGTTCTTTG-3') from VALIUM1[2]. The PCR product was cut with XbaI and MfeI, and then cloned into VALIUM11.

VALIUMp: The *SV40* polyA signal of VALIUM1 w*as* replaced with fill-in nucleotides (fwd: 5'-AATTGAACCGCGGAATCGATTCTGCAGTTGAGCT-3', rev: 5'-CAACTGCAGAATCGATTCCGCGGTTC-3') using SacI and MfeI. The resulting vector was cut with MfeI and PstI, and a 1.7kb *K10* 3'UTR was inserted. The P-element transposase promoter from *UASp* was amplified (fwd: 5'-TCGTCGACAGCCGTAGCTTACCGAAGTATAC-3', rev: 5'-CTGAATTCTGATCCCCGGGCGGGTACCA-3'), the PCR product was cut with SalI and EcoRI, and then cloned into the previous vector.

Luciferase and GFP constructs: To generate the *Luciferase* and *GFP* VALIUM constructs, VALIUM1-*Luciferase* and VALIUM1-*GFP*[1] were cut with EcoRI and XbaI. A 1.8kb DNA fragment containing the *Luciferase* coding region and the small fragment carrying *GFP* were gel purified, and subsequently cloned into VALIUM10, VALIUM11, VALIUM12 and VALIUMp using the same restriction sites. For the *Luciferase* assay, ovaries were dissected in 1XPBS and pooled into groups of 5 in 50 uL Glo Lysis Buffer (Promega). Ovaries were stored frozen at -80^{o}C until further use. Following thawing and homogenizing with an eppendorf pestle, *Luciferase* readings were measured using the Steady Glo Luciferase Assay System (Promega).

VALIUM20: To construct VALIUM20, VALIUM10 was cut with EcoRI and XbaI resulting in five fragments. The largest fragment was gel purified and ligated with an oligonucleotide fragment generated by annealing the two primers (fwd: 5'-AATTGAGATCTGTTGTAGAGTGGACATATGCACCTAGGA-3', rev: 5'-CTAGTCCTAGGTGCATATGTCCACTCTACAACAGATCTC-3'), this resulted in an intermediate vector. pNE3 (gift from Benjamin Haley) was cut with XbaI and NdeI which produced two fragments. The small fragment was gel purified and cloned into the aforementioned intermediate vector that was linearized with XbaI and NdeI.

VALIUM22: To construct VALIUM22, VALIUM2[1] was cut with EcoRI and BamHI. The fragment containing the *P*-element transposase promoter was cloned into pre-linearized VALIUM12. A DNA fragment containing the *miR-1* scaffold was obtained by PCR (fwd: 5'-AATTGAGATCTGTTGTAGAGTG-3', rev: 5'-CTAGGTGCATATGTCCACTCT-3'), and then cloned into the previous vector, which was cut with EcoRI and XbaI, to yield VALIUM22.

shRNA construct: The following steps were used to design and construct the shRNAs:

1. Selection of the 21-nt sequence based on the algorithm of Vert et al.[24];

2. The oligonucleotide design eliminates off target effect at 16nt;

3. Based on *miR-1* scaffold, for the top strand oligo, add ctagcagt to 5' end of passenger strand DNA, add tagttatattcaagcata between passenger strand DNA and guide strand DNA, add gcg to 3' end of guide strand DNA, so the resulting oligo will be:

Nature Methods: doi.10.1038/nmeth.1592

5’-ctagcagtNNNNNNNNNNNNNNNNNNNNtagttatattcaagcataNNNNNNNNNNNNNNNNNNNNgcg-3’;

4. For the bottom strand oligo, add aattcgc to 5’ end of guide strand DNA, add tatgcttgaatataacta between guide strand DNA and passenger strand DNA, add actg to 3’ end of passenger strand DNA, so the resulting oligo will be:

5’-aattcgcNNNNNNNNNNNNNNNNNNNNtatgcttgaatataactaNNNNNN NNNNNNNNNNNNNactg-3’;

5. Annealing top strand with bottom strand oligos, the resulting DNA fragment has overhangs for NheI and EcoRI;

6. Directly clone this DNA fragment into VALIUM20 vector that had been linearized by NheI and EcoRI. Bacteria string TOP10 cells were used as competent cells;

7. PCR select correct clone, and the primers we used are:

fwd: 5'-ACCAGCAACCAAGTAAATCAAC-3'

rev: 5'-TAATCGTGTGTGATGCCTACC-3';

8. DNA sequencing to confirm correct shRNA construct, and the sequencing primer is: 5’-ACCAGCAACCAAGTAAATCAAC-3’.

Nature Methods: doi.10.1038/nmeth.1592

SUPPLEMENTARY NOTE 3: Primer sequences

Northern Blotting: The sequences of the oligonucleotide probes are:

esi-2.1	5'-GGAGCGAACTTGTTGGAGTCAA-3'
miR-bantam	5'-AATCAGCTTTCAAAATGATCTCA-3'
2S rRNA	5'-TACAACCCTCAACCATATGTAGTCCAAGCA-3'
shRNA-E(bx)	5'-CAGCTTGTGGTTCAACAACAA-3'
shRNA-N	5'-CGCGGCGGTTAACAATACCGAA-3'

Transposon qPCR analysis: The sequences of the oligonucleotides used are:

nos-fwd:	5'-GCAACTTAATGCCCATTCCAC-3'
nos-rev:	5'-CGGCTGGTATATACGACATGT-3'
rp49-fwd:	5'-CCGCTTCAAGGGACAGTATCTG-3'
rp49-rev:	5'-ATCTCGCCGCAGTAAACGC-3'
HeT-A-fwd:	5'-CGCCGCAGTCGTTTGGTGAGT-3'
HeT-A-rev:	5'-CGCGCGGAACCCATCTTCAGA-3'
blood-fwd:	5'-CCAACAAAGAGGCAAGACcG-3'
blood-rev:	5'-TCGAGCTGCTTACGCATACTGTC-3'

Nature Methods: doi.10.1038/nmeth.1592

REFERENCES:

1. Ni, J. Q. *et al.*, *Genetics* **182**, 1089-1100 (2009).
2. Ni, J. Q. *et al.*, *Nat Methods* **5**, 49-51 (2008).
3. Fridell, Y. W. and Searles, L. L., *Nucleic Acids Res* **19**, 5082 (1991).
4. Thomason, L. C., Calendar, R., and Ow, D. W., *Mol Genet Genomics* **265**, 1031-1038 (2001).
5. Groth, A. C., Fish, M., Nusse, R., and Calos, M. P., *Genetics* **166**, 1775-1782 (2004).
6. Siegal, M. L. and Hartl, D. L., *Genetics* **144**, 715-726 (1996).
7. Petrella, L. N., Smith-Leiker, T., and Cooley, L., *Development* **134**, 703-712 (2007).
8. Kennerdell, J. R., Yamaguchi, S., and Carthew, R. W., *Genes Dev* **16**, 1884-1889 (2002).
9. Rorth, P., *Mech Dev* **78**, 113-118 (1998).
10. Dietzl, G. *et al.*, *Nature* **448**, 151-156 (2007).
11. Irish, V. F. and Gelbart, W. M., *Genes Dev* **1**, 868-879 (1987).
12. Jiménez, F., Campos-Ortega, J. A. , *Wilhelm Roux's Archives* **191** (1982).
13. King, R. C., *Ovarian Development in Drosophila melanogaster.* . (Academic Press, New York. , 1970).
14. Cook, H. A., Koppetsch, B. S., Wu, J., and Theurkauf, W. E., *Cell* **116**, 817-829 (2004).
15. Cox, D. N. *et al.*, *Genes Dev* **12**, 3715-3727 (1998).
16. Czech, B. *et al.*, *Nature* **453**, 798-802 (2008).
17. Tomari, Y., Du, T., and Zamore, P. D., *Cell* **130**, 299-308 (2007).
18. Forstemann, K. *et al.*, *Cell* **130**, 287-297 (2007).
19. Okamura, K., Liu, N., and Lai, E. C., *Mol Cell* **36**, 431-444 (2009).
20. Czech, B. *et al.*, *Mol Cell* **36**, 445-456 (2009).
21. Ghildiyal, M. *et al.*, *RNA* **16**, 43-56 (2010).
22. Horwich, M. D. *et al.*, *Curr Biol* **17**, 1265-1272 (2007).
23. Saito, K. *et al.*, *Genes Dev* **21**, 1603-1608 (2007).
24. Vert, J. P., Foveau, N., Lajaunie, C., and Vandenbrouck, Y., *BMC Bioinformatics* **7**, 520 (2006).

Nature Methods: doi.10.1038/nmeth.1592

shutdown is a component of the *Drosophila* piRNA biogenesis machinery

Jonathan B. Preall *, **Benjamin Czech***, Paloma M. Guzzardo, Felix Muerdter and Gregory J. Hannon

* authors contributed equally to this work

RNA. 2012 Aug;18(8):1446-57. Epub 2012 Jul 2.

REPORT

shutdown is a component of the *Drosophila* piRNA biogenesis machinery

JONATHAN B. PREALL,[1] BENJAMIN CZECH,[1] PALOMA M. GUZZARDO, FELIX MUERDTER, and GREGORY J. HANNON[2]

Howard Hughes Medical Institute, Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

ABSTRACT

In animals, the piRNA pathway preserves the integrity of gametic genomes, guarding them against the activity of mobile genetic elements. This innate immune mechanism relies on distinct genomic loci, termed piRNA clusters, to provide a molecular definition of transposons, enabling their discrimination from genes. piRNA clusters give rise to long, single-stranded precursors, which are processed into primary piRNAs through an unknown mechanism. These can engage in an adaptive amplification loop, the ping-pong cycle, to optimize the content of small RNA populations via the generation of secondary piRNAs. Many proteins have been ascribed functions in either primary biogenesis or the ping-pong cycle, though for the most part the molecular functions of proteins implicated in these pathways remain obscure. Here, we link *shutdown* (*shu*), a gene previously shown to be required for fertility in *Drosophila*, to the piRNA pathway. Analysis of knockdown phenotypes in both the germline and somatic compartments of the ovary demonstrate important roles for *shutdown* in both primary biogenesis and the ping-pong cycle. *shutdown* is a member of the FKBP family of immunophilins. Shu contains domains implicated in peptidyl-prolyl *cis-trans* isomerase activity and in the binding of HSP90-family chaperones, though the relevance of these domains to piRNA biogenesis is unknown.

Keywords: piRNAs; transposon silencing; RNAi; FKBP; germ cells

INTRODUCTION

Eukaryotic genomes are prone to the accumulation of repetitive sequences, including transposable elements, over evolutionary time (McClintock 1953; Kim et al. 1994; Brennecke et al. 2007; Chambeyron et al. 2008; Feschotte 2008). The genomic instability brought about by transposon activity is a double-edged sword. Low levels of transposition can drive evolution in the long term, but loss of control over mobile elements in any individual can threaten reproductive success. Mechanisms for suppressing transposon activation in the germline are therefore both potent and widely conserved (Grimson et al. 2008). In animals, the PIWI-interacting RNA (piRNA) pathway is key to transposon silencing in reproductive tissues (Aravin et al. 2006; Girard et al. 2006; Lau et al. 2006; Vagin et al. 2006; Malone and Hannon 2009; Khurana and Theurkauf 2010; Senti and Brennecke 2010). In *Drosophila*, piRNAs are active both in the germ cell lineage and in a particular somatic lineage that encysts the germ cells and provides growth and maturation signals (Malone et al. 2009).

piRNA clusters sit at the apex of the pathway and, based upon their sequence content, define transposon targets for repression (Brennecke et al. 2007). piRNA clusters give rise to long, single-stranded transcripts (Brennecke et al. 2007) that are thought to be exported to the cytoplasm and processed into primary piRNAs, most likely in specialized cytoplasmic structures (Saito et al. 2010; Handler et al. 2011). A number of proteins have been implicated in primary piRNA biogenesis and their loading into PIWI-family proteins, including Armitage, Zucchini, Vreteno, and the Yb family (Klattenhoff et al. 2007; Pane et al. 2007; Malone et al. 2009; Szakmary et al. 2009; Haase et al. 2010; Olivieri et al. 2010; Saito et al. 2010; Handler et al. 2011; Zamparini et al. 2011). Yet, almost nothing is known about how each of these promotes the production of primary piRNAs.

The soma relies on a single piRNA cluster, *flamenco* (*flam*) (Brennecke et al. 2007). This ~180 kb, centromere-proximal locus on the X chromosome produces a piRNA population that is strongly enriched for species antisense to the *gypsy* family elements. These elements are active in follicle cells and can propagate by infection of germ cells through

[1]These authors contributed equally to this work.

[2]**Corresponding author**
E-mail hannon@cshl.edu

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.034405.112.

their capability to form virus-like particles (Pelisson et al. 1994; Chalvet et al. 1999). Somatic piRNAs are produced solely through primary biogenesis (Brennecke et al. 2007; Malone et al. 2009). In the germline, a greater variety of clusters targets a broad spectrum of mobile elements and engages an adaptive cycle, termed ping-pong, through which transposon mRNAs help to shape piRNA populations (Brennecke et al. 2007; Gunawardane et al. 2007). Here, antisense-oriented piRNAs derived from genomic clusters are loaded into Aubergine (Aub) and cleave active transposable element transcripts in an RNAi-like reaction. Unlike classical RNAi, this triggers the production of a new small RNA, derived from the target mRNA and with its 5′ end formed by Aub-mediated cleavage. The new, secondary piRNA is loaded into Ago3, which can then use this sense-oriented species to recognize and cleave cluster-derived transcripts, producing more antisense piRNAs via a similar mechanism.

Mutations in the *Drosophila* piRNA pathway generally result in sterility with stereotypical phenotypes in the male and female germline (Schupbach and Wieschaus 1991; Wilson et al. 1996; Gonzalez-Reyes et al. 1997; Lin and Spradling 1997; Cox et al. 2000; Cook et al. 2004). In part, these are thought to result from DNA double-strand breaks induced by element activity (Chen et al. 2007; Klattenhoff et al. 2007). Such breaks trigger meiotic checkpoint activation mediated by the *Drosophila chk2* ortholog, *loki*, which in turn disrupts dorsal-ventral axis formation during oogenesis. Hence, mutations in secondary piRNA genes such as *aubergine* display fused dorsal appendages and other hallmarks of oocyte ventralization (Theurkauf et al. 2006). Transposon silencing is also critical for the maintenance of germline stem cells (Lin and Spradling 1997; Cox et al. 2000; Houwing et al. 2007). In the male germline, loss of *Su(ste)* piRNAs derepresses the repetitive *Stellate* locus, which disrupts spermiogenesis by causing the overproduction and eventual crystallization of Stellate protein within the testis (Bozzetti et al. 1995; Aravin et al. 2001). Several mutants that are now known to affect the *Drosophila* piRNA pathway—including *aubergine*, *zucchini*, *squash*, *vasa*, and *cutoff*—were first described in a female sterility screen by Schüpbach and Wieschaus over 20 yr ago (Schupbach and Wieschaus 1989, 1991). Of the genes identified in that study that would eventually come to be known as piRNA factors, all but *cutoff* were classified phenotypically as having defects in dorsal appendage formation (Schupbach and Wieschaus 1991).

Subsequently, Munn and Steward (2000) mapped another of these female sterile mutations, *shutdown (shu, CG4735),* to an immunophilin gene of the FK506-binding protein (FKBP) family. Mutations in *shu* disrupt germ cell division, eventually causing the germline stem cells to fail entirely. Two strong alleles caused sterility in both males and females, while a third point mutant allele did not affect male fertility. In mutant females, stem cells that successfully divide generally produce faulty egg chambers that arrest mid-oogenesis. Germline clones for strong alleles of *shu* can produce mature oocytes, though they display typical patterning defects such as fused dorsal appendages. Considered together, these observations implicate *shu* as a component of the *Drosophila* piRNA pathway. This conjecture is supported by the presence of FKBP6, the mammalian protein most similar to Shutdown, in complexes with mammalian Piwi-family proteins, Miwi and Miwi2 (Vagin et al. 2009).

FKBPs play diverse biological roles ranging from facilitating protein folding to modulating transport (Ahearn et al. 2011), receptor signaling (Li et al. 2011), and meiotic recombination (Crackower et al. 2003; Kang et al. 2008). The FKBP domain is annotated as a peptidyl-prolyl *cis-trans* isomerase (PPIase), though there are many instances of well-conserved FKBP domains that lack PPIase activity (Gollan and Bhave 2010). The macrolide immunosuppressants FK506 (tacrolimus) and rapamycin (sirolimus) bind with sub-nanomolar affinity to the FKBP domain and block a key protein–protein interaction surface, but as is the case with PPIase activity, many family members display much reduced affinities for these compounds (DeCenzo et al. 1996; Gollan and Bhave 2010).

FKBP-class immunophilins display a variety of domain architectures. One arrangement, conserved from protozoa to humans, places a tetratricopeptide repeat (TPR) domain downstream from the FKBP domain (Pratt et al. 2004). The TPR domain is a protein–protein interaction module that binds heat shock proteins (HSPs), primarily of the HSP90 family in higher eukaryotes (Pratt 1998; Allan and Ratajczak 2011). Several crystal structures are available that highlight key conserved residues that participate in this interaction (Van Duyne et al. 1993; Ward et al. 2002). Connections between small RNA silencing pathways and HSP activity have been observed in several model systems (Smith et al. 2009). In particular, RNA-induced silencing complex (RISC) loading is facilitated by HSP90 and ATP hydrolysis (Iki et al. 2010; Iwasaki et al. 2010; Miyoshi et al. 2010; Iki et al. 2011).

Here, we report that *shutdown* is a critical element of the *Drosophila* piRNA pathway. Tissue-specific depletion of Shu results in derepression of transposon expression and a near-complete loss of mature piRNAs in both the somatic and germline lineages. Shu is cytoplasmically localized, and its loss disrupts the localization of all three piRNA effectors, Piwi, Aub, and Ago3. We hypothesize that Shu is an essential component of both primary and ping-pong–derived piRNA biogenesis, likely acting at a very early step that is shared between both piRNA systems.

RESULTS AND DISCUSSION

Clues to a role for FKBPs in the piRNA pathway

We previously carried out a proteomic analysis of mammalian PIWI proteins, Miwi and Miwi2 (Vagin et al. 2009). Among the components of these complexes were murine

FKBP6 and multiple HSPs. FKBP5 was also detected in Miwi immunoprecipitates with roughly half the coverage seen for FKBP6. Given the greater convenience of manipulating the piRNA pathway in *Drosophila*, we chose to examine potential roles for FKBP proteins in that model system.

The *Drosophila* genome encodes eight FKBP family members (Fig. 1A,B). Three, CG1847, CG5482, and FKBP59, are annotated to share the domain architecture of FKBP6, with their FKBP domains followed by a TPR. Shutdown is a potential fourth member of this group. Though its TPR

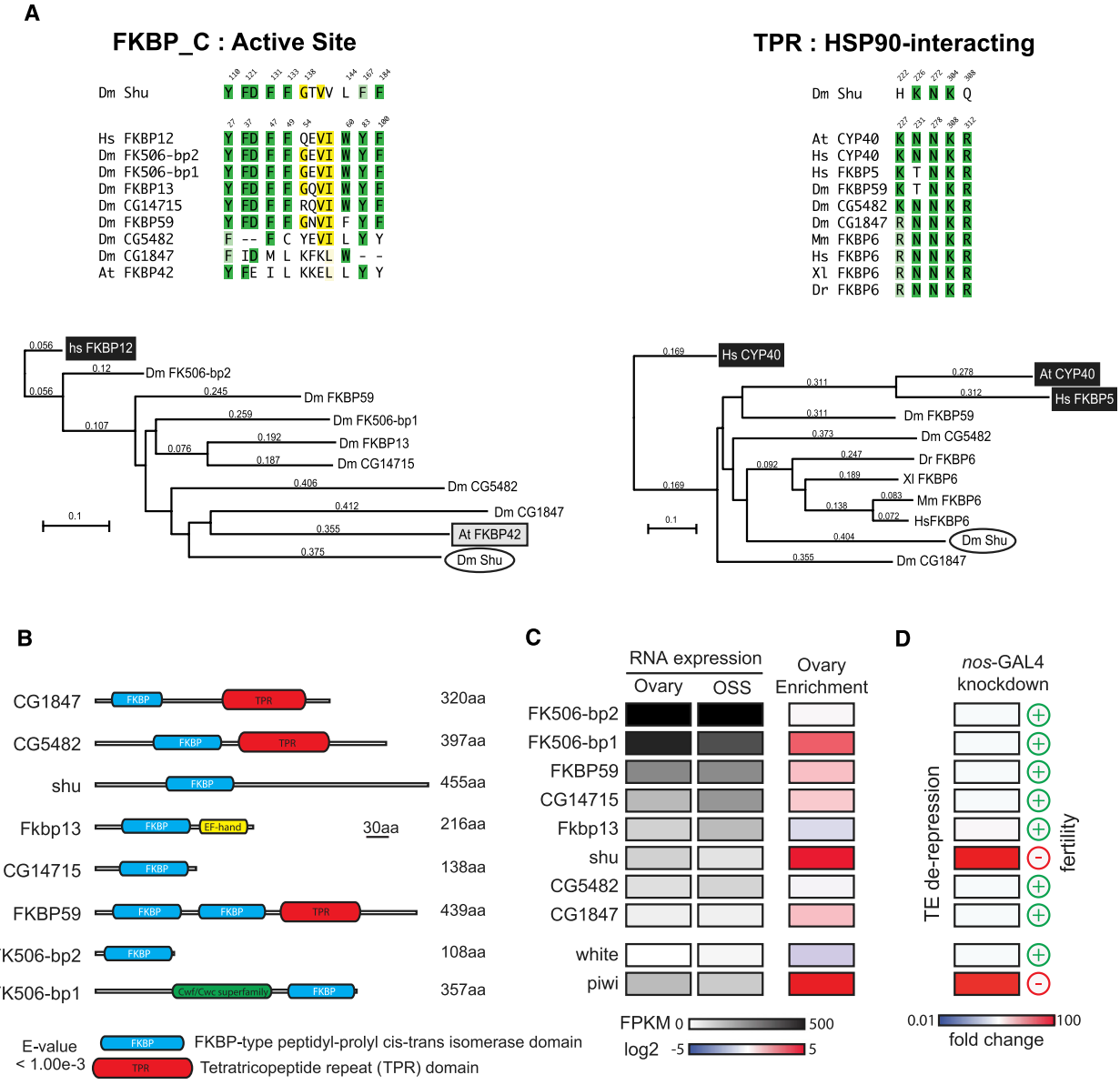


FIGURE 1. Shutdown is the only FKBP-family protein required for transposon silencing. (*A*) *Above* are shown the critical residues for the FKBP family peptidyl-prolyl *cis-trans* isomerase active site and the HSP90-interacting region in the TPR, as indicated, comparing the site in *Drosophila* Shutdown with those present in other family members. Residues in green indicate highly conserved residues with a known impact on PPIase activity, while those in yellow indicate a more poorly conserved region that has also been implicated. *Below* are evolutionary trees comparing each domain with family members present in other species. (*B*) The domain structures of the eight *Drosophila* FKBP family members are shown schematically. (*C*) Relative expression levels of *Drosophila* FKBP family members are shown for ovary and OSS RNAseq data sets. Relative enrichment in ovary versus other tissues is also shown. (*D*) Shown are relative *HetA* expression levels detected in ovaries from *Drosophila* engineered to express dsRNAs corresponding to each family member in the germline lineage. To the *right* is indicated whether dsRNA-expressing females are fertile (+) or sterile (−).

domain has substantially diverged in comparison to its paralogs (TPR_2, Pfam e-value = 0.0026), secondary structure predictions using the Phyre2 algorithm (www.sbg.bio.ic.ac.uk/phyre2) annotate the putative TPR as such with high confidence. Among *Drosophila* FKBPs, Shutdown is most closely related to FKBP6 ($E = 2 \times 10^{-36}$) overall, whereas FKBP59 is a potential FKBP5 ortholog ($E = 2 \times 10^{-46}$).

An examination of RNAseq both from the *Drosophila* ovarian somatic sheet (OSS) cell line and from a published ovarian data set revealed that several FKBPs are expressed in female reproductive tissues (Fig. 1C; Gan et al. 2010). A broader set of published microarray data (Chintapalli et al. 2007) suggested that the expression pattern of *shu* is much more biased to the ovary than is expression of other family members (Fig. 1C), a bias shared by many piRNA pathway components.

The FKBP_C domain is broadly conserved across evolution, though its PPIase activity is not (Kamphausen et al. 2002). Phylogenetic comparison of *Drosophila* FKBP_C domains to known active (*Homo sapiens* FKBP12) and inactive (*Arabidopsis thaliana* FKBP42) PPIase domains suggested that *shu* is more similar to inactive variants (Fig. 1A, bottom). Shutdown does retain more of the active site residues shown to be essential for PPIase activity in human FKBP12 (Fig. 1A, left) than does *AtFKBP42*. A conserved tryptophan residue (W60 in *Hs*Fkbp12) has been replaced by a leucine in Shutdown. Introduction of this change into Fkbp12 reduces PPIase activity by approximately eightfold and rapamycin and FK506 binding affinity by 10- and 75-fold, respectively. It is therefore likely that Shutdown does not represent an optimally active PPIase and may instead utilize the domain as a protein interaction interface, as do other FKBP family members (Gollan and Bhave 2010).

The Shu TPR domain is less well conserved and, in fact, shows little similarity to other TPRs known to bind HSP90 (Fig. 1A, right). In particular, nonconservative amino acid changes at two key residues suggested that the affinity of this domain for the C-terminal MEEVD of HSP90 is likely to be dramatically reduced compared with other family members (Ratajczak and Carrello 1996; Ward et al. 2002). Still, a *shu* allele (shu^{PB70}) bearing a point mutation at a nonconserved residue in the putative TPR is sufficient to cause female sterility, indicating that this region is essential for some aspects of Shu function.

Shutdown is implicated in transposon silencing

Recent work has suggested that Dcr-2 is a limiting factor that prevents conventional dsRNA triggers from inducing potent RNAi in *Drosophila* germ cells, but that this restriction could be overcome by enforced Dcr-2 expression (Handler et al. 2011; Wang and Elgin 2011). We took advantage of this observation by bringing UAS-driven dsRNA constructs from the Vienna *Drosophila* RNAi Center (VDRC) into a background containing a germline-specific GAL4-driver ({GAL4-nos.NGT}40; aka *nos*-GAL4) and a UAS-*Dcr-2* transgene. Among dsRNAs targeting all fly FKBP proteins, only those corresponding to *shu* had significant impacts on expression levels of the *HetA* transposon (Fig. 1D). Moreover, only dsRNAs targeting Shu caused female sterility, a property typical of piRNA mutants (Fig. 1D).

To validate *shu* as a novel piRNA pathway component, we compared the impact of its depletion to knockdowns of known piRNA pathway genes, *armi* and *piwi*. Germline silencing of each gene resulted in a similar level of derepression for 17 transposons, measured by quantitative PCR (qPCR) (Fig. 2A). The tissue specificity of our knockdown strategy was supported by the fact that germline-specific, telomeric transposons *TAHRE*, *HetA*, and *TART* were the most heavily derepressed (greater than 150-fold, $P < 0.01$), whereas RNA levels for primarily somatic elements, such as *ZAM,* remained unchanged (about 1.2- to 1.5-fold).

Shu RNAi also recapitulated the ventralized egg phenotype of shu^{PB70} germline clones, as evidenced by a high incidence of fused or abnormal dorsal appendages (Fig. 2B; Munn and Steward 2000). Surprisingly, the ventralization phenotype was not penetrant in *armi* and *piwi* knockdowns eggs, despite the eggs being nonviable (Fig. 2C). For *armi*, prior studies of mutants produce a clear expectation of ventralization upon potent knockdown (Klattenhoff et al. 2007; Orsi et al. 2010). For *piwi*, the prediction is less clear. Germline *piwi* clones were reported not to show this distinctive phenotype; however, RNAi-mediated *piwi* knockdown did produce eggs with a spindle morphology (Cox et al. 2000; Wang and Elgin 2011). In addition to causing sterility, *shu* depletion also reduced the number of nonviable eggs laid, suggesting that there may be additional requirements for *shu* function outside of piRNA-mediated transposon silencing.

Shu is required for Piwi, Aub, and Ago3 localization

In wild-type tissues, Piwi is localized to the nucleus of germline and somatic cells (Cox et al. 2000; Saito et al. 2006; Brennecke et al. 2007). Aub and Ago3 are expressed exclusively in the germline and are enriched in a perinuclear organelle called nuage (Lim and Kai 2007; Li et al. 2009). Proper localization depends upon normal piRNA production and loading into PIWI family proteins, and disruption of this pattern is an indicator of impaired biogenesis (Malone et al. 2009).

Depletion of *shu* using the *nos*-GAL4 driver resulted in redistribution of Piwi from nurse cell nuclei to the syncytial cytoplasm of the developing egg chamber, while neighboring somatic follicle cells retain proper nuclear Piwi localization (Fig. 2D). Similarly, the ping-pong factors Ago3 and Aub were redistributed from nuage to cytoplasmic foci, while the localization of the core nuage component Vasa was not altered (Fig. 2D). Driving the *shu* dsRNA using

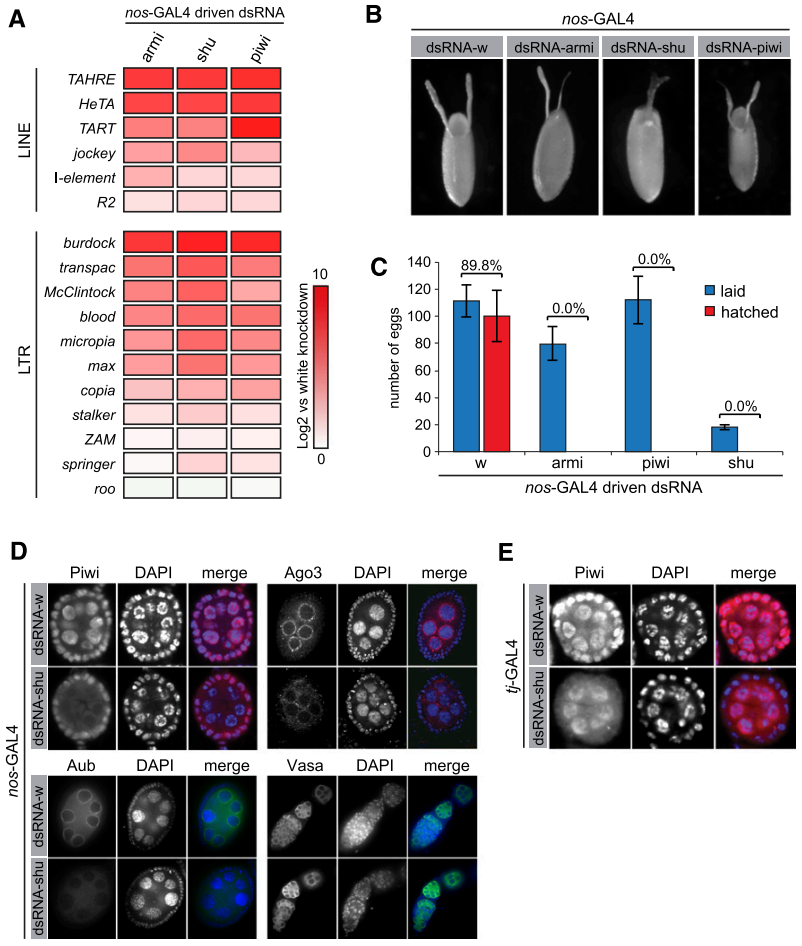


FIGURE 2. Phenotypes of *Drosophila* with germline-specific *shu* knockdown. (*A*) Depletion of *shu* in the germline results in derepression of multiple, unrelated transposons from the LINE and LTR families. Derepression, relative to *white* RNAi, is displayed as log2 fold change in heat map form. Analysis of flies with germline knockdown of Armi and Piwi, two known piRNA components, is displayed for comparison. (*B*) Germline-knockdown of *shu* causes patterning defects as indicated by the presence of fused dorsal appendages. (*C*) Depletion of *shu* causes female sterility. *shu* RNAi females lay fewer eggs compared with controls or animals depleted of other piRNA pathway factors. Hatching rates for all knockdown animals are zero, indicating complete sterility. (*D*) Depletion of *shu* in the germline using *nos*-GAL4 results in Piwi delocalization from nuclei and in Aub and Ago3 delocalization from nuage. Vasa localization is not changed. Depletion of *white* is shown as control. (*E*) *Tj*-GAL4–driven knockdown of *shu* in somatic follicle cells also causes Piwi delocalization. RNAi against *white* is shown as control.

GAL4 expressed from the soma-specific *traffic jam* promoter (*tj*-GAL4) caused delocalization of Piwi from the nuclei of follicle cells, while germline Piwi remained unaffected (Fig. 2E).

Despite its effects on the localization of PIWI-family proteins, we found that the bulk of Shutdown was not associated with domains characteristic of those piRNA pathway components. We generated N- and C-terminal GFP fusions of Shu expressed under the control of the ubiquitous *Actin5c* promoter. We examined the localization of Shutdown fusion proteins by transfection of OSS cells. Control constructs showed the expected localization with GFP-Piwi accumulating in nuclei and with GFP-Armi showing strong perinuclear localization consistent with its association with Yb-bodies. Zucchini features sequence homology with phospholipase D and was reported to localize to the outer membrane of mitochondria. In our studies, it displayed considerable overlap with the mitochondrial stain MitoTracker CMXRos (Supplemental Fig. S2). While cytoplasmic foci of GFP-tagged Shu were visible using both N- and C-terminal constructs, they did not overlap with the previously characterized localization patterns of other piRNA pathway proteins. Considered together, these data indicate that Shu is neither enriched in known structures associated with silencing nor required for assembly of a core nuage component.

Shu is essential for accumulation of both primary and secondary piRNAs

Strong derepression of germline and somatic transposons and the loss of characteristic localization patterns for Piwi-family proteins suggested that *shu* might function as a core piRNA biogenesis component, similar to *armi*. To address this possibility, we cloned and sequenced small RNAs from ovaries in which we drove the expression of *white* (*w*), *shu*, and *piwi* dsRNAs in the germline (*nos*-GAL4) or soma (*tj*-GAL4), as described above. Germline small RNAs libraries were normalized using the number of unique reads mapping to the *flam* locus, which is unaffected by germline-specific knockdowns. Germline-specific *shu* knockdown dramatically reduced the observed piRNA population compared with the *white* knockdown control. Small RNA reads with the characteristic piRNA size (23–29 nucleotides [nt]) mapping to the germline-specific, dual-strand *42AB* cluster were reduced 11.4-fold overall (8.2× on plus strand, 14.4× on minus strand). In contrast, *piwi* knockdown produced only a 2.8-fold overall reduction (2.8× on plus strand, 2.7× on minus strand) (Fig. 3A).

The incomplete loss of piRNAs in the *piwi* knockdown likely reflects the fact that piRNAs from *42AB* are normally loaded into each of the three *Drosophila* PIWI proteins

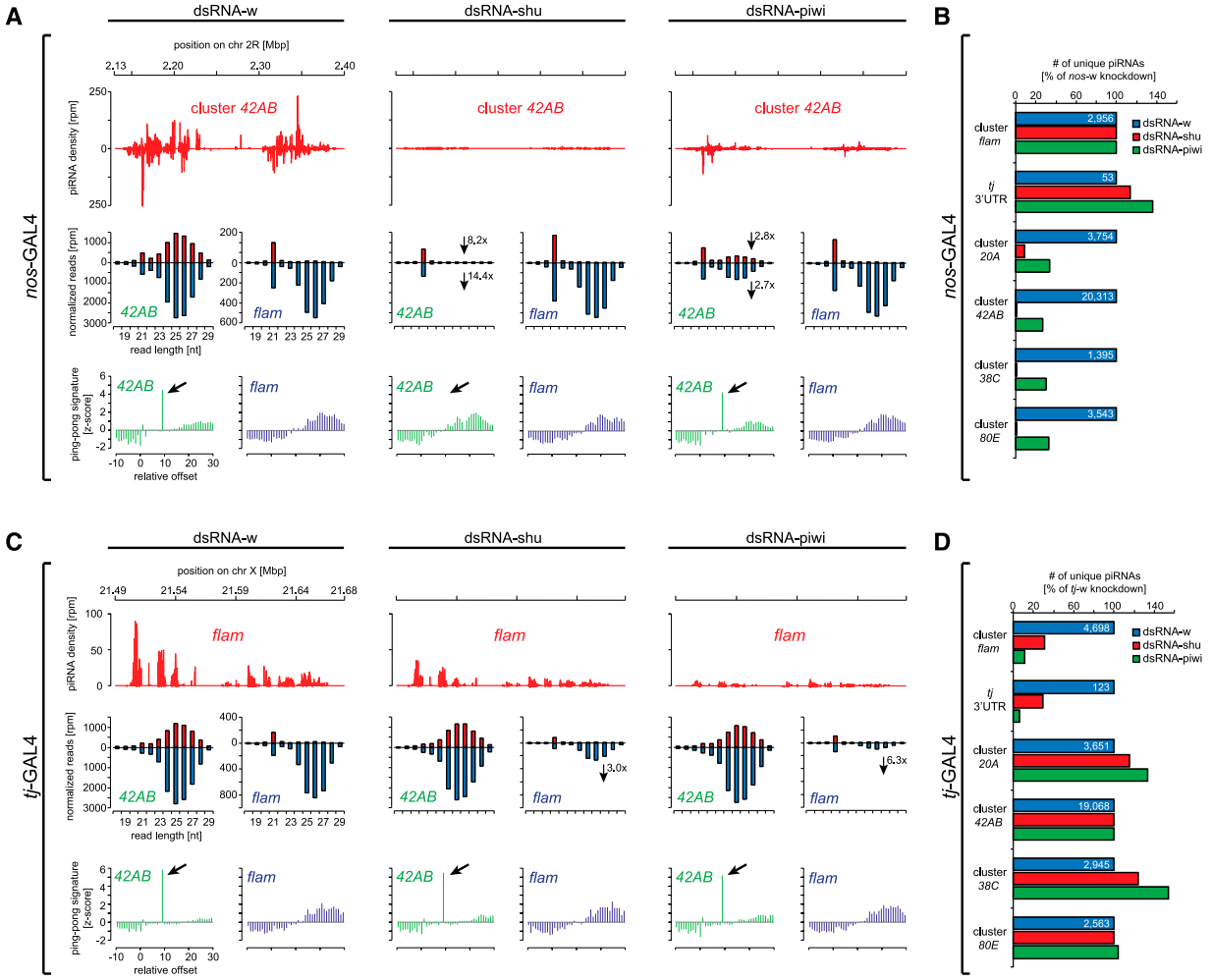


FIGURE 3. Knockdown of *shu* causes loss of cluster-derived piRNAs in both somatic and germline tissues. (*A*) At the *top* is shown a histogram of small RNAs mapping to the germline-specific *42AB* cluster in flies expressing the indicated dsRNAs specifically in germ cells. In the *middle*, the size distribution of RNAs derived from each strand of the *42AB* and *flamenco* clusters is shown as a histogram. At the *bottom* are histograms reflecting the relative enrichment of RNAs overlapping by the indicated number of nucleotides, plotted by *Z*-score, for the *42AB* and *flamenco* clusters in the indicated knockdown animals. The peak at position 9 (arrow) is indicative of a ping-pong interaction. (*B*) A histogram shows relative piRNA levels for a series of germline and somatic clusters. Total reads were normalized across libraries to piRNAs mapping to *flamenco*, which is unaffected in germline-specific knockdowns. For each cluster, changes in mapping piRNAs are shown with reference to the *white* control, which is set to 100%. *C* and *D* are similar to *A* and *B* except that dsRNA expression is driven by a follicle cell–specific *tj*-GAL4 driver. In *C*, at the *top*, reads are shown mapping to the soma-specific *flamenco* cluster. In *D*, reads are normalized across libraries to those derived from *42AB*, whose activity is not affected in the soma-specific knockdown.

(Brennecke et al. 2007), with loading of Aub and Ago3 occurring independently of Piwi function. Small RNAs mapping to this cluster in the *shu*-depleted germline also showed a clear reduction in ping-pong signatures, defined as the frequency of reads with a paired opposite strand read overlapping by 10 nt (Fig. 3A). In contrast, *piwi*, which does not participate significantly in ping-pong amplification, had no effect on ping-ping signatures upon knockdown. The effects of *shu* knockdown appear to be specific to the piRNA pathway. Reads corresponding to miRNAs were not reduced in *shu* knockdown animals (Supplemental Fig. S3). Though enforced Dicer-2 expression generally increased the endo-siRNA fraction, we did not note any further effect of *shu* knockdown, even on endo-siRNAs mapping to piRNA clusters (e.g., *42AB*) (Fig. 3A).

We also analyzed effects of *shu* or *piwi* depletion on other piRNA clusters. We compared reads that could be uniquely mapped to each annotated cluster to the *white* knockdown controls. Reads were set to 100% in the *white* library (normalized read number for the *white* knockdown library is show as a blue bar). piRNAs derived from the 3′ UTR of *traffic jam*, a genic locus that produces piRNAs only in

follicle cells, showed no impact of *shu* and *piwi* knockdowns (Fig. 3B), as expected. In contrast, all germline clusters analyzed showed a dramatic reduction of piRNA levels upon expression of *nos*-GAL4–driven *shu* dsRNAs (<10% remaining as compared to *white* RNAi). Depletion of Piwi had similar effects, although the reduction was less profound (~30% of *white* levels, as seen for *42AB*), probably due to intact Ago3 and Aub loading.

Primary and secondary piRNA biogenesis mechanisms in the germline exhibit some degree of interdependence. For example, disruption of ping-pong in *ago3* mutants or upon Aub knockdown feeds back and reduces the number of primary piRNAs loading into Piwi through unknown mechanisms (Li et al. 2009; Wang and Elgin 2011). Follicle cells, which are of somatic origin, express no detectable Aub or Ago3 and do not use ping-pong amplification. Thus, we directly tested the involvement of *shu* in primary piRNA production by sequencing small RNAs from *tj*-GAL4–driven dsRNA in ovaries. PiRNA-sized small RNAs were normalized using the number of unique reads mapping to the germline-specific *42AB* locus, which is unaffected by *tj*-GAL4–mediated knockdowns.

The sole somatic, unidirectional *flamenco* cluster produces abundant piRNAs that load only the Piwi protein. Thus, as expected, depletion of *piwi* caused a significant reduction in piRNAs derived from this locus (5.2-fold) (Fig. 3C,D). Follicular knockdown of *shu* also produced a marked reduction in *flam* piRNAs (2.9-fold) (Fig. 3C,D). As expected, piRNAs uniquely mapped to *flam* showed no ping-pong signature in any of the somatic knockdowns. Reads corresponding to germline clusters remained unchanged in piRNA abundance, with no shift in size profiles or ping-pong signatures, indicating that, as expected, the pathway remains fully functional in germ cells of animals that have lost *shu* expression only in the soma. As in germ cell–specific knockdowns, miRNA abundance was unaffected (Supplemental Fig. S3).

Mapping small RNAs from our germline-specific knockdown animals to a set of known *Drosophila* transposon consensus sequence further supported a general requirement for *shu* in piRNA accumulation. We retained in our analysis only the 75 transposons with the highest abundance of corresponding piRNAs. Previous reports have demonstrated substantial expression biases for many transposons, with some showing preferential expression in the somatic lineage and others being found predominantly in germline lineages (Malone et al. 2009). For the set of germline-enriched transposons, *nos*-GAL4–driven dsRNA-*shu* substantially affected all known elements, reducing overall piRNA levels (Fig. 4A). In general, sense and antisense piRNAs were depleted to roughly similar extents, suggesting that loading of all three PIWI clade proteins is affected by loss of *shu*. In contrast, only a subset of transposons showed depletion of piRNAs in the *nos*-GAL4–driven *piwi* knockdowns. Elements with a known somatic expression bias, including *ZAM*, *tabor*, *gypsy*, and others (indicated by red dots), show little or no reduction in piRNA levels upon germline knockdown of either *shu* or *piwi* (Fig. 4A). Transposable elements with strong germline signatures (green asterisks), like the LINE element *Rt1b* or the LTR transposon *roo* (pao family), not only showed a severe reduction of their corresponding piRNA levels but also demonstrated a dramatic loss of ping-pong signatures (Fig. 4B). In contrast, soma-specific elements retain their piRNA levels and generally lack ping-pong signatures. As an example of such an element, piRNA levels for the LTR element *ZAM* (gypsy family) are shown (Fig. 4B, bottom).

Summary

A combination of biochemical and genetic approaches are beginning to link a substantial number of proteins to functions in the piRNA pathway. Some act exclusively in primary piRNA biogenesis and affect small RNAs in both the germline and somatic compartments of the *Drosophila* ovary (Malone et al. 2009; Haase et al. 2010; Olivieri et al. 2010; Saito et al. 2010; Handler et al. 2011; Zamparini et al. 2011). Others function exclusively in the germline, and these tend to selectively affect the ping-pong cycle that hones piRNA populations in response to the expression of transposon mRNAs or factors implicated in germline cluster transcription (Klattenhoff et al. 2009; Li et al. 2009; Patil and Kai 2010; Pane et al. 2011; Zhang et al. 2011; Anand and Kai 2012). Here, we followed clues initially provided by proteomic analysis of Piwi-family protein complexes in mice to link *shutdown*, a gene previously shown to be required for fertility in *Drosophila* females (Schupbach and Wieschaus 1991; Munn and Steward 2000), to the piRNA pathway.

Analysis of transposon expression patterns and small RNA libraries in *shu* knockdown cells and animals suggests a role either in piRNA biogenesis or in piRNA stabilization, perhaps by fostering loading of piRNAs into PIWI-family proteins. Shutdown is a member of the FKBP family and its constituent domains have been ascribed PPIase activity and the ability to interact with the HSP90 family of chaperone proteins. Either of these activities could underlie the role of Shutdown in the piRNA pathway. In particular, studies of the Argonaute clade have implicated HSP family chaperones as critical cofactors for small RNA loading (Iki et al. 2010; Iwasaki et al. 2010; Miyoshi et al. 2010; Iki et al. 2011). However, evolutionary comparisons indicate that both the PPI and HSP90-binding domains harbor variations that reduce activity when introduced into other well-studied FKBP family members. Thus, understanding the true role of Shutdown in both primary biogenesis and the ping-pong cycle will await further genetic analysis and the development of biochemical systems that recapitulate aspects of the piRNA pathway in vitro.

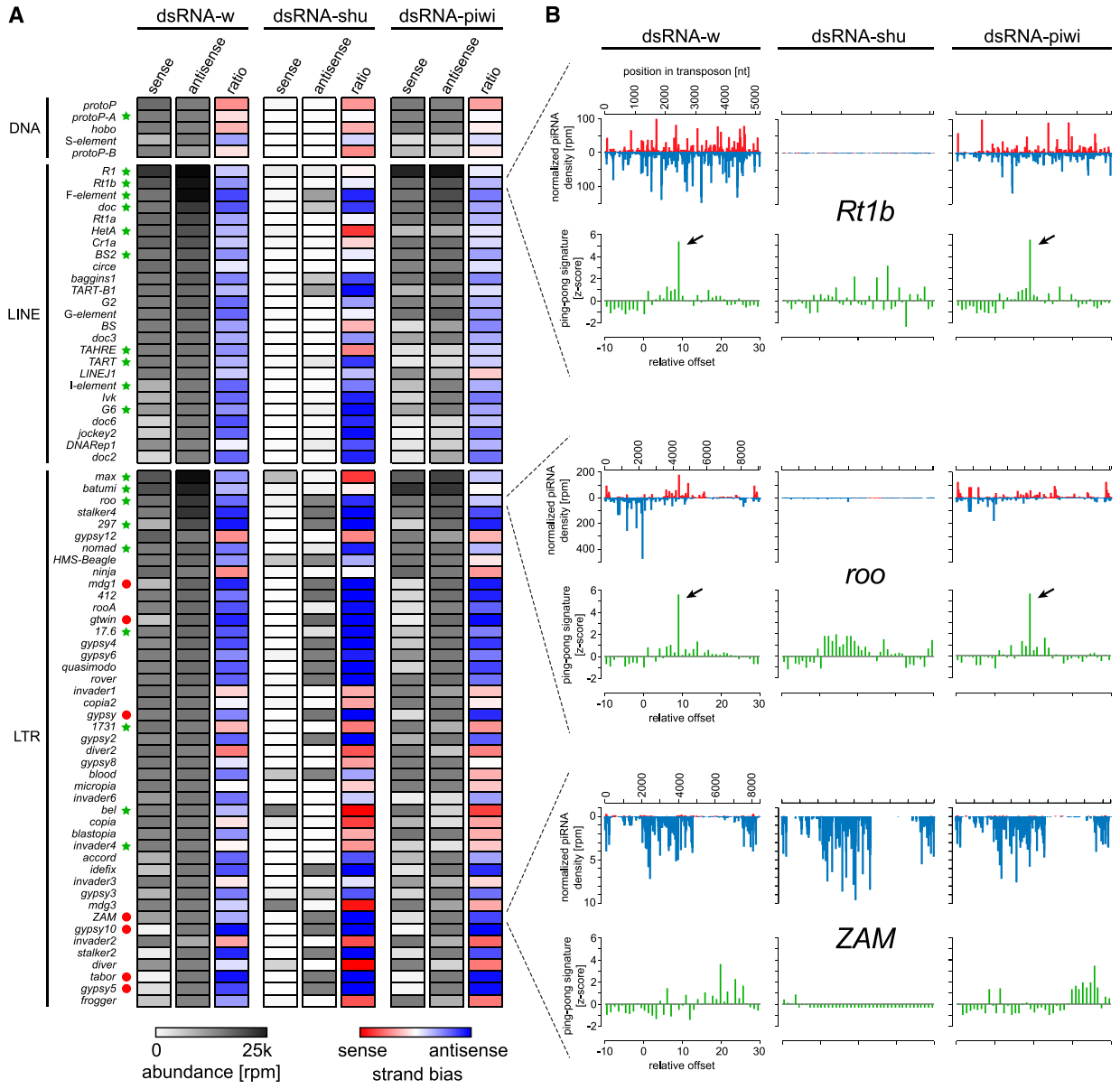


FIGURE 4. Loss of transposon control in *shu* knockdowns is a consequence of piRNA loss. (*A*) The heat map displays changes in piRNA abundance for each germline knockdown (as indicated) for the 75 elements most heavily targeted in our strain. Sense and antisense, with respect to the transposon coding strand, are quantified separately (gray heat maps), and their ratio is also indicated (red-blue heat map). (*B*) For three transposons, piRNAs are plotted along the length of the consensus sequence (*upper*) and a histogram of overlap between sense and antisense species (*lower*) is presented to indicate the degree of ping-pong (arrow highlights peak at position 9). Data are presented for *shu* and *piwi* knockdown and for a control (*white*). Two transposons with strong expression in the germline, *Rt1b* and *roo* (*top* and *middle*), are shown in comparison to a somatically biased element, *ZAM*. Since knockdown is germline specific, *ZAM* piRNAs are unaffected.

MATERIALS AND METHODS

Fly stocks and handling

Drosophila UAS-dsRNA strains were obtained from the VDRC. *nos*-GAL4 and *tj*-GAL4 driver lines were obtained from Bloomington and Kyoto, respectively (see Supplemental Table S1). For knockdown experiments, five males from dsRNA stocks were crossed with five virgin females expressing the desired GAL4 driver. Fertility of the FKBP-family F1 knockdown females was estimated by counting the number of eggs laid and crawling larvae 7 d post transfer to fresh media. Quantitative fertility measurements (shown in Fig. 1D) were obtained by transferring 3-d-old male and female F1 offspring (10 each) to grape-agar plates for

4 h and counting the eggs laid. Hatching frequencies were ascertained after 24 h (measurements were carried out in triplicate). For qPCR, small RNA libraries, and immunofluorescence experiments, ovaries were dissected from 2- to 3-d-old females fed with fresh yeast paste.

Expression of tagged transgenes in OSS

Full-length coding sequences of Shu, Piwi, Armi-RB, and Zuc were amplified from *Drosophila* ovary cDNA, cloned into pENTR/D-TOPO, and recombined into N- or C-terminal GFP destination vectors of the *Drosophila* Gateway collection (Terence Murphy, Carnegie Institute of Washington, Baltimore, MD). *Shu* was cloned into pAGW and pAWG, Zuc into pUWG, and Piwi and Armi into pUGW. Cells were transfected using Xfect reagent (Clontech) and costained with DAPI and MitoTracker Red CMXRos (Invitrogen).

RNA isolation, reverse transcription, qPCR

Ovaries were dissected into cold 1× PBS. Total RNA was extracted using TRIzol reagent (Invitrogen) following the manufacturer's instructions. One microgram RNA was treated with DNase I Amplification Grade (Invitrogen) according to the manufacturer's instructions. Complementary DNA was prepared by reverse transcription using oligo$(dT)_{20}$ primer and SuperScript III Reverse Transcriptase (Invitrogen). qPCR was carried out using SYBR Green PCR Master Mix (Applied Biosystems) and primers listed in Supplemental Table S2 on a Chromo4 Real-Time PCR Detector (BioRad). Transcripts were quantitated by the $\Delta\Delta C_t$ method (Livak and Schmittgen 2001), and normalized to transcript levels of *rp49*. Fold changes are expressed relative to control dsRNA-white knockdown RNA. Significance was calculated using a one-tailed heteroscedastic Student *t*-test of *rp49*-subtracted transposon c(t) values. All experiments were carried out in triplicates, with the average results shown.

Immunofluorescence

Ovaries were fixed in freshly prepared 4% paraformaldehyde for 20 min at room temperature. Blocking and permeabilization were carried out simultaneously in wash buffer (50 mM Tris at pH 6.8, 150 mM NaCl, 0.5% NP-40) supplemented with bovine serum albumin (5 mg/mL). All primary antibodies were diluted 1:1000 and incubated overnight at 4°C in wash buffer plus 1 mg/mL BSA. Anti-Ago3 and Anti-Piwi were generated in our laboratory (Brennecke et al. 2007); monoclonal mouse anti-Aub was provided by Mikiko Siomi (Nishida et al. 2007); and rabbit anti-Vasa (d-260) was purchased from Santa Cruz. Secondary AlexaFluor-488 and -568 antibodies were purchased from Invitrogen and used at 1:1000. Images were acquired on a Perkin Elmer UltraVIEW spinning disk confocal microscope.

RNAseq data analysis

For transcriptome libraries, 1 μg of total RNA from OSS cells transfected with GFP control dsRNA was used as input for the Illumina mRNA-Seq sample prep kit (catalog no. RS-930-1001). Libraries were made following the instructions by the manufacturer and sequenced on the Illumina GAII platform. RNAseq data were deposited to the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo/) under accession no. GSE38090. Publically available ovarian RNAseq data (GEO accession no. GSM424751) (Gan et al. 2010) were reanalyzed for this study. Raw sequence reads were iteratively mapped to the *Drosophila* genome (version dm3) using Bowtie (Langmead et al. 2009) with a tolerance of up to two mismatches. Remaining reads were also mapped to RefGene-annotated exon junctions with TopHat (Trapnell et al. 2009). Transcripts were quantitated using Cufflinks (Trapnell et al. 2010) and expressed as fragments per kilobase per million reads (fkpm) for relative comparison of FKBP family mRNA expression in the ovary.

Small RNA libraries and bioinformatic analysis

Small RNAs were cloned as described (Brennecke et al. 2007). For this study, the following small RNA libraries from total RNAs were prepared:

19–28 nt from *tj*-GAL4–driven dsRNA against white,
19–28 nt from *tj*-GAL4–driven dsRNA against shu,
19–28 nt from *tj*-GAL4–driven dsRNA against piwi,
19–28 nt from *nos*-GAL4–driven dsRNA against white,
19–28 nt from *nos*-GAL4–driven dsRNA against shu, and
19–28 nt from *nos*-GAL4–driven dsRNA against piwi.

Libraries were sequenced in-house using the Illumina GAII sequencing platform. Small RNA sequences were deposited in the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo/) under accession no. GSE38089. The analysis of small RNA libraries was performed similarly as described (Czech et al. 2008). In brief, Illumina reads were stripped of the 3′ linker and collapsed, and the resulting small RNA sequences were matched to the *Drosophila* release 5 genome (version dm3) without mismatches. Only reads that met these conditions were subjected to further analyses. For annotations we used a combination of UCSC (repeats/transposons; noncoding RNAs), miRBase (microRNAs), and FlyBase (protein coding genes; noncoding RNAs) tracks, as well as custom tracks (for synthetic markers, endo-siRNAs from structured loci, and miR and miR* strands) with different priorities. For comparison of small RNA counts between samples, libraries of dsRNA-white samples were set to 1 million reads. Next, all libraries were normalized based on unique piRNA-size mappers to the *flamenco* (for *nos*-GAL4 knockdowns) or *42AB* (for *tj*-GAL4 knockdowns) piRNA clusters. Heat maps were created by plotting the abundance of sequences (all piRNAs to a given transposable element or individual miRNA strands) as well as their strand bias within the indicated libraries.

Ping-pong analysis

For each piRNA, the relative frequency (*Z*-score) of an existing "neighbor" piRNA on the opposite strand within a certain window (10-nt upstream of and 30-nt downstream from each 5′ end of a piRNA) was calculated. In the case of germline and somatic piRNA clusters, this analysis was based on genomic mapping coordinates. For transposons, the 5′ coordinate of each mapping event to the respective transposon consensus sequence was used. Calculated frequencies were based on total cloning count. A spike at position 9 indicates more than average partners with a 10-nt overlap and is a signature of ping-pong amplified piRNAs.

DATA DEPOSITION

RNAseq data and small RNA sequences were deposited in the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo/) under accession no. GSE38098.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank members of the Hannon laboratory for helpful discussion, Yang Yu for providing DNA constructs, and Assaf Gordon for help with ping-pong analysis. We thank the VDRC, Kyoto, and Bloomington stock centers for fly stocks, and Julius Brennecke (IMBA, Austria) and Mikiko Siomi (Keio University, Japan) for antibodies. J.P. is supported by the American Cancer Society (award no. 121614-PF-11-277-01-RMC). B.C. was supported by the Boehringer Ingelheim Fonds. P.M.G. is supported by the NIH (grant 5T32GM065094) and by a William Randolph Hearst Foundation Scholarship. Work in the Hannon laboratory is supported by grants from the NIH and by a kind gift from Kathryn W. Davis. G.J.H. is an investigator of the HHMI.

Received May 14, 2012; accepted May 15, 2012.

REFERENCES

Ahearn IM, Tsai FD, Court H, Zhou M, Jennings BC, Ahmed M, Fehrenbacher N, Linder ME, Philips MR. 2011. FKBP12 binds to acylated H-ras and promotes depalmitoylation. *Mol Cell* **41:** 173–185.

Allan RK, Ratajczak T. 2011. Versatile TPR domains accommodate different modes of target protein recognition and function. *Cell Stress Chaperones* **16:** 353–367.

Anand A, Kai T. 2012. The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in *Drosophila*. *EMBO J* **31:** 870–882.

Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr Biol* **11:** 1017–1027.

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442:** 203–207.

Bozzetti MP, Massari S, Finelli P, Meggio F, Pinna LA, Boldyreff B, Issinger OG, Palumbo G, Ciriaco C, Bonaccorsi S, et al. 1995. The *Ste* locus, a component of the parasitic *cry-Ste* system of *Drosophila melanogaster*, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the β subunit of casein kinase 2. *Proc Natl Acad Sci* **92:** 6067–6071.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128:** 1089–1103.

Chalvet F, Teysset L, Terzian C, Prud'homme N, Santamaria P, Bucheton A, Pelisson A. 1999. Proviral amplification of the Gypsy endogenous retrovirus of *Drosophila melanogaster* involves *env*-independent invasion of the female germline. *EMBO J* **18:** 2659–2669.

Chambeyron S, Popkova A, Payen-Groschene G, Brun C, Laouini D, Pelisson A, Bucheton A. 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc Natl Acad Sci* **105:** 14964–14969.

Chen Y, Pane A, Schupbach T. 2007. *Cutoff* and *aubergine* mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*. *Curr Biol* **17:** 637–642.

Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet* **39:** 715–720.

Cook HA, Koppetsch BS, Wu J, Theurkauf WE. 2004. The *Drosophila* SDE3 homolog *armitage* is required for *oskar* mRNA silencing and embryonic axis specification. *Cell* **116:** 817–829.

Cox DN, Chao A, Lin H. 2000. *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127:** 503–514.

Crackower MA, Kolas NK, Noguchi J, Sarao R, Kikuchi K, Kaneko H, Kobayashi E, Kawai Y, Kozieradzki I, Landers R, et al. 2003. Essential role of Fkbp6 in male fertility and homologous chromosome pairing in meiosis. *Science* **300:** 1291–1295.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453:** 798–802.

DeCenzo MT, Park ST, Jarrett BP, Aldape RA, Futer O, Murcko MA, Livingston DJ. 1996. FK506-binding protein mutational analysis: defining the active-site residue contributions to catalysis and the stability of ligand complexes. *Protein Eng* **9:** 173–180.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9:** 397–405.

Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, Chen X. 2010. Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. *Cell Res* **20:** 763–783.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442:** 199–202.

Gollan PJ, Bhave M. 2010. Genome-wide analysis of genes encoding FK506-binding proteins in rice. *Plant Mol Biol* **72:** 1–16.

Gonzalez-Reyes A, Elliott H, St Johnston D. 1997. Oocyte determination and the origin of polarity in *Drosophila*: the role of the spindle genes. *Development* **124:** 4927–4937.

Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455:** 1193–1197.

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315:** 1587–1590.

Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, Chen C, Gordon A, Hannon GJ. 2010. Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*. *Genes Dev* **24:** 2499–2504.

Handler D, Olivieri D, Novatchkova M, Gruber FS, Meixner K, Mechtler K, Stark A, Sachidanandam R, Brennecke J. 2011. A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J* **30:** 3977–3993.

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129:** 69–82.

Iki T, Yoshikawa M, Nishikiori M, Jaudal MC, Matsumoto-Yokoyama E, Mitsuhara I, Meshi T, Ishikawa M. 2010. In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol Cell* **39:** 282–291.

Iki T, Yoshikawa M, Meshi T, Ishikawa M. 2011. Cyclophilin 40 facilitates HSP90-mediated RISC assembly in plants. *EMBO J* **31:** 267–278.

Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y. 2010. Hsc70/Hsp90 chaperone machinery mediates

ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* **39:** 292–299.

Kamphausen T, Fanghanel J, Neumann D, Schulz B, Rahfeld JU. 2002. Characterization of *Arabidopsis thaliana At*FKBP42 that is membrane-bound and interacts with Hsp90. *Plant J* **32:** 263–276.

Kang CB, Hong Y, Dhe-Paganon S, Yoon HS. 2008. FKBP family proteins: immunophilins with versatile biological functions. *Neurosignals* **16:** 318–325.

Khurana JS, Theurkauf W. 2010. piRNAs, transposon silencing, and *Drosophila* germline development. *J Cell Biol* **191:** 905–913.

Kim A, Terzian C, Santamaria P, Pelisson A, Purd'homme N, Bucheton A. 1994. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc Natl Acad Sci* **91:** 1285–1289.

Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE. 2007. *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell* **12:** 45–55.

Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A, et al. 2009. The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138:** 1137–1149.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25. doi: 10.1186/gb-2009-10-3-r25.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. 2006. Characterization of the piRNA complex from rat testes. *Science* **313:** 363–367.

Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137:** 509–521.

Li L, Lou Z, Wang L. 2011. The role of FKBP5 in cancer aetiology and chemoresistance. *Br J Cancer* **104:** 19–23.

Lim AK, Kai T. 2007. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc Natl Acad Sci* **104:** 6714–6719.

Lin H, Spradling AC. 1997. A novel group of *pumilio* mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* **124:** 2463–2476.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25:** 402–408.

Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136:** 656–668.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137:** 522–535.

McClintock B. 1953. Induction of instability at selected loci in maize. *Genetics* **38:** 579–599.

Miyoshi T, Takeuchi A, Siomi H, Siomi MC. 2010. A direct role for Hsp90 in pre-RISC formation in *Drosophila*. *Nat Struct Mol Biol* **17:** 1024–1026.

Munn K, Steward R. 2000. The *shut-down* gene of *Drosophila melanogaster* encodes a novel FK506-binding protein essential for the formation of germline cysts during oogenesis. *Genetics* **156:** 245–256.

Nishida KM, Saito K, Mori T, Kawamura Y, Nagami-Okada T, Inagaki S, Siomi H, Siomi MC. 2007. Gene silencing mechanisms mediated by Aubergine–piRNA complexes in *Drosophila* male gonad. *RNA* **13:** 1911–1922.

Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J* **29:** 3301–3317.

Orsi GA, Joyce EF, Couble P, McKim KS, Loppin B. 2010. *Drosophila I-R* hybrid dysgenesis is associated with catastrophic meiosis and abnormal zygote formation. *J Cell Sci* **123:** 3515–3524.

Pane A, Wehr K, Schupbach T. 2007. *zucchini* and *squash* encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev Cell* **12:** 851–862.

Pane A, Jiang P, Zhao DY, Singh M, Schupbach T. 2011. The Cutoff protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *EMBO J* **30:** 4601–4615.

Patil VS, Kai T. 2010. Repression of retroelements in *Drosophila* germline via piRNA pathway by the Tudor domain protein Tejas. *Current biology: CB* **20:** 724–730.

Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG. 1994. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO J* **13:** 4401–4411.

Pratt WB. 1998. The hsp90-based chaperone system: involvement in signal transduction from a variety of hormone and growth factor receptors. *Proc Soc Exp Biol Med* **217:** 420–434.

Pratt WB, Galigniana MD, Harrell JM, DeFranco DB. 2004. Role of hsp90 and the hsp90-binding immunophilins in signalling protein movement. *Cell Signal* **16:** 857–872.

Ratajczak T, Carrello A. 1996. Cyclophilin 40 (CyP-40), mapping of its hsp90 binding domain and evidence that FKBP52 competes with CyP-40 for hsp90 binding. *J Biol Chem* **271:** 2961–2965.

Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* **20:** 2214–2222.

Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* **24:** 2493–2498.

Schupbach T, Wieschaus E. 1989. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. I. Maternal effect mutations. *Genetics* **121:** 101–117.

Schupbach T, Wieschaus E. 1991. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129:** 1119–1136.

Senti KA, Brennecke J. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet* **26:** 499–509.

Smith MR, Willmann MR, Wu G, Berardini TZ, Moller B, Weijers D, Poethig RS. 2009. Cyclophilin 40 is required for microRNA activity in *Arabidopsis*. *Proc Natl Acad Sci* **106:** 5424–5429.

Szakmary A, Reedy M, Qi H, Lin H. 2009. The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in *Drosophila melanogaster*. *J Cell Biol* **185:** 613–627.

Theurkauf WE, Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA. 2006. rasiRNAs, DNA damage, and embryonic axis specification. *Cold Spring Harb Symp Quant Biol* **71:** 171–180.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511–515.

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313:** 320–324.

Vagin VV, Wohlschlegel J, Qu J, Jonsson Z, Huang X, Chuma S, Girard A, Sachidanandam R, Hannon GJ, Aravin AA. 2009. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev* **23:** 1749–1762.

Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229:** 105–124.

Wang SH, Elgin SC. 2011. *Drosophila* Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc Natl Acad Sci* **108:** 21164–21169.

Ward BK, Allan RK, Mok D, Temple SE, Taylor P, Dornan J, Mark PJ, Shaw DJ, Kumar P, Walkinshaw MD, et al. 2002. A structure-based mutational analysis of cyclophilin 40 identifies key residues in the core tetratricopeptide repeat domain that mediate binding to Hsp90. *J Biol Chem* **277:** 40799–40809.

Wilson JE, Connell JE, Macdonald PM. 1996. *aubergine* enhances *oskar* translation in the *Drosophila* ovary. *Development* **122:** 1631–1639.

Zamparini AL, Davis MY, Malone CD, Vieira E, Zavadil J, Sachidanandam R, Hannon GJ, Lehmann R. 2011. Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in *Drosophila*. *Development* **138:** 4039–4050.

Zhang Z, Xu J, Koppetsch BS, Wang J, Tipping C, Ma S, Weng Z, Theurkauf WE, Zamore PD. 2011. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol Cell* **44:** 572–584.

Supplementary Figure Legends

Figure S1. Density of siRNAs over genes targeted by dsRNAs. Shown are 21-nt siRNAs mapping to the indicated genes (top: piwi; middle: shu; bottom: white) in small RNA libraries from total RNA of knockdown ovaries. Reads mapping to the genomic plus strand are shown in red, those derived from the minus strand in blue-grey. Blue arrows indicate orientation of transcription. Exon (thick bars) and intron (thin lines) structures are displayed. Note that siRNAs are restricted to the regions targeted by the VDRC dsRNA construct used for the indicated knockdowns.

Figure S2. Sub-cellular localization of piRNA pathway components in OSS cells. Transient transfection of OSS cells with GFP-tagged constructs expressing indicated piRNA pathway factors under the ubiquitous *Actin5c or ubiquitin* promoters are shown. MitoTracker CMXRos and DAPI were used as co-stains for mitochondria and nuclei, respectively. GFP-Armi localizes to cytoplasmic, perinuclear structures consistent with the expected Yb body localization. GFP-Piwi is predominantly nuclear, while Zuc-GFP shows significant overlap with mitochondria. Both N- and C-terminal tagged Shu variants localize to the cytoplasm without significant overlap with mitochondria. Minor enrichments of Shu can be seen in perinuclear foci, but the majority is diffusely cytoplasmic.

Figure S3. MicroRNA levels in knockdown libraries are very similar. Shown are size profiles (upper part) and heat maps (lower part) displaying the top 15 abundant miRNAs in the indicated knockdown libraries (driven either by *nos*-GAL4 or *tj*-GAL4). Size profiles show normalized miRNA reads over their length (18- to 26-nt). Heat maps show total number of reads of the 15 most abundant miRNAs in *white* control knockdowns compared to *shu* and *piwi* RNAi. Abundance in reads per million (rpm) is shown in grey scale.

Supplementary Tables

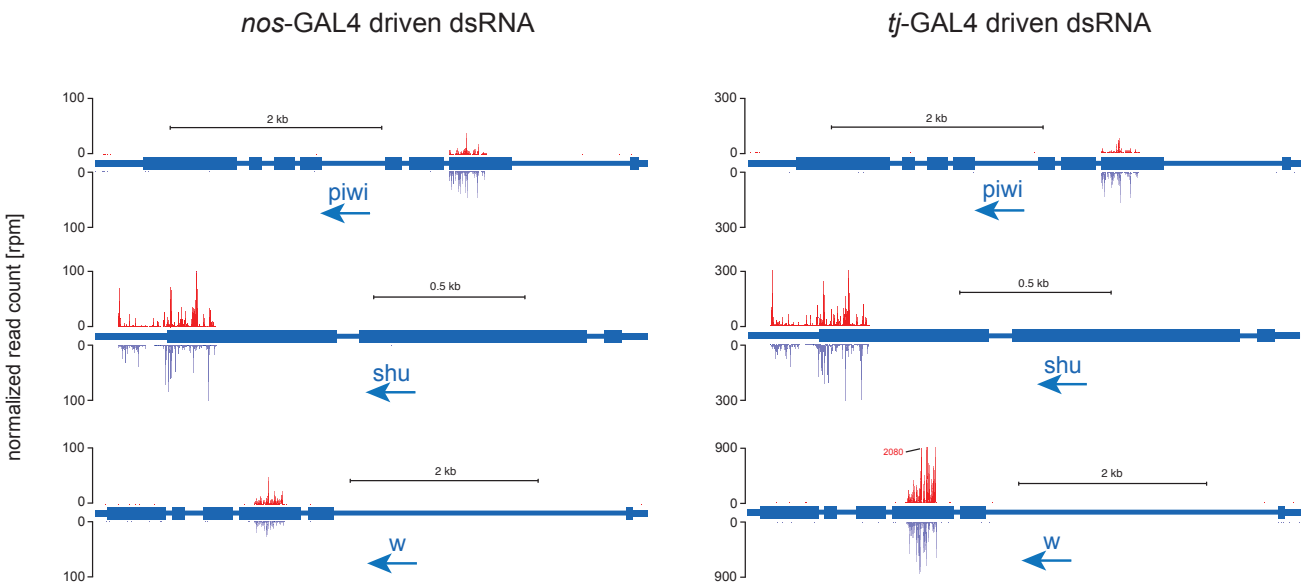
Supplementary Table S1: Fly lines used in this study

Stock Name	Genotype	Source	Stock Number
nos-GAL4	P{w[+mC]=UAS-Dcr-2.D}1, w[1118]; P{w[+mC]=GAL4-nos.NGT}40	Bloomington	25751
tj-GAL4	y* w*; P{GawB}NP1624 / CyO, P{UAS-lacZ.UW14}UW14	Kyoto	104055
dsRNA(*white*)	w^{1118}; ; P{GD14981}v30033	VDRC	v30033
dsRNA(*piwi*)	w^{1118}; P{KK105350}VIE-260B		v101658
dsRNA(*armi*)	w^{1118}; P{KK101517}VIE-260B		v101517
dsRNA(*shu*)	w^{1118}; P{KK102092}VIE-260B		v102092

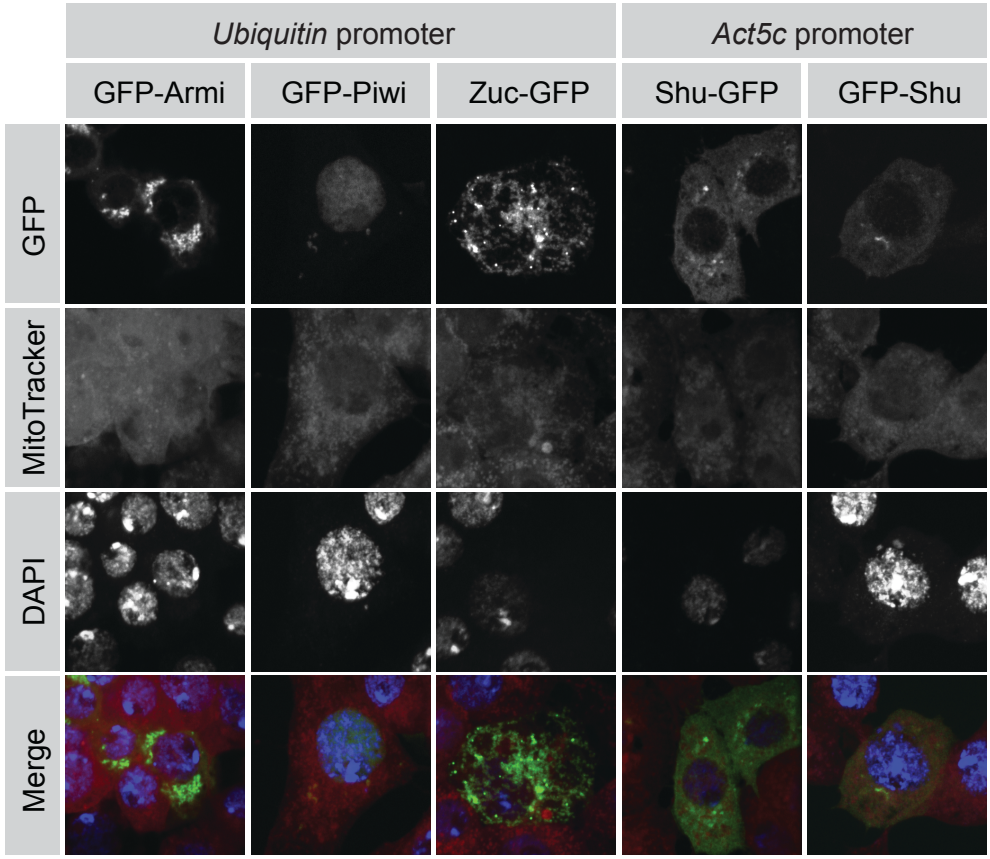
Supplementary Table S2: Primer sequences used for SYBR Green qPCR

Name	forward	reverse
rp49	ATCTCGCCGCAGTAAACGC	CCGCTTCAAGGGACAGTATCTG
Burdock	AGGGAAATATTTGGCCATCC	TTTTGGCCCTGTAAACCTTG
TAHRE	CTGTTGCACAAAGCCAAGAA	GTTGGTAATGTTCGCGTCCT
HetA	CGCGCGGAACCCATCTTCAGA	CGCCGCAGTCGTTTGGTGAGT
Transpac	GGAACGCACCTTCAACATTT	GCAAACTCGCATTTGTCTGA
TART	ACCAGGGAAAAGTGTGAACG	GGTGCAGTGGTATGGCTTTT
McClintock	CCCTAATCCGTTTTCCCAAT	CTGGTCGGTTCTGGTCAAAT
blood	TGCCACAGTACCTGATTTCG	GATTCGCCTTTTACGTTTGC
micropia	CGAATGTTACGCGGTGTATG	CTGGTCAGGTCCAAGGTTGT
Max	ATCTAGCCAGTCGAGGCGTA	TGGAAGAGTGTCGCTTTGTG
Jockey	TCTGCGGTCTCCAGCTTAAT	GTTGGGCAAATGCTAGTGGT
I-element	TGAAATACGGCATACTGCCCCCA	GCTGATAGGGAGTCGGAGCAGATA
Copia	AGCAAACAACCCCTCATGTC	GCAAACCCAATTTGTCTCGT
R2	ATGCTCCCGAAACAACAAAC	GCACTGCAGACTTGGTTCAA
stalker	TTATCAGGCTAGCCACATCTCTG	TTGGCAGATATCACTTCTACCGATTC
ZAM	ACTTGACCTGGATACACTCACAAC	GAGTATTACGGCGACTAGGGATAC
Springer	TGAAGAGCAAGAACCGGAGT	TCCTCCAGCAAAGCTTGTTT
roo	TCCTTTAAGCATCTTACAGCTAAAGG	TTTAGCTGTAAGATGCTTAAAGGAGCT

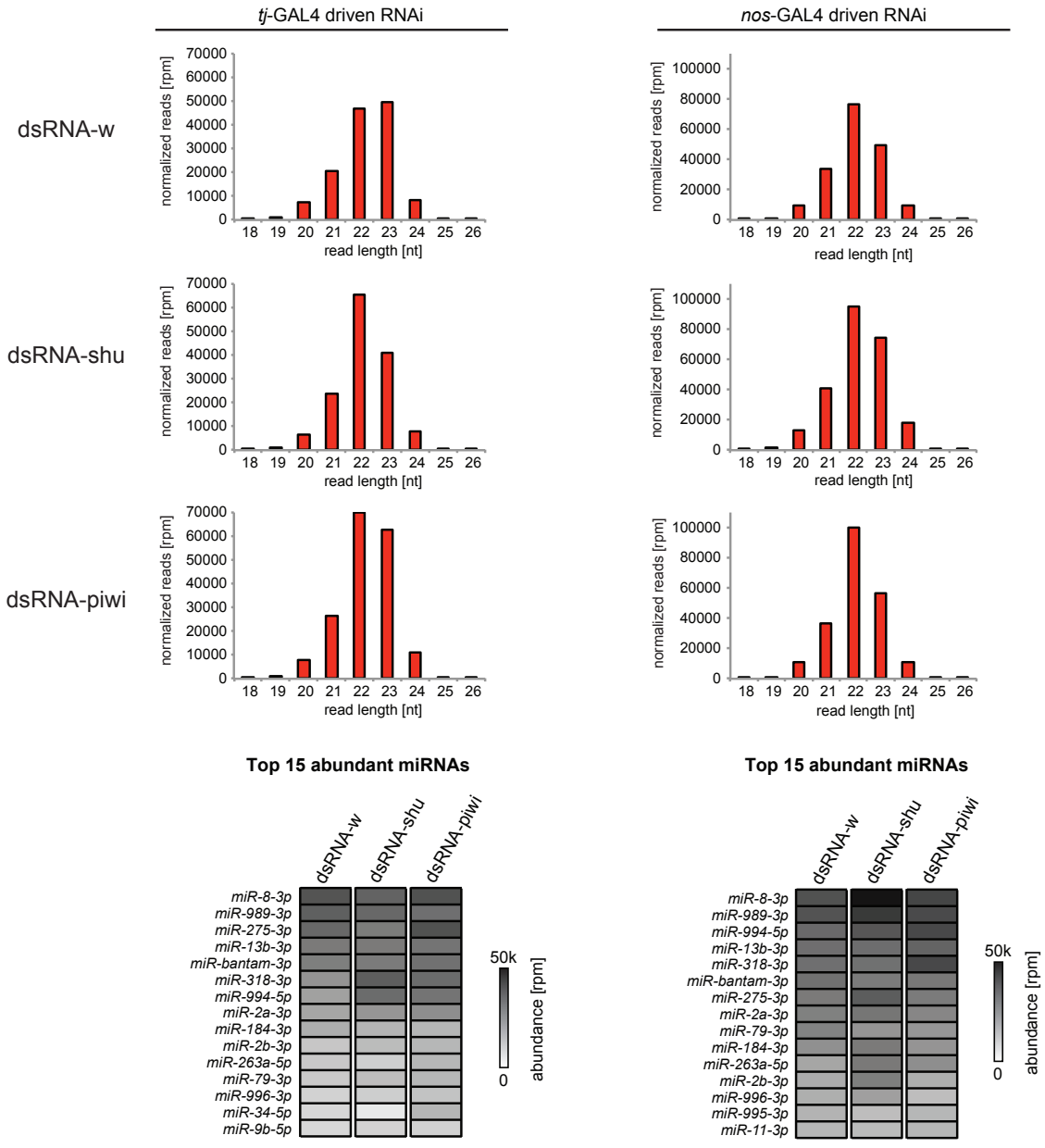
Preall, Czech et al., Fig. S1



Preall, Czech et al., Fig. S2



Preall, Czech et al., Fig. S3



A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals novel factors of the germline piRNA pathway

Benjamin Czech*, Jonathan B. Preall*, Jon T. McGinn, Simon R. Knott and Gregory J. Hannon

* authors contributed equally to this work

In preparation for submission to Cell.

A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals novel factors of the germline piRNA pathway

Benjamin Czech[1,2,3], Jonathan B. Preall[1,2,3], Jon T. McGinn[1,2], Simon R. Knott[1,2] and Gregory J. Hannon[1,2]

[1] Watson School of Biological Sciences
[2] Howard Hughes Medical Institute
Cold Spring Harbor Laboratory
Cold Spring Harbor, New York, USA

[3] contributed equally

Contact: hannon@cshl.edu

Keywords: *Drosophila*, RNAi, piRNA pathway, transposon silencing, germ cells

Summary

The *Drosophila* piRNA pathway provides an RNA-based immune system that defends the germline against selfish genetic elements. Two inter-related branches of the piRNA system exist: one in the somatic cells that support oogenesis and one in the germline. PiRNA-mediated silencing of transposons centers around the action of three gonad-specific Argonaute proteins: *P*-element induced wimpy testis (Piwi), Aubergine (Aub), and Argonaute3 (AGO3). Piwi is common to both somatic and germline pathways, whereas Aub and AGO3 act exclusively in germline cells. While several key factors of each branch have been identified, a great deal remains unknown about their respective inputs and outputs, and the extent to which they overlap. Here, we report a reverse genetic screen that spans the ovarian transcriptome in an attempt to uncover genes required for piRNA-mediated transposon silencing in the female germline. Our results reveal new key factors of piRNA silencing as well as transposon-specific requirements.

Introduction

Eukaryotic organisms of all phyla are constantly challenged by a myriad of genomic parasites such as transposable elements (Malone and Hannon, 2009; Siomi et al., 2011). These mobile elements are broadly classified based on their transposition strategy. Retrotransposons employ an RNA intermediate during mobilization, whereas DNA transposons directly excise and insert their genomic information (Slotkin and Martienssen, 2007). While expression patterns for elements of both classes vary widely over a range of spatial and temporal niches, the reproductive system invariably suffers the majority of transposable element activity in the host organism. Alterations to the host genome caused by transposition, such as disruptions to coding or regulatory regions, double-strand breaks, or chromosomal rearrangements, are generally deleterious and reduce reproductive fitness when they occur in the germ cell lineage (Khurana and Theurkauf, 2010; Malone and Hannon, 2009; Senti and Brennecke, 2010; Siomi et al., 2011).

Small RNAs play a central role in transposon control in eukaryotes. Efficient suppression of mobile elements in animal germ cells relies on the conserved piRNA pathway, whose core is composed of Argonaute proteins of the PIWI-clade and their 23- to 28-nt small RNA partners known as piRNAs (Piwi-interacting RNAs) (Khurana and Theurkauf, 2010; Malone and Hannon, 2009; Senti and Brennecke, 2010; Siomi et al., 2011). To a rough approximation, the piRNA pathway can be seen as RNA-based immune system with innate and adaptive components to its defense response (Brennecke et al., 2007). The targets of the primary piRNA response are genetically hardwired in the form of piRNA clusters, which contain truncated transposon remnants that witness prior exposure to mobile elements (Aravin et al., 2006; Brennecke et al., 2007; Girard et al., 2006). Additionally, an adaptive amplification mechanism known as the ping-pong cycle attunes the intensity of the piRNA response by feeding back sequence information from the active elements themselves to generate new silencing triggers (Brennecke et al., 2007; Gunawardane et al., 2007; Li et al., 2009; Malone et al., 2009).

The piRNA pathway is probably best understood in the ovary of *Drosophila melanogaster,* which comprises two gonadal tissue types each deploying its own unique piRNA system (Malone et al., 2009). The germline is composed of the transcriptionally inactive oocyte and the syncytial nurse cells, which express all three *Drosophila* PIWI-clade proteins, *P*-element induced wimpy testis (Piwi), Aubergine (Aub), and Argonaute3 (AGO3). Germ cells generate piRNAs derived from specific piRNA clusters as well as from active transposable elements via the ping-pong cycle (Brennecke et al., 2007; Gunawardane et al., 2007; Li et al., 2009; Malone et al., 2009). In contrast, the somatic cell lineage that surrounds and supports the germline only expresses Piwi, transcribes piRNAs derived from separate, soma-specific clusters, and does not engage in ping-pong amplification (Brennecke et al., 2007; Lau et al.,

2009; Malone et al., 2009; Saito et al., 2009). The varying expression and sub-cellular localization patterns of the PIWI-family proteins reflect differences in the biogenesis of the piRNAs that fuel them as well as the molecular modes of silencing they employ.

Primary piRNA biogenesis appears to follow similar routes in somatic and germline cells and is initiated with transcription of precursors from lineage-specific piRNA clusters. Processing of cluster transcripts is thought to take place after export to the cytoplasm, likely at specialized perinuclear foci marked by known components of the piRNA biogenesis machinery. Current models suggest that cluster transcripts are parsed linearly into multiple intermediates that feature the 5' end of the mature piRNA but are extended at their 3' end. Following loading into Piwi (and in the germline, Aub), which takes place at in perinuclear structures called Yb-bodies in the soma and nuage in the germline (Handler et al., 2011; Olivieri et al., 2010; Qi et al., 2011; Saito et al., 2010; Szakmary et al., 2009), the bound piRNA is matured through 3' trimming by a-yet to identified exonuclease and methylated by Hen1 (Horwich et al., 2007; Kawaoka et al., 2011; Saito et al., 2007). Mature Piwi-RISC enters the nucleus to silence complementary transposons via an unknown mechanism likely involving chromatin modifications.

The ping-pong amplification loop, in contrast, only operates in germline cells and serves to detect and post-transcriptionally silence highly active transposon threats. Input for the ping-pong cycle is received either from primary biogenesis of cluster transcripts or from maternally deposited piRNAs associated with Aub (Brennecke et al., 2008; Malone et al., 2009). PiRNAs in antisense orientation to transposons guide Aub to appropriate mRNA substrates and leads to target cleavage. The resulting cleavage product loads into AGO3, and following exonucleolytic trimming and methylation yields mature AGO3-RISC associated with sense piRNA (Brennecke et al., 2007; Gunawardane et al., 2007). These transposon-derived piRISCs complete the ping-pong cycle by pairing with and cleaving more cluster transcripts to produce further Aub-loaded antisense piRNAs (Brennecke et al., 2007; Gunawardane et al., 2007). Numerous factors have been linked to the ping-pong cycle, with most of them localizing to nuage and a prominent subset being members of the Tudor protein family (Anand and Kai, 2012; Handler et al., 2011; Lim and Kai, 2007; Malone et al., 2009; Patil and Kai, 2010; Zamparini et al., 2011; Zhang et al., 2011).

The vast majority of today's known piRNA pathway components were originally identified by forward genetic approaches in classic screens aimed to uncover mutations that affect oogenesis, female fertility or spindle formation (Schupbach and Wieschaus, 1989, 1991), and only linked to the piRNA pathway later. Although reverse genetics utilizing transgenic RNAi have proven powerful to reveal novel factors in many studied pathways in somatic tissues, technical limitations prevented the application of RNAi in the female germline until recently. However, through overexpression of Dcr2, we and others have recently

demonstrated the effectiveness of transgenic RNAi in germ cells of the female ovary (Handler et al., 2011; Preall et al., 2012; Wang and Elgin, 2011), enabling us to apply this technology to uncover missing piRNA pathway components on a large scale.

Although we are beginning to understand basic concepts of piRNA-mediated silencing, many important aspects of the pathway are still enigmatic due to gaps in our knowledge of central factors such as nucleases and silencing effectors. Thus, we designed an RNAi-based screen to systematically probe for missing components of the piRNA system *in vivo*. To cover the broadest range of new factors, we specifically targeted the germline pathway, which comprises both primary biogenesis and the ping-pong cycle (Brennecke et al., 2007; Malone et al., 2009). Here, we report the results of this screen, which spans the ovarian transcriptome and found 74 genes (including already known piRNA pathway components), whose knockdown caused strong de-repression of four individual transposon types. Secondary assays enabled us to distinguish factors involved in the production of primary and/or secondary piRNAs from those required for different silencing mechanisms carried out by Piwi or Aub/AGO3 complexes. The identification of these genes, followed by a detailed characterization will uncover their specific functions and ultimately shed light on piRNA biogenesis and effectors functions.

Results

Screening assays to monitor transposon de-repression

In order to limit the number of labor-intensive crosses, we first cataloged the observable ovarian transcriptome by RNAseq. Using Cufflinks (Trapnell et al., 2010) to quantify transcripts in replicate samples (R^2 > 0.95) prepared from ovaries dissected from our screening stock that carries a *Dcr2* transgene and the *nos*-GAL4 driver, we found significant expression levels for 8,396 protein-coding genes (average FPKM > 1) (Figure 1A), which corresponds to 60.86% of *Drosophila melanogaster* genes (total of 13,795 unique protein-coding genes; version dm3, refseq release 4.48). As the ovary consists of both somatic and germline tissue, we compared our ovarian RNAseq data with experiments carried out on RNA from Ovarian Somatic Sheet (OSS) cells (Niki et al., 2006), early embryos (indicator for germline RNAs as only those are deposited), and ovariextomized carcasses with the goal to identify germline-enriched genes. In general, we found strong correlations between entire ovaries and the other two ovarian data sets (embryo: R^2 = 0.457; OSS: R^2 = 0.603), with little correlation of expression levels in data sets derived from ovaries and carcasses (R^2 = 0.151) (Figure S1A). The high correlation between the ovarian transcriptome and RNAseq from embryos and OSS is also evident by the pronounced overlap of expressed genes with FPKM > 1 (Figure S1B). Comparing datasets, we see no

genes that are significantly ($P < 0.01$) expressed at FPKM < 1 in ovary versus carcass, suggesting that we are unlikely to miss any ovary-specific genes by selecting this value as cutoff for screening (Figure S1C). We conclude that our transcriptome analysis identified ovarian-expressed genes with sufficient stringency to serve as a filter for selecting screening targets.

Of the 8,396 genes expressed in ovaries, 97.3% (8,171 genes) were covered by RNAi lines by the Vienna *Drosophila* RNAi Center (VDRC) library (Figure 1A, bottom) (Dietzl et al., 2007). To deplete factors specifically in germ cells, we used our previously reported knockdown strategy: Males carrying the UAS-driven dsRNA were crossed to virgin females containing the germline-specific driver *nos*-GAL4 and a UAS-*Dcr2* transgene, which was previously shown to increase the knockdown efficiency (Handler et al., 2011; Preall et al., 2012; Wang and Elgin, 2011). Applying this strategy to deplete Piwi and Armi from germ cells resulted in dramatic and selective de-repression of germline-dominant transposons, similar to earlier results (Figure 1B) (Preall et al., 2012). Due to their consistent and robust fold increases compared to *white* knockdowns, we based our screen on two LINE-like elements *HeTA* and *TAHRE*, as well as the LTR elements *burdock* and *blood*. For these preliminary measurements, we used RNA isolated from dissected ovaries and standard SYBR green-based qPCR assays.

To achieve the throughput necessary for a large-scale screen, we isolated RNA in 96-well format from whole adult females. Combination of this strategy with optimized, sensitive *TaqMan* qPCR assays provided comparable sensitivity and dynamic range to dissected ovaries, making it an ideal screening platform (Figure 1C). Furthermore, we confirmed the specificity of our system by germline knockdown of the soma-specific component Yb, which caused no significant change in transposon levels (Handler et al., 2011; Olivieri et al., 2010; Qi et al., 2011; Saito et al., 2010). We further optimized our workflow by using of Taqman assays that allowed multiplexed qPCRs with the normalization standard *rp49* and the probed transposons in the same reaction tube.

To identify piRNA pathway components required for transposon silencing, we setup trays containing 96 crosses, each containing *white* and *armi* control knockdowns. Flies were allowed to mate for 7 days and then discarded. Offspring flies were transferred to fresh food vials 5 days later and allowed to mature for 2.5 additional days. Of each cross, we transferred six females into 96-well collection tubes, and performed RNA isolation, reverse transcription and multiplexed qPCR. Data was analyzed and *z*-scores were calculated for each plate (Figure 1D).

Comprehensive identification of factors required for proper transposon silencing

We screened a collection of 8,171 dsRNA lines *in vivo* for genes involved in piRNA silencing for four individual transposons. To increase the stringency for potential candidates, we also calculated the average of the four individual transposon *z*-scores and computed heat maps for average and individual transposon data (Figure 2A). Requiring an average *z*-score of -1.5 or lower, our primary screen only identified 216 out of all 8,171 probed dsRNA lines (2.64%) to result in robust transposon de-repression. Strikingly, using this threshold for candidate selection, all of the positive *armi* controls were included, whereas none of the negative *white* controls was (Figure 2C), with similar results obtained when transposons were assessed individually (Figure S2). Based on the magnitude of transposon de-repression, we further categorized these candidates into "weak" and "strong", with knockdown of 74 genes that resulted in average *z*-scores of -2 or lower in the "strong" class (Figure 2B). Using these more stringent criteria, we found all known piRNA pathway components (except egg/dSETDB1), thus providing internal validation for our screen.

Validation of candidate genes identified from primary screening

We chose the 216 strongest candidates based on transposon de-repression, and repeated our experimental workflow by re-crossing the same dsRNA line followed by multiplexed qPCR. Whenever available, we extended our analysis by including an independent RNAi line (dsRNA or shRNA, derived from either the VDRC or TRiP stock centers) targeting a different region of the candidate gene. In addition to the four transposons, we also probed the levels of two genes highly expressed in germline cells (*nos*, *yTub37c*). For better comparison across data sets, we calculated all expression changes relative to the average of ten independent negative control knockdowns (6x *white*, 3x *yb*, 1x *GFP*), and normalized to the average of two *armi* control knockdowns included on each plate. Furthermore, we analyzed potential fertility defects by counting the number of larvae and pupae. For an improved overview, heat maps summarizing all assayed parameters are shown only for the 75 strongest hits (Figure 3). Overall, we found significant overlap between our primary screen data, the re-screening using the identical dsRNA line, and an independent RNAi line, evidencing the high reproducibility of our data. Notably, compromised fertility correlates either with reduced expression of the germline markers *nos* and *yTub37c*, or with general transposon de-repression with changes similar to *armi* knockdowns.

Our re-screening validated the identification of all known piRNA pathway components that have been reported to affect the germline pathway (Shu, AGO3, Vas, Rhi, Squ, Piwi, Tej, Mael, Cuff, Vret, Zuc, Armi, Spn-E, BoYb, Mino/CG5508, Aub, and Qin), with the exception of Krimper, for which no dsRNA line was available, and the *Drosophila* SETDB1 ortholog, Egg. In addition to genes already implicated in the piRNA pathway, our screen uncovered numerous novel candidates that cause dramatic transposon de-repression, often

accompanied by fertility defects, when depleted from germline cells. The novel candidate with the highest average level transposon up-regulation, *deadlock* (*del*), was previously reported as gene important for germline maintenance and female fertility, with lesions resulting in defects during early oogenesis (Schupbach and Wieschaus, 1991; Wehr et al., 2006). Another gene that dramatically de-repressed all four measured transposable elements was *CG2183*, which we named *GASZ* after its nearest vertebrate counterpart (GASZ, Germ cell-specific protein with Ankyrin repeats, Sterile alpha motif, and a putative basic leucine Zipper domain) (Yan et al., 2004; Yan et al., 2002). *Drosophila* GASZ shares the ankyrin repeats and a predicted sterile alpha motif of vertebrate GASZ (Figure S3). Using our stringency (E-value threshold = 0.1) and domain prediction algorithm (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), we did not detect the basic leucine zipper motif in *Drosophila* or mouse GASZ. Murine GASZ is a nuage component, with its loss resulting in transposon de-repression, sterility and reduced piRNA levels (Ma et al., 2009). Another candidate, Ars2, was previously shown to impact the biogenesis of miRNAs and siRNAs (Gruber et al., 2009; Sabin et al., 2009), but no function in the piRNA pathway is yet reported. In all, we identified a total of 74 factors with pronounced effects on all transposons tested.

Placement of candidate hits within the piRNA pathway

To characterize GASZ and Del at the molecular level, we first studied the subcellular localization pattern of selected piRNA pathway components upon germline-specific knockdown (Figure 4). Depletion of *white* resulted in expected localization of all investigated proteins: Piwi was strongly distributed within the nurse cell nuclei, whereas Aub and AGO3 were both enriched in nuage granules, and Armi was detected in diffuse perinuclear structures reminiscent of nuage (Brennecke et al., 2007; Cox et al., 1998; Cox et al., 2000; Gunawardane et al., 2007; Lim and Kai, 2007; Nishida et al., 2007; Saito et al., 2010; Saito et al., 2006). A dsRNA specific to *armi* resulted in loss of protein below detectable levels in germline cells, while Armi staining in the Yb-bodies of adjacent follicle cells remained intact. Consistent with its function in primary piRNA biogenesis, depletion of Armi caused the redistribution of Piwi from nurse cell nuclei, but had no effect on the localization of Aub and AGO3. Upon knockdown of *gasz*, nuclear localization of Piwi was severely compromised, whereas Aub and AGO3 appeared normal. Interestingly, we also observed a redistribution of Armi from nuage granules into cytoplasmic speckles, suggesting that GASZ might act in primary piRNA biogenesis. Depletion of *del*, in contrast, did not affect Piwi or Armi localization. Instead, we detected a pronounced re-distribution of Aub from nuage, while AGO3 remained normal. Thus, *del* could play specific roles in the ping-pong cycle.

GASZ and Del affect different steps of piRNA biogenesis

Transposon de-repression, reduced fertility, and disturbed protein localization of PIWI-clade members are hallmarks of piRNA pathway mutants. To directly probe possible effects on piRNA levels, we cloned and sequenced small RNA libraries from ovaries with germline-specific depletion of GASZ or Del. For comparison, we also prepared small RNA libraries from germline knockdowns of several known piRNA pathway factors. We normalized small RNA libraries to the number of unique piRNA reads derived from the *flam* cluster, which were not affected by germline knockdowns. Germline-specific depletion of Yb behaved highly similar to *white* knockdowns and caused no alterations in piRNA populations, whereas all other dsRNAs resulted in significantly reduced piRNA levels (Figure 5A). Consequently, while the fraction of reads corresponding to miRNAs was indistinguishable between *white* and *yb* knockdowns, slightly increased miRNA fractions coincided with reduced piRNA levels in all other knockdowns (Figure 5A).

The piRNA levels (23- to 29-nt) of unique reads derived from the *42AB* locus, in contrast, were dramatically reduced upon knockdown of *armi* (35.3x), *gasz* (57.4x), *spn-E* (26.9x), *aub* (6.0x), or *del* (13.3x), as shown by length profiles. The abundance of endo-siRNAs, which are produced and loaded by independent machineries, was similar for all knockdowns, with a marginal increase observed in cases where piRNAs were depleted from the small RNA cloning pool. Notably, knockdown of *del* resulted in a near-complete loss of endo-siRNAs.

Next, we analyzed the effect of depleting either of these factors on the ping-pong signature, which is defined as the frequency of reads from opposite strands overlapping by 10-nt, of *42AB*-derived piRNAs (Figure 5B). Knockdown of *armi* and *gasz* resulted in increased ping-pong signatures, compared to *white* or *yb* knockdown, whereas depletion of Spn-E and Aub, known components of the ping-ping amplification loop, as well as Del resulted in significantly reduced frequency of ping-pong pairs (Figure 5B). The reciprocal effects on ping-pong signatures observed in *gasz* and *del* knockdowns further support the conjecture that these genes play roles specific to primary and secondary (ping-pong) piRNA biogenesis pathways, respectively.

We also analyzed effects on piRNAs corresponding to a set of 80 established transposons (Figure 5C). Based on previous data, we separated these elements into those that dominate in somatic cells, intermediate transposons and mobile elements predominantly active in germline cells (Malone et al., 2009). In comparison to *white* knockdown, depletion of Yb caused only minimal changes in piRNAs derived from any transposon (Figure 5D, top). All other knockdowns showed varying reduction of piRNA levels from germline dominant elements. Intermediate transposons showed mildly reduced piRNA levels, and soma-enriched mobile elements were not changed significantly (less than 2-fold). Levels of germline dominant piRNAs were affected to similar extent

upon knockdown of *armi* and *gasz*, suggesting related functions in primary biogenesis (Figure 5D, middle). Knockdown of *armi* and *aub* resulted in highly correlating piRNA levels (Figure 5D, bottom), in agreement with Aub receiving inputs from primary biogenesis, but despite unaltered Aub localization upon *armi* knockdown (Figure 4).

These similar relations were confirmed by mapping piRNAs over the LTR transposon *batumi*, which is predominantly active in the germline lineage (Figure 5E). While knockdown of *yb* showed similar piRNA levels to *white* controls, depletion of Armi and GASZ resulted in severe reduction of piRNAs, with some ping-pong-derived pairs persisting. Knockdown of *spn-E*, in contrast, caused a dramatic loss of piRNA populations. Depletion of Aub resulted in significantly reduced piRNA levels, with the remaining sequences probably associated with the other PIWI-clade proteins Piwi and AGO3. Knockdown of *del* also showed severely reduced levels of piRNAs matching the *batumi* transposon, but had distinct patterns compared to all other knockdowns.

Differential silencing requirements for distinct transposon types

Transposons are cataloged based on their sequence, transposition strategy, and replication intermediates and can be separated coarsely into DNA elements and retrotransposons (Slotkin and Martienssen, 2007). The latter are further subdivided into LTR transposons (including *blood* and *burdock*), and non-LTR elements to which the LINE-like *HeTA* and *TAHRE* transposons belong. Each element class displays its own unique genome and life cycle within the fly, thus it is likely that transposon-specific adaptions evolved for efficient silencing of each type. To analyze whether specific genes are only required for repression of certain transposons, we compared de-silencing phenotypes between elements. For this analysis, we relied on the *z*-scores obtained from the primary, transcriptome-wide screen, as they provide a more robust measure of the de-silencing phenotype due to intrinsic variable in the fold induction seen from line to line. As expected by their similar replication cycle, *z*-scores for *HeTA* and *TAHRE* highly correlated amongst the top 500 genes scored as hits (R^2 = 0.62) (Figure 6A). In contrast, correlations between *HeTA* and the LTR elements *burdock* (Figure 6B) or *blood* (Figure 6C) are much weaker (R^2 = 0.01 and 0.08, respectively), probably reflecting important differences in the silencing determinants for these elements. Importantly, all factors known to affect piRNA biogenesis showed strong transposon de-repression for all four elements, with an ensemble average *z*-score of -3.66 (±0.57). Thus, knockdowns that robustly de-silence all four elements are highly enriched for core piRNA pathway components. After inspecting our secondary screen (Figure 3) as a filter for false positives, we can classify 17 new genes with an average transposon *z*-score in this range, including both *gasz* and *del*.

One of the strengths of our screen setup is that it also reveals an array of new factors that participate in silencing individual mobile elements or transposon families, but lay outside the core piRNA pathway. For example, *CG5694* and *lsd1,* also known as *Su(var)3-3*, are critical for silencing both *HeTA* and *TAHRE*, but have no effect on *burdock* or *blood* (compare Figure 6A with panels B and C). Interestingly, *lsd1* codes for a histone demethylase, suggesting a specific requirement for certain chromatin modifiers for the suppression of LINE-like elements. Notably, *Actr13E* was specifically required for the suppression of *blood*, with little to no effect on the other tested transposons. A comparison of *blood* and *burdock* only found few factors that compromised silencing of one mobile element, without affecting the other (R^2 = 0.07) (Figure 6D).

Discussion

With very few exceptions, piRNA pathway components were originally identified by classic reverse genetic screens and only linked to piRNA biogenesis or effector mechanisms through later experiments (Schupbach and Wieschaus, 1989, 1991). Forward genetic approaches have been powerful in identifying new pathway components in somatic tissues of *Drosophila*, and were recently engaged in mini-screens to probe the requirement of known factors for somatic piRNA silencing or to analyze Tudor and FKBP protein biology in somatic and germline cells (Handler et al., 2011; Olivieri et al., 2010; Preall et al., 2012). However, until today, no systematic screen has been published that aimed to find novel components of the much more complex germline piRNA pathway. Here, we report an *in vivo* screen carried out in germ cells of the *Drosophila* ovary designed to uncover new core piRNA machinery as well as factors involved in general mechanisms of transposon silencing. Overall, we identified 74 genes that severely affect silencing of a set of four transposons. Importantly, of 17 known piRNA pathway components included in the screen, all but Egg were present in this enriched set, providing a powerful validation of the screen approach (false-negative rate at ~5.9% or one in 17).

By measuring de-silencing of an array of transposon types, we were able to further narrow down our candidate list to factors that are very likely to be core components of the piRNA response. Knockdowns of known piRNA related genes also share the properties of inducing sterility while grossly maintaining the integrity of the germline, as measured by specific marker mRNAs *nos* and *yTub37c.* Our screen identifies ~15 genes that show properties exactly like established piRNA genes. Among these, we confirmed *CG2183*, now named *GASZ*, which shows a similar domain structure to the murine piRNA pathway component GASZ (Ma et al., 2009), as putative primary biogenesis factor. Germline-specific depletion of GASZ displays molecular phenotype highly reminiscent of Armi, with piRNA populations dramatically reduced, while ping-pong signatures remain intact reflecting a disruption of primary, but not

secondary, piRNA biogenesis. Moreover, loss of GASZ perturbs localization of Armi to nuage-like granules, suggesting that GASZ may act upstream of Armi.

In addition, we provide evidence that Del, which was previously found to be essential for female fertility, oogenesis, and germline maintenance (Schupbach and Wieschaus, 1991; Wehr et al., 2006), is a novel piRNA component that participates in secondary biogenesis. In addition to a robust activation of transposable elements, *del* knockdown results in a severe reduction of piRNA populations and reduced ping-pong amplification. Consistent with this observation, depletion of Del displaces Aub, but not AGO3, from nuage while retaining nuclear localization of Piwi. The protein product of *del* is not conserved outside of drosophilids and contains no predicted domains or motifs. Pending further investigation, we speculate that Del may be required for the expression or stabilization of Aub. While we have confidently added two new factors to the milieu of the piRNA biogenesis machinery, further experiments will be necessary to discriminate piRNA producers from effectors amongst the remainder of our candidate core components.

Simultaneously with this effort, our lab also conducted an independent, genome-wide RNAi screen for primary piRNA components in *Drosophila* OSS cells (Muerdter, Guzzardo et al., submitted). This complementary screen reports several of the same components identified here, namely *CG2183* (now named GASZ), *CG3893,* and *windei (wde).* Somatic and germline piRNA pathways face distinct threats with regards to transposon activation, and thus it is to be expected that many factors will be unique to each lineage. Among the overlap, we expect to find primarily genes involved in the biogenesis of cluster-derived primary piRNAs (transcription, precursor export, import of Piwi, or transcriptional activation of clusters by Piwi), though the clusters in question differ significantly. Additionally, each tissue may utilize similar general effector mechanisms coupled to Piwi-bound piRNAs.

Clues to silencing effector mechanisms

Small RNA-based gene silencing pathways have been suggested to regulate targets at the transcriptional level in a variety of organisms, including *Drosophila*. In particular, piRNA-mediated silencing has been shown to promote the deposition of repressive chromatin marks, in particular H3K9me3, at certain transposon loci (Klenov et al., 2007). Curiously, the same mark has also been shown to be essential for the robust transcription of piRNA clusters (Rangan et al., 2011). As a nuclear protein, Piwi is the likely candidate for the principal orchestrator of such chromatin modifications, but its cohort of co-repressor proteins remain largely mysterious. While the mechanism of transcriptional-level silencing by small RNAs has been extensively characterized in fission yeast, most of the key players are not conserved in *Drosophila*. This screen classifies several known regulators of transcriptional output as “strong” transposon silencers, including the H3K9 methyltransferase *Su(var)3-9*, *HP1*/*Su(var)205*,

His2Av, *mof*, and *TfIIA-S*. A significant number of further chromatin-related genes seem to exert an effect on individual elements, most notably the putative H3K4-specific demethylase Lsd1.

Transposable elements that threaten the germline genome can found within all classes of chromatin, and are thus embedded within a variety of molecular contexts. It seems plausible that the specific effectors modulating a given element insertion are a function of chromatin context, with Piwi acting as a general purpose local silencing trigger by targeting a nascent transcript and providing a cooperative binding surface for a limited set of nearby co-repressors. Another factor that may contribute to the variability we observe across candidates and element classes are the idiosyncrasies of a given element's life cycle and transposition strategy. The telomeric transposons *HeTA* and *TAHRE*, for example, play an important cellular role that is presumably maintained in a delicate balance: too little activity could lead to telomere shortening and chromosomal aberrations, while too much could lead to mutagenic levels of transposition. These elements are very likely regulated by a number of cellular processes independent of the piRNA pathway. Similarly, many LTR elements exploit host machinery to assemble into virus-like particles. Recognition of structures harboring transposon RNAs may be a key strategy employed by the piRNA machinery for finding its targets.

In summary, our transcriptome-wide RNAi screen has provided a rich new set of genes involved in transposable element suppression in the *Drosophila* female germline. Our aim was to produce a comprehensive list of core components of the primary and secondary piRNA pathways, and indeed we have confirmed two hits as such with many more yet to pursue in more detail. A second benefit of this dataset is to provide insights into the cellular niche occupied by a subset of highly abundant transposable elements in *Drosophila.* Understanding how an individual transposable element exploits its host machinery is likely to lead to important discoveries regarding general mechanisms of gene silencing and RNA trafficking. Most importantly, this study in combination with the accompanying somatic screen advances our knowledge of the key players in the piRNA pathway, with the ultimate goal of having the blueprints for building a fully operationally piRNA silencing machine nearly in hand.

Experimental Procedures

Fly stocks and husbandry

All fly strains for the primary screen were purchased from the VDRC. RNAi lines for re-screening were obtained from the VDRC or TRiP (Harvard). Strains carrying the X-chromosome linked UAS-*Dcr2* (P{UAS-Dcr-2.D}1) transgene and female germline driver *nos*-GAL4 (P{GAL4-nos.NGT}40) (# 25751), as well as the stock containing a *hs-hid* transgene on the Y-chromosome (P{w[+mC]=hs-hid}Y) (# 8846) were obtained from Bloomington. For facilitated virgin collection, we generated a conditional virginator stock that contains the *Dcr2* transgene and *nos*-GAL4 driver, as well as the *hs-hid* transgene (males: *Dcr2*/*hs-hid*(Y); *nos*-GAL4 and females: *Dcr2*; *nos*-GAL4).

Flies were kept on standard media at 25°C. Virgin females were obtained by two consecutive 2 hr heat shocks at 37°C (spaced by 24 hrs) of bottles containing 0-24 hr old embryos. For germline-knockdown experiments, five males expressing RNAi constructs were crossed to five virgin females (3-5 day old) expressing *nos*-GAL4 and UAS-*Dcr2*. After 7 days of mating, parental flies were discarded. F1 offspring flies were transferred to fresh food media for 2.5 days before collection of six females into 96-well collection tubes (seeded with one 5 mm stainless steel bead per well, Qiagen) and stored at -80°C until further processing. For immunofluorescence experiments and small RNA libraries, F1 offspring flies were additionally supplemented with yeast paste and ovaries were dissected in cold 1x PBS.

RNA isolation, reverse transcription, qPCR

For screening experiments, total RNAs were isolated using the RNeasy 96 Universal Tissue kit (Qiagen) according to the manufacturer's instructions. In brief, 750 μl QIAzol Lysis Reagent were added to each collection tube and tissue was ground using a Mixer Mill MM400 (Retsch) twice for 5 min at 25 Hz. 150 μl chloroform were added to each sample, mixed and centrifuged. The aqueous phase containing the RNA was transferred to square blocks and mixed with 400 μl of 70% EtOH. The mixture was transferred to RNeasy 96-well plates placed on a vacuum manifold (Qiagen). DNase treatment was performed on column following the manufacturer's instructions. Following DNase digest, RNAs were washed once with 800 μl Buffer RW1, and twice with 800 μl Buffer RPE. RNAs were eluted in 100 μl RNase-free water by centrifugation.

Complementary DNA was obtained by reverse transcription using 2.5 μl RNA as input, oligo$(dT)_{20}$ primers, and the SuperScript III Reverse Transcriptase system (Invitrogen). Multiplexed qPCRs were carried out using TaqMan Universal Master Mix II, no UNG (Applied Biosystems) and primers listed in Table S1. Experiments were performed on a Chromo4 Real-Time PCR Detector (BioRad), and data was quantified using the Opticon Monitor software. *Z*-scores

for transposon expression were calculated on ΔC_T values ($C_T^{[\text{transposon}]} - C_T^{[rp49\ \text{control}]}$).

For dissected ovaries, TRIzol reagent (Invitrogen) was used to extract total RNAs. 1 μg RNA was treated with DNase I Amplification Grade (Invitrogen) according to the manufacturer's instructions. Reverse transcription was carried out with oligo(dT)20 primers using the SuperScript III Reverse Transcriptase system (Invitrogen). QPCRs were either carried out using SYBR Green PCR Master Mix or TaqMan Universal Master Mix II, no UNG (Applied Biosystems) with primers listed in Table S1. Transposon levels were quantified using the $\Delta\Delta C_T$ method (Livak and Schmittgen, 2001), normalized to *rp49* and fold changes were calculated relative to knockdown of *white*. All experiments were carried out in triplicates, with the average and standard deviation shown.

Immunofluorescence

The immunofluorescence experiments were carried out as described (Preall et al., 2012). In brief, ovaries were fixed in freshly prepared 4% paraformaldehyde for 20 min at room temperature. Blocking and permeabilization were carried out simultaneously in wash buffer (50 mM Tris pH6.8, 150 mM NaCl, 0.5% NP-40) supplemented with bovine serum albumin (BSA) (5 mg/mL). Primary antibodies were diluted 1:1000 and incubated overnight at 4°C in wash buffer supplemented with 1 mg/mL BSA. Anti-Piwi and anti-AGO3 were generated in the Hannon laboratory (Brennecke et al., 2007). Anti-Aub and anti-Armi were gifts from Mikiko Siomi (Nishida et al., 2007; Saito et al., 2010). Secondary antibodies (AlexaFluor-488 and -568) were purchased from Invitrogen and used at 1:1000 dilutions. Images were acquired on a Perkin Elmer UltraVIEW spinning disk confocal microscope.

Small RNA libraries

Small RNA libraries were constructed similar as described (Brennecke et al., 2007), with slightly modified adapters that enabled multiplexed sequencing. The below small RNA libraries were prepared for this study:

19- to 28-nt from *nos*-GAL4-expressed dsRNA-*white*
19- to 28-nt from *nos*-GAL4-expressed dsRNA-*yb*
19- to 28-nt from *nos*-GAL4-expressed dsRNA-*armi*
19- to 28-nt from *nos*-GAL4-expressed dsRNA-*gasz*
19- to 28-nt from *nos*-GAL4-expressed dsRNA-*spn-E*
19- to 28-nt from *nos*-GAL4-expressed dsRNA-*aub*
19- to 28-nt from *nos*-GAL4-expressed dsRNA-*del*

Libraries were sequenced in-house using the Illumina HiSeq platform.

Bioinformatic analysis of small RNA libraries

The analysis of small RNA libraries was performed similar as described (Czech et al., 2008; Preall et al., 2012). Briefly, Illumina reads were stripped of the 3' linker and collapsed. The resulting small RNA sequences were matched to release 5 of the *Drosophila* genome (version dm3). Only reads that met these conditions were used for further analyses. Sequences were annotated using a combination of miRBase (microRNAs), Flybase (protein coding genes; non-coding RNAs), and UCSC (transposons; non-coding RNAs) tracks, as well as custom annotations for synthetic cloning markers, endo-siRNAs from structured loci and individual miR and miR* strands. Following removal of reads corresponding to structural RNAs and synthetic markers, total small RNA counts (18- to 29-nt) were set to one million reads in the dsRNA-*white* library. For comparison between samples, all libraries were normalized based on unique piRNA-sized (23- to 29-nt) mappers to the *flam* cluster. Size profiles for *flam* and *42AB* piRNA clusters were obtained by extracting the abundance and read length of sequences uniquely matching to these loci. For ping-pong analysis of *42AB*-derived piRNAs, we extracted uniquely mapping reads. For each piRNA, we recorded the abundance of all neighboring piRNAs on the opposite genomic strand within a certain window (20-nt upstream and 10-nt downstream) as well as their relative 5' end distance, with each sequence only counted once per offset even if multiple ping-pong pairs were detected. We calculated *z*-scores to display ping-pong signatures within the probed window. A peak at position 9 is indicative of more piRNA partners at the opposite strand that overlap by precisely 10-nt and serves as signature for ping-pong amplified piRNAs. To compare piRNA levels matching to individual transposons across libraries, we extracted all mapping reads based on annotations and calculated the ratio relative to *white* control knockdowns ($\log_2$ transformation was used for improved display). Density plots were generated by matching all piRNA reads against the *batumi* consensus sequence allowing zero mismatches.

RNAseq experiments and transcriptome analysis

For transcriptome libraries, 6 μg of total RNA were extracted from ovaries of 2-3 day old females, carcasses of the same ovariectomized females, 0-1 hr dechoronized embryos, or OSS cells using TRIzol (Invitrogen). All flies were of the genotype (UAS-*Dcr2*; *nos*-GAL4) used as the driver line for our screen. Two independent biological replicates were analyzed for each tissue type. The samples were depleted of rRNA using Ribo-Zero Gold (Epicentre) and used as input for the ScriptSeq v2 kit (Epicentre). Libraries were multiplexed using TruSeq barcoding PCR primers and sequenced in-house using the Illumina HiSeq platform.

Raw sequencing reads were mapped iteratively to release 5 of the *Drosophila* genome (version dm3, refseq release 4.48 from 9/13/2011) using Bowtie (Langmead et al., 2009) tolerating up to two mismatches. The remaining reads were mapped to the RefGene-annotated exon junctions using TopHat (Trapnell et al., 2009). Transcripts were quantified using Cufflinks (Trapnell et al.,

2010) and converted in fragments per kilobase of exon per million reads (FPKM), which was used to select candidates for initial screening. For differential expression analysis, Bowtie-mapped reads aligning to *Drosophila* RefSeq genes were counted using HTSeq and processed using the DESeq R package (Anders and Huber, 2010). For the estimation of ovary-expressed genes, we only considered protein-coding mRNAs, and removed all non-coding transcripts (i.e., snoRNAs, snRNAs, tRNAs).

Acknowledgements

We thank members of the Hannon laboratory for helpful discussion, Maria Mosquera and Steven Sau for help with fly husbandry and qPCR, and Assaf Gordon and Felix Muerdter for computational support. We thank the VDRC, TRiP and Bloomington stock centers for fly stocks. For part of this work, BC was supported by a PhD fellowship from the Boehringer Ingelheim Fonds. JP is supported by the American Cancer Society (award number 121614-PF-11-277-01-RMC). This work was supported in part by grants from the NIH and a kind gift from Kathryn W. Davis. GJH. is an investigator of the HHMI.

Accession Numbers

RNAseq and small RNA data were deposited in the Gene Expression Omnibus database under accession number GSExxxxx (pending).

References

Anand, A., and Kai, T. (2012). The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in Drosophila. Embo J *31*, 870-882.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T.*, et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science *322*, 1387-1392.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev *12*, 3715-3727.

Cox, D.N., Chao, A., and Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Development *127*, 503-514.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R.*, et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S.*, et al.* (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature *448*, 151-156.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Gruber, J.J., Zatechka, D.S., Sabin, L.R., Yong, J., Lum, J.J., Kong, M., Zong, W.X., Zhang, Z., Lau, C.K., Rawlings, J.*, et al.* (2009). Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. Cell *138*, 328-339.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011). A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. Embo J *30*, 3977-3993.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol *17*, 1265-1272.

Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. (2011). 3' end formation of PIWI-interacting RNAs in vitro. Mol Cell *43*, 1015-1022.

Khurana, J.S., and Theurkauf, W. (2010). piRNAs, transposon silencing, and Drosophila germline development. J Cell Biol *191*, 905-913.

Klenov, M.S., Lavrov, S.A., Stolyarenko, A.D., Ryazansky, S.S., Aravin, A.A., Tuschl, T., and Gvozdev, V.A. (2007). Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. Nucleic Acids Res *35*, 5430-5438.

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

Lau, N.C., Robine, N., Martin, R., Chung, W.J., Niki, Y., Berezikov, E., and Lai, E.C. (2009). Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. Genome Res *19*, 1776-1785.
Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M*., et al.* (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell *137*, 509-521.
Lim, A.K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A *104*, 6714-6719.
Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.
Ma, L., Buchold, G.M., Greenbaum, M.P., Roy, A., Burns, K.H., Zhu, H., Han, D.Y., Harris, R.A., Coarfa, C., Gunaratne, P.H*., et al.* (2009). GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. PLoS Genet *5*, e1000635.
Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.
Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656-668.
Niki, Y., Yamaguchi, T., and Mahowald, A.P. (2006). Establishment of stable cell lines of Drosophila germ-line stem cells. Proc Natl Acad Sci U S A *103*, 16325-16330.
Nishida, K.M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., Siomi, H., and Siomi, M.C. (2007). Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. Rna *13*, 1911-1922.
Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. Embo J *29*, 3301-3317.
Patil, V.S., and Kai, T. (2010). Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas. Curr Biol *20*, 724-730.
Preall, J.B., Czech, B., Guzzardo, P.M., Muerdter, F., and Hannon, G.J. (2012). shutdown is a component of the Drosophila piRNA biogenesis machinery. Rna *18*, 1446-1457.
Qi, H., Watanabe, T., Ku, H.Y., Liu, N., Zhong, M., and Lin, H. (2011). The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. J Biol Chem *286*, 3789-3797.
Rangan, P., Malone, C.D., Navarro, C., Newbold, S.P., Hayes, P.S., Sachidanandam, R., Hannon, G.J., and Lehmann, R. (2011). piRNA production requires heterochromatin formation in Drosophila. Curr Biol *21*, 1373-1379.
Sabin, L.R., Zhou, R., Gruber, J.J., Lukinova, N., Bambina, S., Berman, A., Lau, C.K., Thompson, C.B., and Cherry, S. (2009). Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. Cell *138*, 340-351.
Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., and Siomi, M.C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature *461*, 1296-1299.
Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K.M., Siomi, H., and Siomi, M.C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev *24*, 2493-2498.
Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev *20*, 2214-2222.
Saito, K., Sakaguchi, Y., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev *21*, 1603-1608.

Schupbach, T., and Wieschaus, E. (1989). Female sterile mutations on the second chromosome of Drosophila melanogaster. I. Maternal effect mutations. Genetics *121*, 101-117.
Schupbach, T., and Wieschaus, E. (1991). Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. Genetics *129*, 1119-1136.
Senti, K.A., and Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. Trends Genet *26*, 499-509.
Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol *12*, 246-258.
Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet *8*, 272-285.
Szakmary, A., Reedy, M., Qi, H., and Lin, H. (2009). The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. J Cell Biol *185*, 613-627.
Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.
Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.
Wang, S.H., and Elgin, S.C. (2011). Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc Natl Acad Sci U S A *108*, 21164-21169.
Wehr, K., Swan, A., and Schupbach, T. (2006). Deadlock, a novel protein of Drosophila, is required for germline maintenance, fusome morphogenesis and axial patterning in oogenesis and associates with centrosomes in the early embryo. Dev Biol *294*, 406-417.
Yan, W., Ma, L., Zilinski, C.A., and Matzuk, M.M. (2004). Identification and characterization of evolutionarily conserved pufferfish, zebrafish, and frog orthologs of GASZ. Biol Reprod *70*, 1619-1625.
Yan, W., Rajkovic, A., Viveiros, M.M., Burns, K.H., Eppig, J.J., and Matzuk, M.M. (2002). Identification of Gasz, an evolutionarily conserved gene expressed exclusively in germ cells and encoding a protein with four ankyrin repeats, a sterile-alpha motif, and a basic leucine zipper. Mol Endocrinol *16*, 1168-1184.
Zamparini, A.L., Davis, M.Y., Malone, C.D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G.J., and Lehmann, R. (2011). Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. Development *138*, 4039-4050.
Zhang, Z., Xu, J., Koppetsch, B.S., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W.E., and Zamore, P.D. (2011). Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol Cell *44*, 572-584.

Figure Legends

Figure 1. Screen workflow and summary of preliminary experiments. (a) Relative expression levels of protein-coding genes in *Drosophila melanogaster* are shown for ovarian RNAseq data as histogram. Green bars highlight ovary-expressed genes with FPKM > 1 (top). Doughnut diagram showing screened genes where dsRNA line was available from the VDRC (bottom). **(b)** Histograms show the relative expression levels of indicated transposons detected in ovaries from *Drosophila* that express dsRNA against *piwi* or *armi* in germline cells. Fold changes are relative to knockdown of *white*. Measurements were carried out on ovary-dissected total RNA. Error bars indicate standard deviation (n = 3). **(c)** Relative expression levels of the indicated mobile elements upon germline-specific knockdown of *piwi* and *armi* are shown. Depletion of Yb served as control. Fold changes relative to dsRNA against *white* (indicated by red line) were calculated. Measurements were carried out on RNA extracted from whole female flies. Error bars indicate standard deviation (n = 3). **(d)** Scheme of the screen setup. A germline-specific driver, *nos*-GAL4, was used to express UAS-dsRNA constructs in germ cells of the developing oocyte. UAS-*Dcr2* was co-expressed specifically in germ cells to enhance the RNAi response. Two and a half day old female offspring flies were collected and following RNA isolation and reverse transcription probed for de-repression of four transposons by multiplexed qPCR. Crosses were carried out in trays of 96 that contained a positive (*armi*) and negative (*white*) control knockdown.

Figure 2. Summary of primary screen and determination of candidate hits. (a) Heat map displaying transposon de-repression (as *z*-scores) for all 8,171 investigated ovary-expressed genes in red-blue scale. The average of the four tested transposons is shown along the separate z-scores of *HeTA*, *TAHRE*, *blood*, and *burdock*. Negative z-scores indicate transposon de-repression (shown in red). **(b)** Close-up of the heat map for 216 candidate hits with *z*-scores < -1.5. **(c)** Box plots summarizing *z*-scores of all screened genes, positive *armi* controls, and negative *white* controls. Average data from four transposons was used for the analysis.

Figure 3. Identification and validation of strong candidate genes. Heat maps summarizing transposon de-repression, germline marker gene expression, and sterility phenotypes upon germline-specific knockdown of indicated genes. Data is presented relative to depletion of Armi. Yellow boxes highlight known piRNA pathway components.

Figure 4. Subcellular localization phenotypes upon depletion of the piRNA pathway candidate factors GASZ and Del. Knockdown of *armi* and *gasz* in the germline using *nos*-GAL4 causes Piwi delocalization from nuclei and

redistribution of Armi from nuage-like sites. The localization of Aub and AGO3 is not changed. *Nos*-GAL4-driven dsRNA against *del* results in redistribution of Aub, whereas the localization of Piwi, AGO3 and Armi is not affected. Knockdown of *white* is shown as control.

Figure 5. Knockdown of *gasz* and *del* affects different steps of piRNA biogenesis. (a) Size distribution of 18- to 29-nt small RNAs derived from each strand of the *flam* and *42AB* clusters are shown as histogram (red sense; blue antisense). The fraction of miRNAs (green) for the indicated libraries is highlighted in the cake diagrams. **(b)** Histograms showing the relative enrichment of piRNAs overlapping by the indicated number of nucleotides are plotted for *42AB*-derived sequences in the indicated knockdowns (using *nos*-GAL4). The peak at position 9 (arrow, the number corresponds to the *z*-score) is suggestive of a ping-pong signature. **(c)** Histograms showing the abundance of piRNAs mapping to soma dominant (green), intermediate (grey), or germline dominant (orange) transposons in *white* knockdowns compared to depletion of the indicated genes ($\log_2$ scale). **(d)** Scatter plots ($\log_{10}$ scale) comparing the levels of piRNAs mapping to germline dominant (orange), intermediate (grey), or soma dominant (green) transposons in the indicated knockdown libraries. **(e)** Histograms of piRNAs mapping to the consensus sequence of the germline dominant *batumi* LTR transposon, are shown for the indicated knockdowns.

Figure 6. Specific requirements for silencing of different transposon types. (a) Scatter plot comparing de-repression (as *z*-scores) for *HeTA* and *TAHRE* transposons (top 500 candidates from the primary screen are shown in red, median 500 candidates are indicated in blue). Known piRNA pathway components are highlighted in green. **(b)** Similar as in (a), but *HeTA* de-repression is compared to the LTR element *burdock*. **(c)** Similar as in (a), but comparing *HeTA* to *blood*. **(d)** Similar as in (a) except levels of *blood* are compared to levels of *burdock*.

Supplemental Information

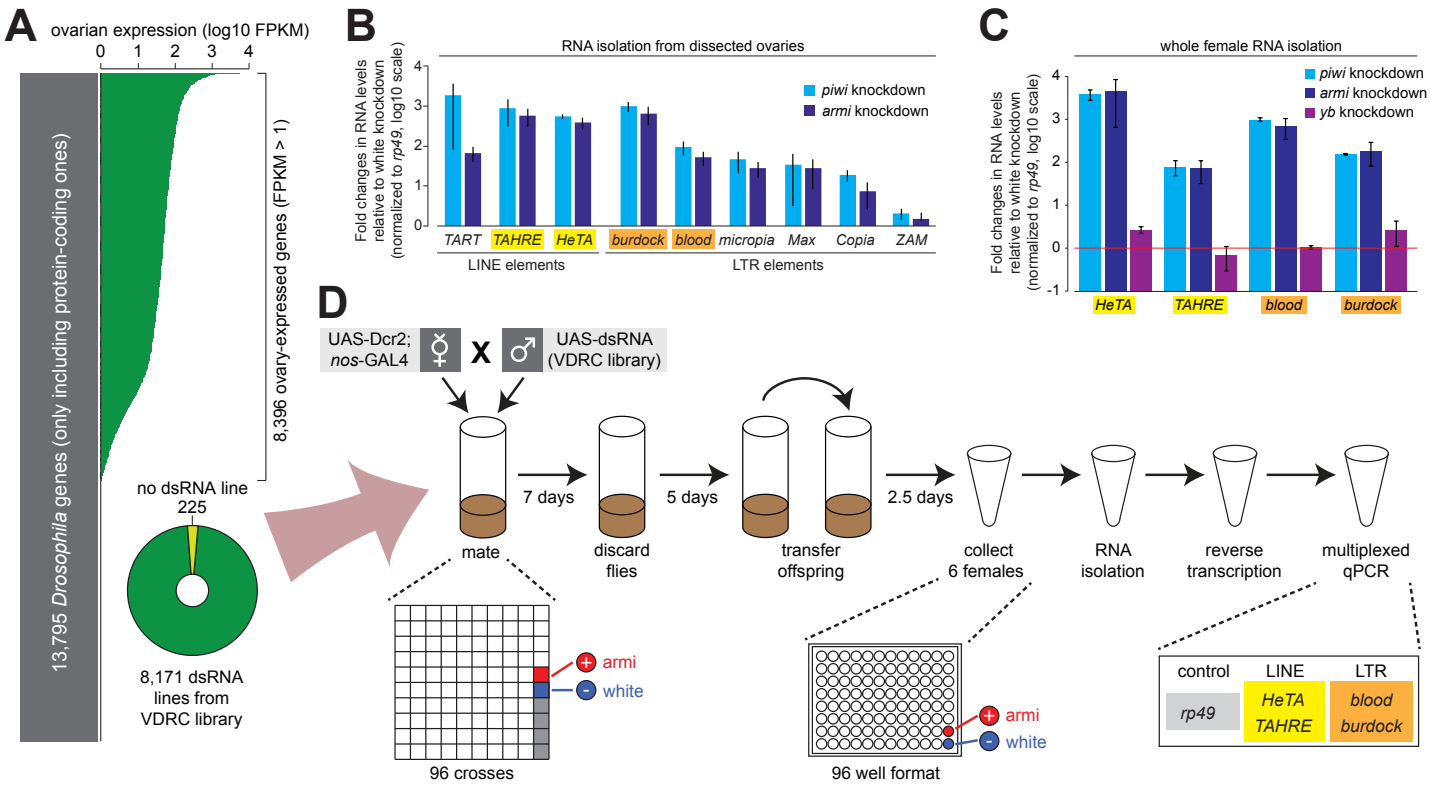
Supplemental Figure Legends

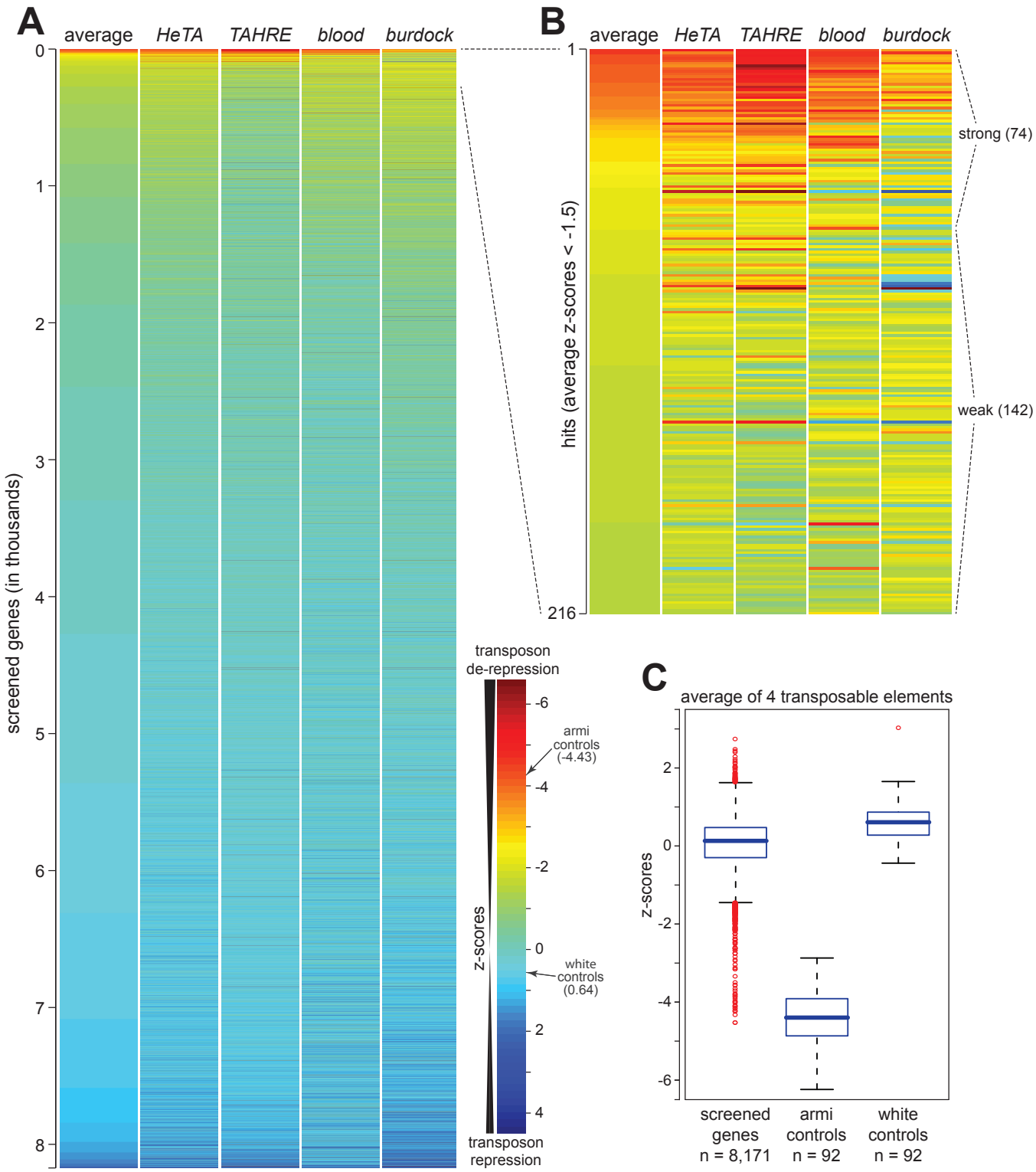
Figure S1. Enrichment analysis of RNAseq from ovaries and other tissues. **(a)** Scatter plots comparing the relative expression levels of protein-coding genes in *Drosophila* ovaries to carcasses, embryos or OSS cells (log_{10} FPKM). **(b)** VENN diagram showing genes with shared and specific expression. Only genes with FPKM > 1 in either tissue were included to the analysis. **(c)** Scatter plot displaying relative expression ratio of all genes in ovary versus carcass RNAseq data as a function absolute expression level in ovary (log_{10} FPKM). Statistically significant ($P < 0.01$) or sub-significant ($P > 0.01$) levels of differential expression are denoted with black dots and grey dots, respectively. Genes with ovary FPKM > 1, which is the range screened in our study, are highlighted by yellow background.

Figure S2. Candidate hit calling. Box plots summarizing *z*-scores of all screened genes, positive *armi* controls, and negative *white* controls, respectively. Data for each of the tested transposons (*HeTA*, *TAHRE*, *blood*, *burdock*) is shown separately.

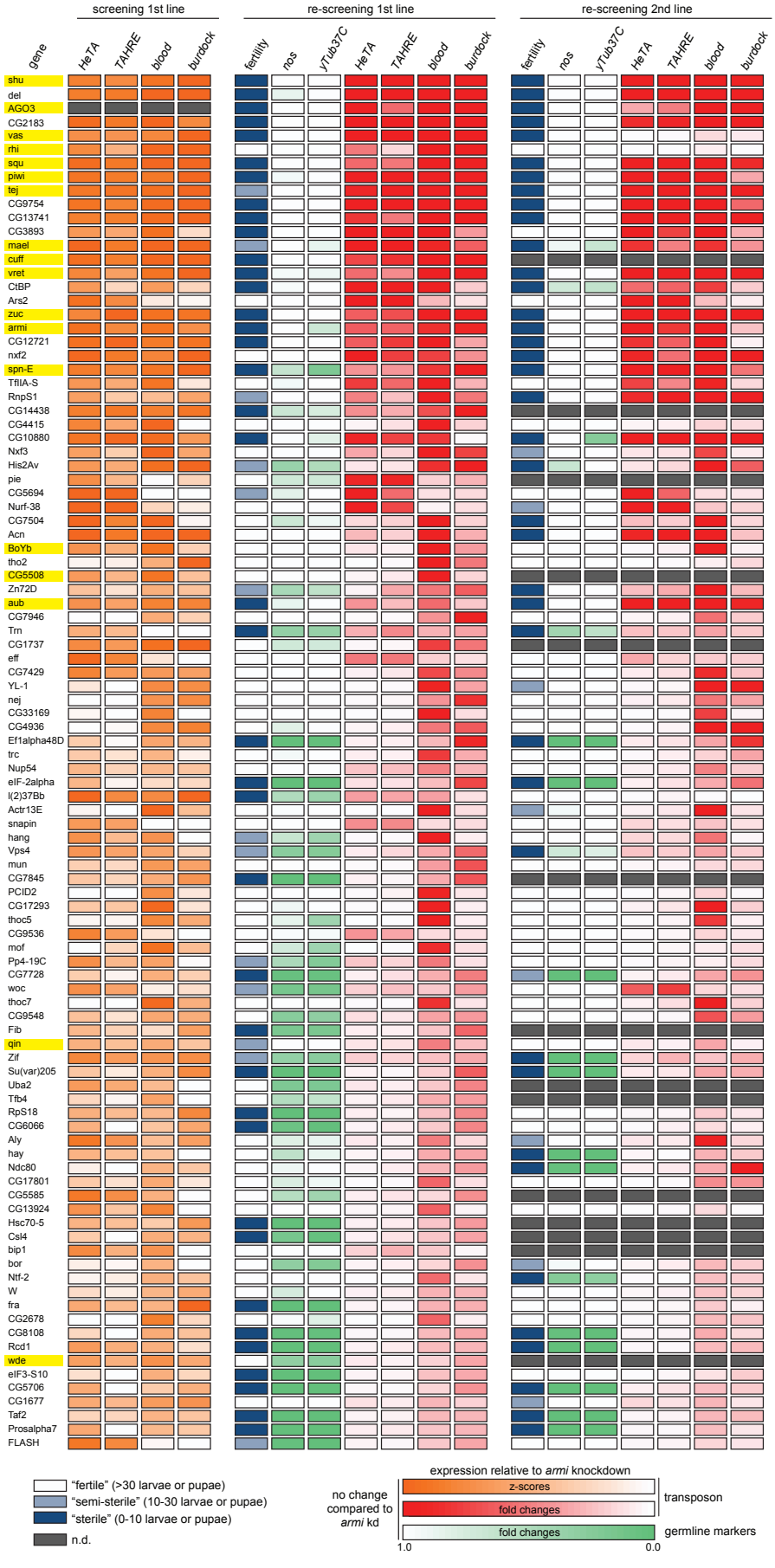
Figure S3. Schematic domain structure of murine and *Drosophila* GASZ. The ankyrin repeat domain (green), sterile alpha motif (SAM) (red), and putative basic leucine zipper domain (yellow) are indicated. E-values represent homology with consensus sequence for the indicated domains. *Dm* = *Drosophila melanogaster* ; *Mm* = *Mus musculus*.

Czech, Preall *et al.*, Fig. 1

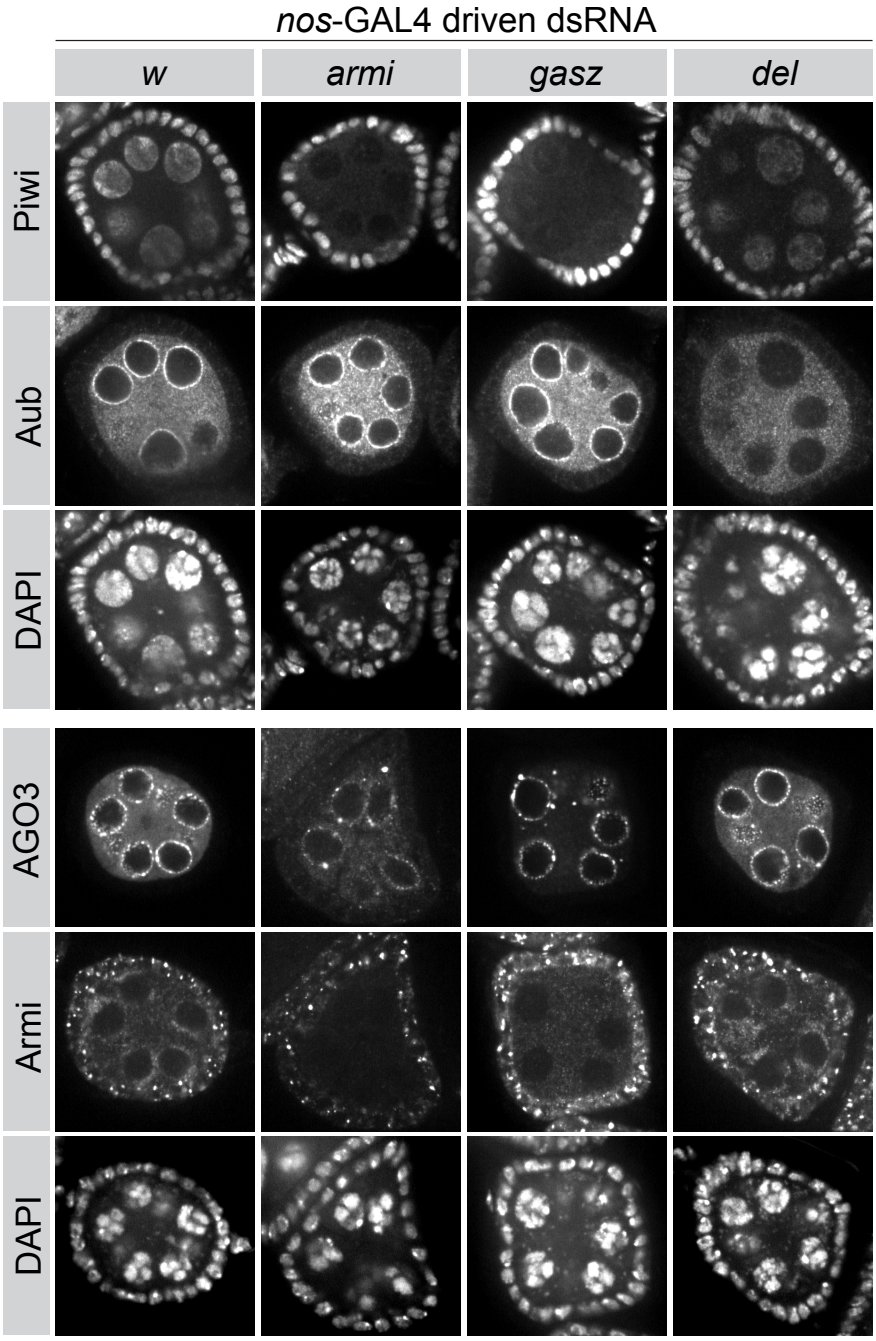


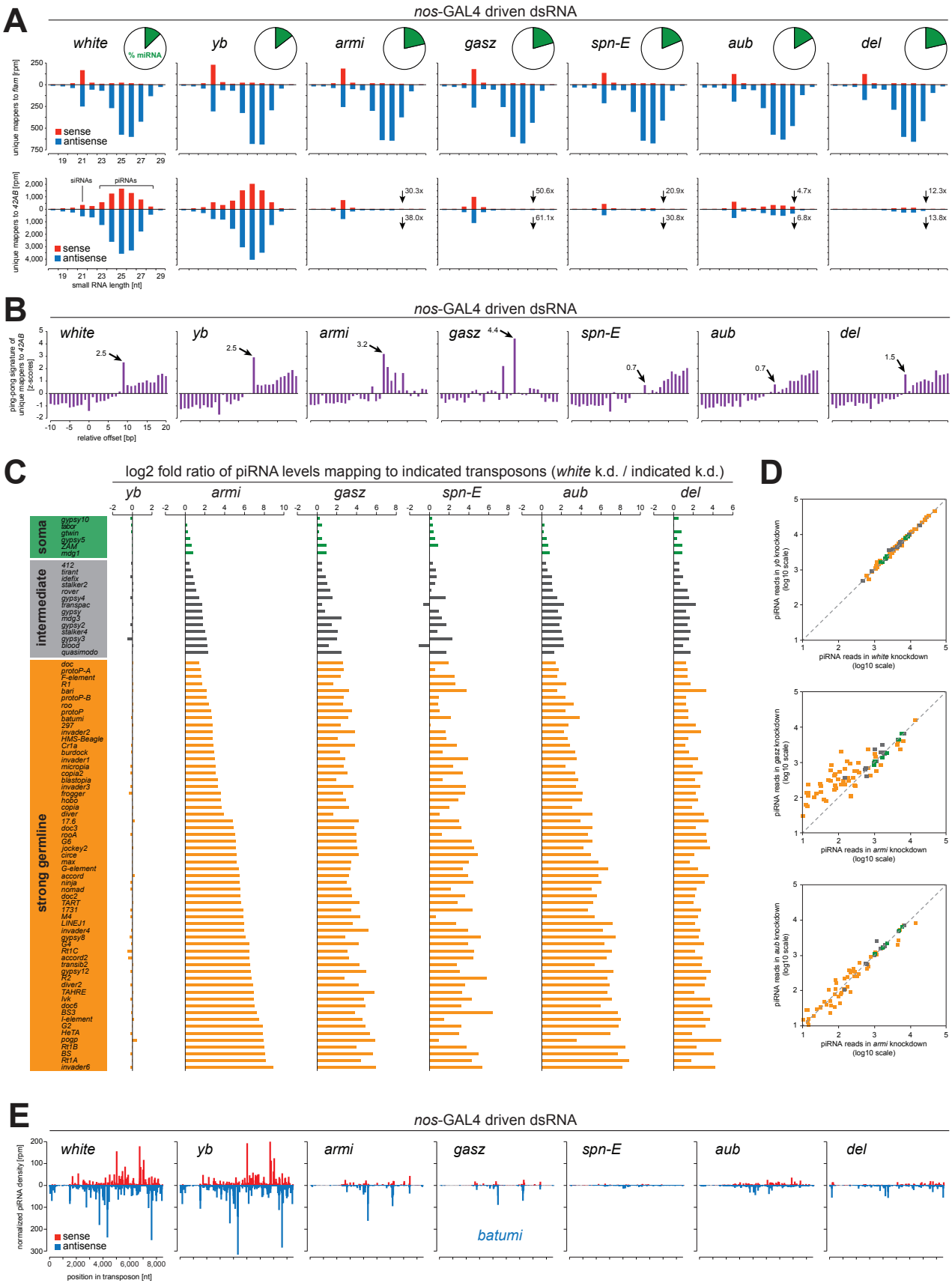

A
average
HeTA
TAHRE
blood
burdock
screened genes (in thousands)
0
1
2
3
4
5
6
7
8
B
average
HeTA
TAHRE
blood
burdock
hits (average z-scores < -1.5)
1
216
strong (74)
weak (142)
transposon de-repression
z-scores
-6
-4
-2
0
2
4
armi controls (-4.43)
white controls (0.64)
transposon repression
C
average of 4 transposable elements
z-scores
2
0
-2
-4
-6
screened genes n = 8,171
armi controls n = 92
white controls n = 92

Czech, Preall *et al.*, Fig. 3

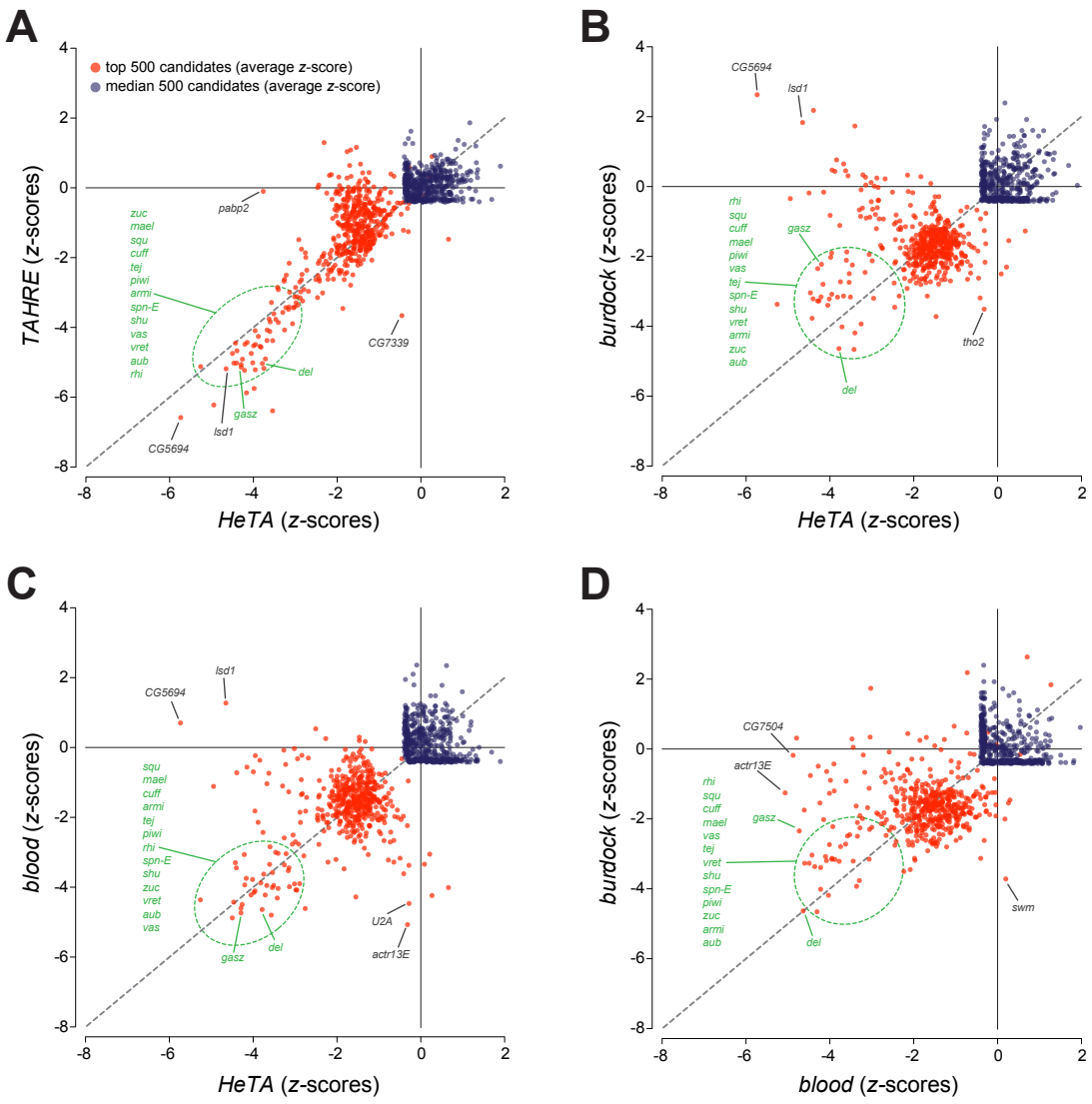


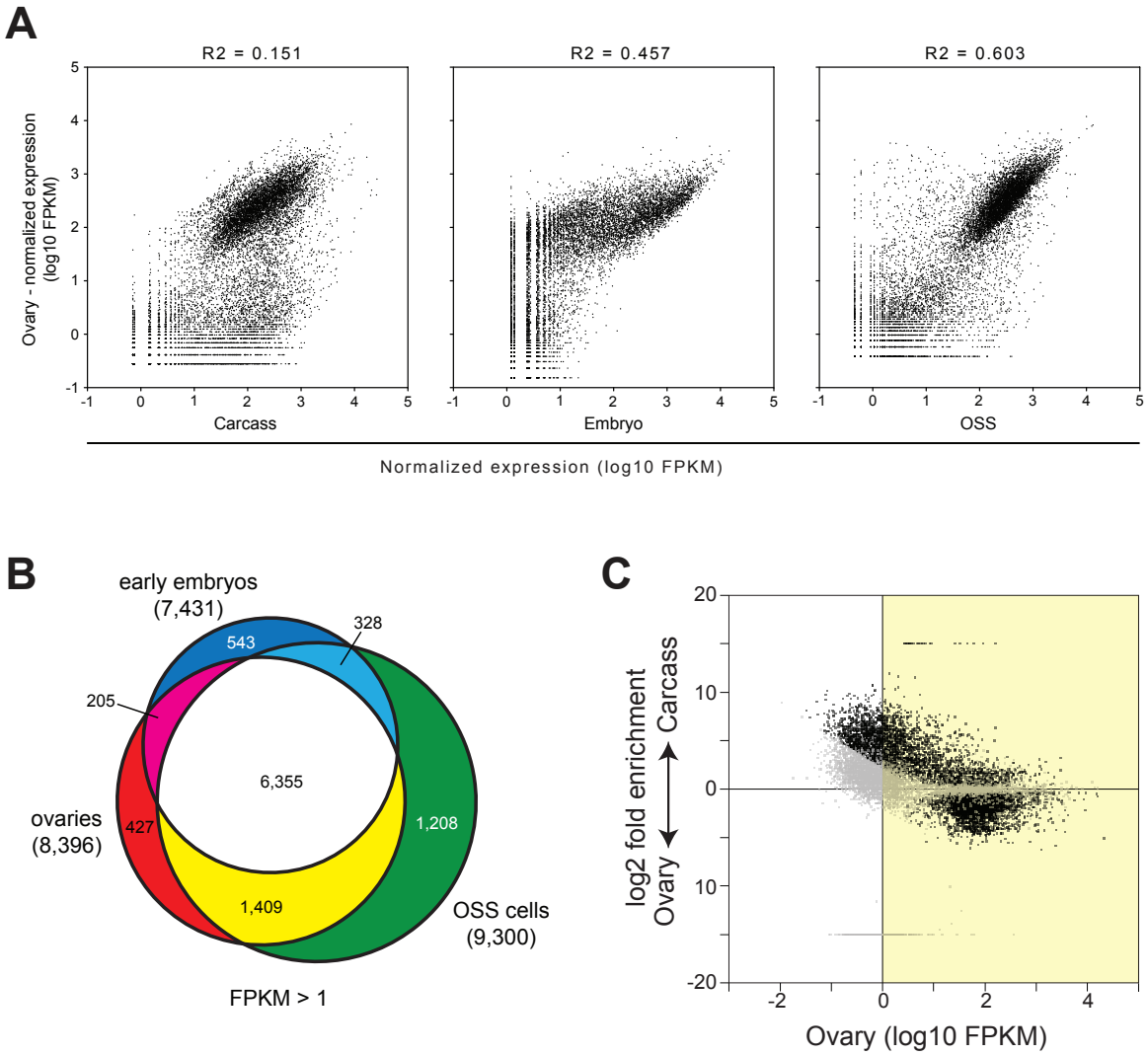
Czech, Preall *et al.*, Fig. 4



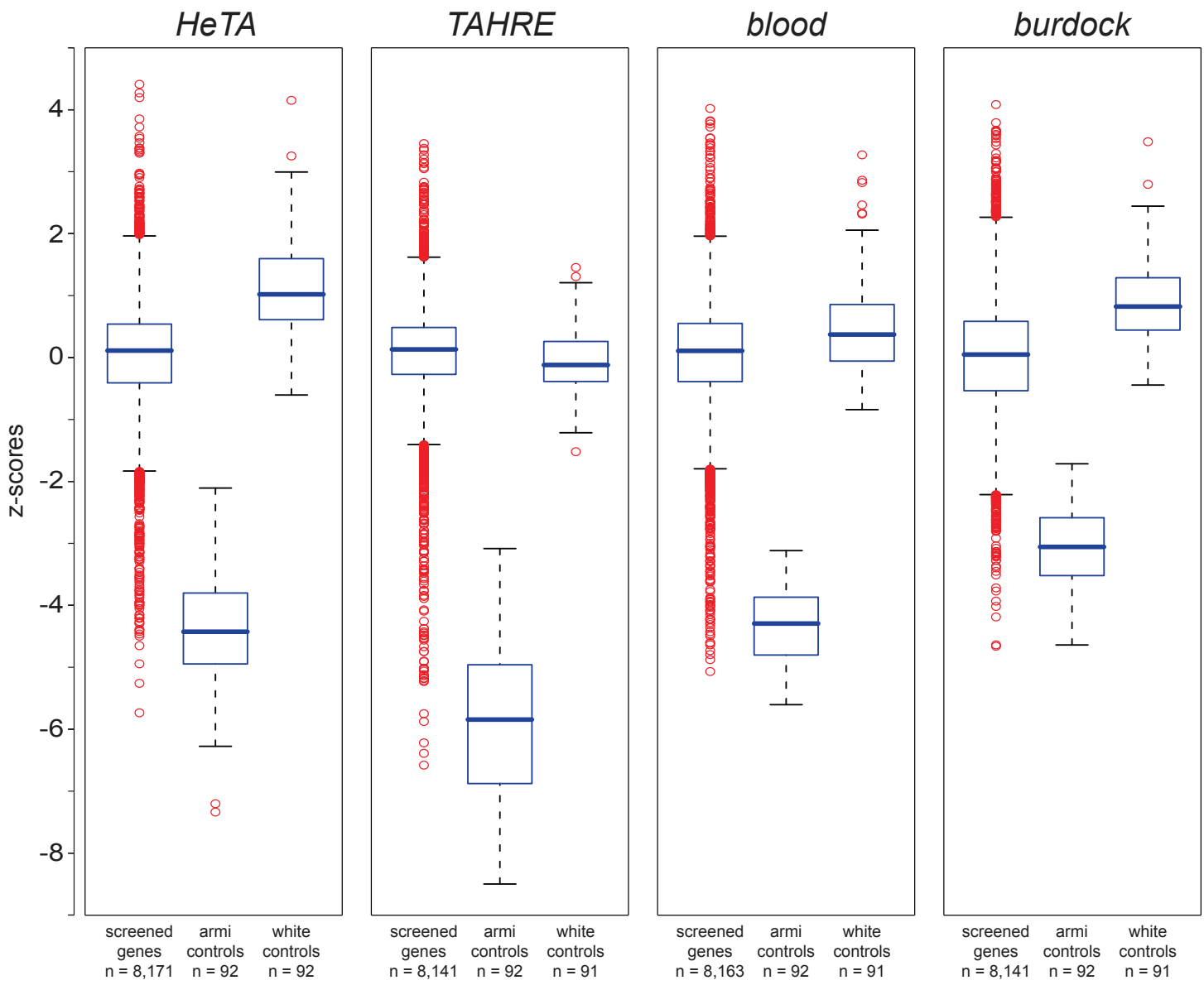

A
nos-GAL4 driven dsRNA
white
yb
armi
gasz
spn-E
aub
del
% miRNA
unique mappers to flam [rpm]
sense
antisense
unique mappers to 42AB [rpm]
siRNAs
piRNAs
30.3x
38.0x
50.6x
61.1x
20.9x
30.8x
4.7x
6.8x
12.3x
13.8x
small RNA length [nt]
B
nos-GAL4 driven dsRNA
ping-pong signature of unique mappers to 42AB [z-scores]
relative offset [bp]
C
log2 fold ratio of piRNA levels mapping to indicated transposons (white k.d. / indicated k.d.)
soma
intermediate
strong germline
D
piRNA reads in yb knockdown (log10 scale)
piRNA reads in white knockdown (log10 scale)
piRNA reads in gasz knockdown (log10 scale)
piRNA reads in armi knockdown (log10 scale)
piRNA reads in aub knockdown (log10 scale)
E
nos-GAL4 driven dsRNA
normalized piRNA density [rpm]
batumi
position in transposon [nt]

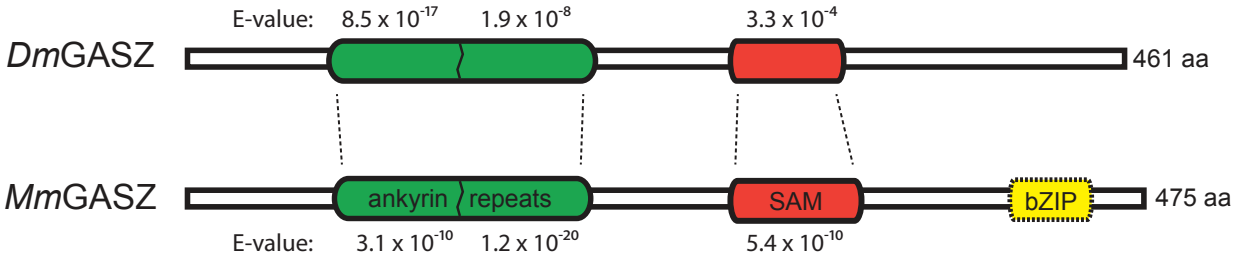
Czech, Preall *et al.*, Fig. 6




A
R2 = 0.151
R2 = 0.457
R2 = 0.603
Ovary - normalized expression (log10 FPKM)
Carcass
Embryo
OSS
Normalized expression (log10 FPKM)
B
early embryos (7,431)
543
328
205
6,355
ovaries (8,396)
427
1,208
1,409
OSS cells (9,300)
FPKM > 1
C
log2 fold enrichment Ovary Carcass
Ovary (log10 FPKM)

Czech, Preall *et al.*, Fig. S2




E-value: 8.5 x 10-17 1.9 x 10-8
3.3 x 10-4
DmGASZ
461 aa
MmGASZ
ankyrin repeats
SAM
bZIP
475 aa
E-value: 3.1 x 10-10 1.2 x 10-20
5.4 x 10-10

Taqman probes and primers for qPCRs

Name	Sequence	ratio (probe:primer)	dye/quencher
rp49_probe	TTGTCGATACCCTTGGGCTTGCG	1:3	CY5/IBRQ
rp49_forward	GTCGGATCGATATGCTAAGCTG	1:3	CY5/IBRQ
rp49_reverse	CAGATACTGTCCCTTGAAGCG	1:3	CY5/IBRQ
nos_probe	CGACGGGCGCAGCAGAAAATG	1:2	HEX/ZEN/IBFQ
nos_forward	AACTCCTGCATCACATCCTG	1:2	HEX/ZEN/IBFQ
nos_reverse	CGCGATCCTTGAAAATCTTTGC	1:2	HEX/ZEN/IBFQ
yTub37c_probe	CTGCACCTTCTCACCCTGGCTAAA	1:2	HEX/ZEN/IBFQ
yTub37c_forward	GTCCAGAATGTCGAACACCTC	1:2	HEX/ZEN/IBFQ
yTub37c_reverse	ATGACCTCTCCGTACTCCAA	1:2	HEX/ZEN/IBFQ
HeTA_probe	CCGCTCATCACTCCCGTTTTCAAGA	1:1	6-FAM/ZEN/IBFQ
HeTA_forward	TCCAACTTTGTAACTCCCAGC	1:1	6-FAM/ZEN/IBFQ
HeTA_reverse	TTCTGGCTTTGGATTCCTCG	1:1	6-FAM/ZEN/IBFQ
TAHRE_probe	TCACCAGAGCAGTTGACGCAGG	1:1	HEX/ZEN/IBFQ
TAHRE_forward	GCACAAAGCCAAGAACCAC	1:1	HEX/ZEN/IBFQ
TAHRE_reverse	GCATCCCTTGTACGCACTAA	1:1	HEX/ZEN/IBFQ
blood_probe	TGCCTCACGGTCGCCATGTAAT	1:1	6-FAM/ZEN/IBFQ
blood_forward	GTTGTCGCTGTTTAGAATTCCC	1:1	6-FAM/ZEN/IBFQ
blood_reverse	CCAACTGCAAGGAAAACACG	1:1	6-FAM/ZEN/IBFQ
burdock_probe	CGCATCGCAACCCCAACAGC	1:1	6-FAM/ZEN/IBFQ
burdock_forward	GCCATCCCAACAGCAAAATTC	1:1	6-FAM/ZEN/IBFQ
burdock_reverse	CTGAGCCTGACTTGTGTTTTG	1:1	6-FAM/ZEN/IBFQ

Primers for SYBR qPCRs

Name	Sequence
rp49_forward	ATCTCGCCGCAGTAAACGC
rp49_reverse	CCGCTTCAAGGGACAGTATCTG
TART_forward	ACCAGGGAAAAGTGTGAACG
TART_reverse	GGTGCAGTGGTATGGCTTTT
TAHRE_forward	CTGTTGCACAAAGCCAAGAA
TAHRE_reverse	GTTGGTAATGTTCGCGTCCT
HeTA_forward	CGCGCGGAACCCATCTTCAGA
HeTA_reverse	CGCCGCAGTCGTTTGGTGAGT
burdock_forward	AGGGAAATATTTGGCCATCC
burdock_reverse	TTTTGGCCCTGTAAACCTTG
micropia_forward	CGAATGTTACGCGGTGTATG
micropia_reverse	CTGGTCAGGTCCAAGGTTGT
Max_forward	ATCTAGCCAGTCGAGGCGTA
Max_reverse	TGGAAGAGTGTCGCTTTGTG
Copia_forward	AGCAAACAACCCCTCATGTC
Copia_reverse	GCAAACCCAATTTGTCTCGT
Zam_forward	ACTTGACCTGGATACACTCACAAC
ZAM_reverse	GAGTATTACGGCGACTAGGGATAC