



universität
wien

DISSERTATION

Titel der Dissertation

Expression and Function of the Mouse Argonaute Protein Miwi2.

angestrebter akademischer Grad

Doktorin der Naturwissenschaften (Dr. rer.nat.)

Verfasserin	Dubravka Pezic
Matrikel-Nummer:	0547845
Dissertationsgebiet (lt. Studienblatt):	Molekularbiologie
Betreuerin / Betreuer:	Dr. Javier Martinez

Wien, am 28. Juni 2009

1. SUMMARY.....	3
2. ZUSAMMENFASSUNG.....	4
3. INTRODUCTION.....	5
3. I. RNA SILENCING: THE CONCEPTS.....	5
3. II. THE HISTORY OF RNA SILENCING.....	7
3. III. AT THE CORE OF RNA SILENCING: ARGONAUTES AND SMALL RNAS.....	9
3. III.a. Argonaute proteins.....	9
3. III.b. Small RNAs.....	11
3. IV. PIWIS AND PIRNAS.....	17
3. IV.a. Amplification of piRNA-mediated silencing of transposons: the ping-pong cycle.....	18
3. IV. b. Piwis in the mouse: Miwi, Mili, and Miwi2.....	21
3. IV. c. Miwi2 and Mili in male fetal gonads.....	25
3. V. DNA METHYLATION IN CONTROL OF TRANSPOSITION AND IMPRINTING.....	28
4. MATERIALS AND METHODS.....	31
5. RESULTS AND DISCUSSION.....	41
5. I. ANALYSIS OF MIWI2 MRNA EXPRESSION.....	42
5. I. 1. The ubiquitous expression of miwi2 mRNA.....	42
5. I. 2. Analysis of the 5' end of miwi2: detection of two differentially expressed full-length transcript variants in testis and retention of intron 1.....	48
5. I. 3. Analysis of transcription from the miwi2 locus in the Miwi2 knockout mice.....	53
5. I. 4. 5'Rapid Amplification of cDNA Ends (RACE) analysis of miwi2.....	61
5. I.5. Analysis of the miwi2 locus: presence of alternative transcriptional starting sites.....	63
5. I. 6. Identified miwi2 transcriptional and splicing variants.....	66
5. II. ANALYSIS OF PUTATIVE DIFFERENT FORMS OF MIWI2, THEIR EXPRESSION OUTSIDE OF THE GERM LINE, SUB-CELLULAR LOCALIZATION, AND POTENTIAL FUNCTIONS.....	70
5. II. 1. The Miwi2 antibodies.....	70
5. II. 2. Detection of Miwi2 using the N-terminal antibody.....	71
5. II. 3. siRNA-mediated knockdown of Miwi2 in MEFs and protein detection using the C-terminal antibody.....	74
5. II. 5. Immuno-detection of Miwi2 in tissues.....	79
5.II. 6. Miwi2 in the sex body: co-localization with markers.....	84
5. II.7. N-terminal antibody specificity.....	89
5. II. 8. Detection of the putative alternative form of Miwi2 in tissues and cells using the C-terminal antibody.....	91
5. III. A POSSIBLE ALTERNATIVE FUNCTION OF MIWI2 IN REGULATION OF IMPRINTING.....	94
6. CONCLUSIONS.....	100
7. REFERENCES.....	102
7. ACKNOWLEDGMENTS.....	117
8. CURRICULUM VITAE.....	119
9. APPENDIX.....	123
PUBLICATION.....	123

1. Summary.

RNA silencing mechanisms employ small RNA molecules, 20-30 nucleotides long, to down-regulate the expression of genes. Small RNAs are always associated with proteins from the Argonaute family, constituting the central effector complex of silencing mechanisms, RNA-induced silencing complex (RISC).

Members of the Piwi subfamily of Argonaute proteins are expressed predominantly in the animal gonadal tissues, and are essential for fertility. They are associated with a specific class of small RNAs, the Piwi-interacting RNAs (piRNAs), and control the expression of transposable elements. The mouse genome encodes three Piwi proteins, called Miwis, and the absence of either of them results in spermatogenesis arrest and consequent male sterility.

In this Thesis, I report a comprehensive analysis of expression and potential alternative functions of Miwi2, in both somatic and germ line adult mouse tissues. I found that the expression of *miwi2* at the RNA level is not restricted to the germ line, as several different coding and non-coding transcripts originating from the *miwi2* locus can be detected at the low levels in the majority of tissues. One of the coding transcripts, potentially originating from an alternative promoter within the *miwi2* locus and corresponding to the 3' region of the full-length *miwi2* mRNA, could potentially give rise to a putative shorter Miwi2 protein form. Furthermore, expression of an imprinted gene, *H19*, seems to be affected by the absence of Miwi2, both in the germ line and somatic tissues.

My results reveal transcriptional complexity of the *miwi2* locus, could indicate the existence of an alternative protein form of Miwi2 expressed in at least some somatic tissues, and suggest that Miwi2 might have additional functions in regulating the expression of imprinted genes.

2. Zusammenfassung.

In RNA „silencing“ Mechanismen finden kleine, 20 bis 30 Nukleotide lange RNA Moleküle Verwendung um die Expression von Genen zu verringern. Diese kleinen RNAs sind immer mit Proteinen der Argonautefamilie assoziiert und bilden so den zentralen Effektor-Komplex (RNA induced silencing complex, RISC) dieser „silencing“ Mechanismen.

Piwi Proteine, eine Unterfamilie der Argonaute-Proteine, werden hauptsächlich in tierischen Gonaden exprimiert und sind für die Fruchtbarkeit essentiell. In Komplexen mit den sogenannten Piwi-assoziierten RNAs (piRNAs) kontrollieren sie die Expression von transposablen Elementen. Das Mausgenom kodiert für drei Piwi-Proteine, die sogenannten Miwis. Die Abwesenheit eines dieser drei Proteine führt zu einem Spermatogenese-Arrest und daraus folgend zur Sterilität.

In dieser Doktorarbeit analysiere ich die Expression und potentielle alternative Funktionen von Miwi2 sowohl in somatischem Gewebe als auch in adultem Keimbahngewebe der Maus. Die Expression von Miwi2 RNA scheint nicht auf die Keimbahn beschränkt zu sein und verschiedene kodierende und nicht-kodierende, vom miwi2 locus stammende Transkripte konnten in den meisten Geweben nachgewiesen werden. Eins der kodierenden Transkripte, welches der 3' Region der vollständigen miwi2 mRNA entspricht und von einem alternativen Promotor zu stammen scheint, könnte zur Expression einer kürzeren Miwi2 Form führen. Ausserdem scheint die Expression eines imprinteten Gens, H19, sowohl in somatischem als auch in Keimbahngewebe in der Abwesenheit von Miwi2 verändert zu sein.

Meine Ergebnisse zeigen erstens die Komplexität des miwi2 locus, könnten zweitens die Präsenz einer alternativen Proteinform in manchen somatischen Geweben bedeuten und implizieren drittens, dass Miwi2 zusätzliche Funktionen bei der Regulierung von imprinteten Genen haben könnte.

3. Introduction.

Correct regulation of gene expression is a central theme in the development of multicellular organisms. Different cells within an organism use a wide variety of mechanisms to ensure that a certain gene is expressed at the right time and place. Dysregulation of gene expression leads to developmental defects, disease, and reduced fitness.

The discovery that double-stranded RNA molecules are very potent inducers of gene silencing was made 11 years ago (Fire *et al. Nature* 1998), and it dramatically changed the understanding of gene expression regulation. The phenomenon, named RNA interference (RNAi), drew attention to the wide variety of silencing mechanisms induced by and executed through RNA molecules (reviewed by Ghildiyal and Zamore, *Nat. Rev. Genetics* 2009; Carthew and Sontheimer, *Cell* 2009).

3. 1. RNA silencing: the concepts.

RNA silencing mechanisms are present in all five kingdoms of life and operate on all levels of gene expression regulation: from chromatin structure and organization, through mRNA localization and stability, to translational control. They orchestrate development, combat virus infections and guard the integrity of the genome. Disruption or dysregulation of silencing mechanisms can result in severe developmental defects, infertility, and disease.

The trigger of silencing is an RNA molecule, typically a long double-stranded (dsRNA), which is processed by a conserved machinery into the central species of RNA silencing: small RNAs, 20-30 nucleotides (nt) long. Small RNAs associate with proteins from the Argonaute family to form the main effector complex of all silencing mechanisms: the RNA-induced silencing complex (RISC). Within RISC, small RNAs are able to complementary pair to other nucleic acids - most commonly other RNA molecules, such as cellular messenger RNAs – and induce their silencing. In

canonical pathways, perfect complementarity results in target cleavage, and imperfect complementarity leads to translational silencing of the target (Fig. 1).

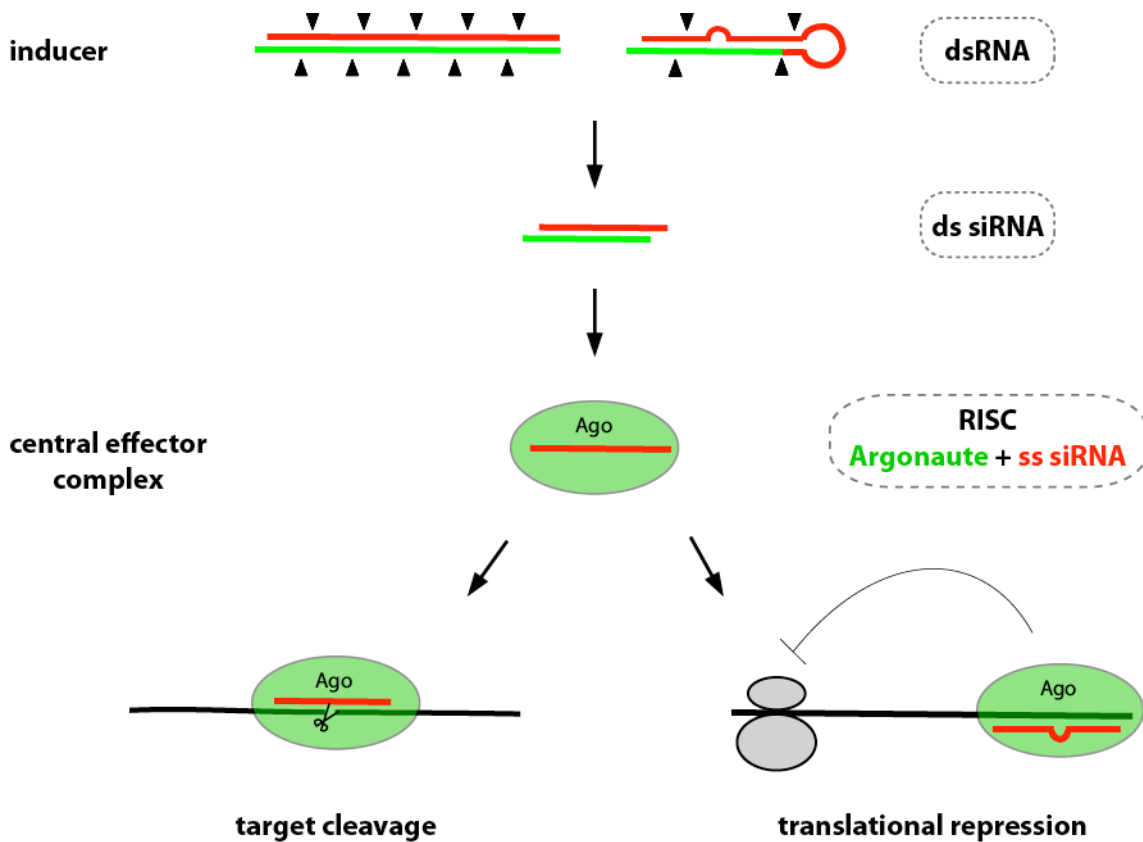


Figure 1. The RNA silencing paradigm.

RNA silencing mechanisms are induced by a long double-stranded (ds) RNA molecule. dsRNA is processed into a short interfering dsRNA (ds siRNA). One of the strands of the ds siRNA, called the guide strand, and a protein from the Argonaute family are assembled into the central effector complex: the RNA-induced silencing complex (RISC). Single-stranded small RNA (ss siRNA) enables recognition of the target RNA via complementary base-pairing. RISC exerts target silencing, by either cleaving the target RNA in the case of full complementarity between the ss siRNA and the target, or repressing its translation in the case of partial complementarity.

All RNA silencing mechanisms depend on the complementarity between small RNAs and their targets, a feature that provides exquisite specificity and potency in

regulating gene expression (reviewed by Ghildiyal and Zamore, *Nat. Rev. Genetics* 2009; Carthew and Sontheimer, *Cell* 2009).

3. II. The history of RNA silencing.

The story of RNA silencing mechanisms converged in the last decade from two main directions. One was the discovery of microRNAs (miRNAs), genome-encoded small RNAs – about 21 nucleotides (nt) long - that act mainly by repressing translation of their cognate mRNAs (Lee *et al. Cell* 1993; Wightman *et al. Cell* 1993; Lagos-Quintana *et al. Science* 2001; Lee and Ambros, *Science* 2001; Lau *et al. Science* 2001). The second was the discovery of RNA interference, a process in which small interfering RNAs (siRNAs) – at the average size of 22 nt - guide the cleavage of complementary mRNAs (Fire *et al. Nature* 1998; Zamore *et al. Cell* 2000; Elbashir *et al. Genes Dev* 2001; Elbashir *et al. EMBO J.* 2001).

It took several years and contributions from several different fields of research to bring the attention to the existence of RNA-induced silencing mechanisms, and a few more to decipher their common characteristics, core machinery, and a variety of ways in which they operate.

Two papers from the lab of Richard Jorgensen published in 1990 (Napoli *et al. Plant Cell* 1990; van der Krol *et al. Plant Cell* 1990) are probably the first written record of RNA silencing. The investigators observed that upon introduction of a transgene into petunias that was supposed to induce deepening of flower color, the flowers unexpectedly showed variegated pigmentation, and some lacked pigment altogether. This meant that not only was the transgene not active, but it also induced silencing of the endogenous locus. Similar observations were made in the filamentous fungus *Neurospora crassa*, where the phenomenon was termed 'quelling' (Romano and Macino, *Mol. Microbiol.* 1992; Cogoni and Macino, *PNAS* 1997). At the same time, several labs observed that plants responded to replicating RNA viruses by post-transcriptionally silencing viral RNA, as well as endogenous genes that shared homology to sequences included in the viral replicon (Dougherty *et al. Mol. Plant Microbe Interact.* 1994; Angell & Baulcombe *EMBO J.* 1997; Ruiz *et al. Plant Cell*

1998), and it was proposed that the two phenomena might be mechanistically related (Ratcliff *et al. Science* 1997). But it was only when Andrew Fire and Craig Mello observed that dsRNA was much more potent in inducing heritable gene-specific silencing in the nematode *C. elegans* than the antisense single-stranded RNA (Fire *et al. Nature* 1998) that researchers started realizing similarities between responses to dsRNA in different organisms: it was able to induce specific silencing of any sequence that it shared homology to. They named the phenomenon RNA interference (RNAi). A year later, the new class of RNAs, ~25 nt long small RNAs, derived from long dsRNAs upon introduction of a transgene or viral infection, were discovered in plants (Hamilton and Baulcombe, *Science* 1999). Soon afterwards, it became clear that these unique molecules guide the silencing of their complementary sequences, explaining the great specificity of RNA silencing mechanisms (Zamore *et al. Cell* 2000).

The importance of RNA silencing was acknowledged by awarding the Nobel Prize to Andrew Fire and Craig Mello in 1996, only 8 years after their milestone paper.

Concomitantly, the labs of Victor Ambros and Gary Ruvkun discovered in 1993 that *lin-4*, essential for normal development of *C.elegans*, is a non-protein-coding locus producing transcripts that negatively regulate levels of LIN-14 protein (Lee *et al. Cell* 1993; Wightman *et al. Cell* 1993). They anticipated that this could represent the paradigm for a new way of gene expression control, but they probably did not think that they had just hit the tip of the iceberg. They identified two species of *lin-4*, a ~60-nucleotide (nt) long *lin-4L* and 22-nt long *lin-4S*. They discovered several sites in the 3'UTR of *lin-14* transcript that *lin-4S* could complementary pair to. The region of complementarity was short – only 10 nt – but absolutely essential for *lin-4* function. Also, the levels of *lin-14* transcript were stable, so they suggested that binding of *lin-4* to its complementary sites in the 3'UTR of *lin-14* inhibits its translation.

Today, we know that they discovered the first miRNA (*lin-4S*), its precursor (*lin-4L*), and its target (*lin-14*). They discovered what is today known as 'the seed' of miRNAs

– 10 nucleotides of complementarity between *lin-4* and its complementary sites in the 3'UTR. They also discovered the hairpin structure of miRNA precursors. And they did discover a novel pathway of gene regulation: inhibition of translation by miRNAs binding to their complementary sites in the 3'UTRs of target transcripts.

Argonaute proteins, the second essential component of RISC, were discovered in plants and flies as factors affecting stem cell differentiation and growth (Lin and Spradling, *Development* 1997; Cox *et al. Genes Dev.* 1998; Bohmert *et al. EMBO J.* 1998). Subsequent genetic and biochemical studies established them as the main protein component required for target repression in all RNA silencing mechanisms (Tabara *et al. Cell* 1999; Fagard *et al. PNAS* 2000; Hammond *et al. Science* 2001; Carmell *et al. Genes Dev.* 2002; Pal-Bhadra *et al. Mol. Cell* 2002, Williams and Rubin, *PNAS* 2002).

3. III. At the core of RNA silencing: Argonautes and small RNAs.

3. III.a. Argonaute proteins.

Argonaute proteins and their role as the main endonucleases of RNA silencing mechanisms are conserved throughout plant and animal kingdoms, as well as in fungi (Cox *et al. Genes Dev.* 1998; Fagard *et al. PNAS* 2000; Carmell *et al. Genes Dev.* 2002). They are grouped into three subfamilies: Argonaute, Piwi, and WAGO (Faehle and Joshua-Tor, *Curr. Opinion Chem. Biol.* 2007). The Argonaute subfamily contains proteins with strongest homology to *Arabidopsis* Ago1; these proteins are ubiquitously expressed, and are involved in classical RNAi and microRNA pathways (Carmell *et al. Genes Dev.* 2002; Ghildiyal and Zamore, *Nat. Rev. Genetics* 2009). Piwi proteins – sharing strongest homology to *Drosophila* Piwi – are animal-specific, and show mainly germ line-restricted expression (Seto *et al. Mol. Cell* 2007; Malone and Hannon, *Cell* 2009). They were discovered as factors involved in stem cell maintenance in the *Drosophila* germ line (Lin and Spradling, *Development* 1997; Cox *et al. Genes Dev.* 1998; Cox *et al. Development* 2000) and genetic and biochemical studies have shown that they are essential for the control of transposon expression in the germ line (Vagin *et al. Science* 2006; Saito *et al. Genes Dev.* 2005;

Carmell *et al. Dev. Cell* 2007; Houwing *et al. Cell* 2007). Like the other Argonautes, Piwis function in complexes with small RNAs – the Piwi-interacting RNAs or piRNAs. The members of the third subfamily have been found only in *C. elegans*, and are therefore termed WAGOs – worm Argonautes (Yigit *et al. Cell* 2006). They bind secondary siRNAs and are important for amplification of RNAi.

The genomes of different organisms encode various numbers of Argonautes from different subfamilies (Carmell *et al. Genes Dev.* 2002; Tolia and Joshua-Tor; *Nat. Chem. Biol.* 2007). Fission yeast has only one Argonaute protein. Plants have ten Argonautes and no Piwis (Vaucheret, *Trends Plant Sci.* 2008). Humans have four Argonautes and four Piwis, and mice have five Argonautes and three Piwis. There are three Piwis and two Argonautes in the *Drosophila* genome. *C. elegans* genome contains 27 different Argonautes, two of which are Piwis, and ~18 belonging to the WAGO subfamily.

The proteins belonging to the Argonaute family share two signature domains: PAZ and Piwi (Cerutti *et al. Trends Biochem. Sci.* 2000); (Fig. 2). Crystal structures of both domains, as well as of the whole protein, have been solved, and they were instrumental in deciphering the roles of the two domains and the mechanistic details of silencing (Lingel *et al. Nature* 2003; Song *et al. Nat. Struct. Mol. Biol.* 2003; Song *et al. Science* 2004). The PAZ domain is named after proteins that were known to contain it, namely Piwi, Argonaute and Zwiille (Cerutti *et al. Trends Biochem. Sci.* 2000). Dicer proteins also have a PAZ domain (Bernstein *et al. Nature* 2001); this domain is characteristic for proteins involved in RNA silencing. The PAZ domain of Dicer binds 3' two-nucleotide overhangs of long dsRNA duplexes that initiate silencing (MacRae *et al. Science* 2006); in Argonautes, it binds the 3' end of the small RNA (Ma *et al. Nature* 2004). Piwi domain has an RNaseH-like fold and bears the nuclease activity of RISC – it has therefore been termed the 'Slicer' (Song *et al. Science* 2004; Parker *et al. EMBO J.* 2004). The residues responsible for the nuclease activity are two aspartates and a histidine – the so-called DDH motif, or the 'catalytic triad' (Song *et al. Science* 2004; Parker *et al. EMBO J.* 2004). As it has the RNaseH-like fold, the most suitable substrate for Piwi domain is a nucleic acid duplex, in this case an RNA duplex. Small RNAs are therefore pivotal for Piwi function, as they

guide Argonautes to their single-stranded targets and complementary pair with them through Watson-Crick base-pairing, forming duplexes. Piwi catalyzes the hydrolysis of a phosphodiester bond between the nucleotides of the target mRNA that are paired to nucleotides 10 and 11 of the small RNA guide strand (Elbashir *et al. Genes Dev.* 2001; Elbashir *et al. EMBO J.* 2001).



Figure 2. Schematic view of the two main domains of the Argonaute proteins, PAZ and Piwi.

PAZ domain is located N-terminally to the Piwi domain, and binds the single-stranded 3' end of small RNAs. Piwi domain is C-terminal and an RNaseH-like fold and bears endonucleolytic activity responsible for target silencing. The 5' end of the small RNA is anchored in the mid-domain on the N-terminal end of the Piwi domain.

Not all Argonaute proteins are catalytically active: the only human Argonaute known to have 'slicer' activity is Ago2 (Liu *et al. Science* 2004). Secondary Argonautes (SAGOs) from the WAGO subfamily lack the catalytic DDH motif altogether (Faehnle and Joshua-Tor, *Curr. Opinion Chem. Biol.* 2007).

3. III.b. Small RNAs.

A small RNA is defined by its origin, mechanism of biogenesis, length, Argonaute protein it associates with, its targets, mechanisms of target recognition and repression, and finally, its function.

Three major classes of small RNAs, siRNAs, miRNA and piRNAs, share some common characteristics, and also have many distinctive, defining features.

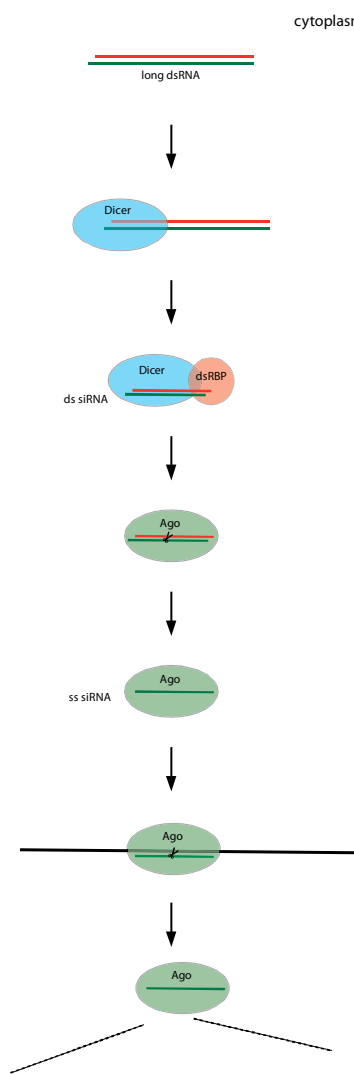
3. III. b.1. siRNAs.

Derived from either exogenously introduced or endogenous long dsRNA, siRNAs are perfect duplexes, 21-22 nt long, with 2-nucleotide overhangs on the 3' ends and a 5' monophosphate (Zamore *et al. Cell* 2000; Elbashir *et al. EMBO J* 2001; Elbashir *et al. Genes Dev.* 2001). They are generated in the cytoplasm in a Dicer-dependant manner (Bernstein *et al. Nature* 2001; Knight *et al. Science* 2001; Ketting *et al. Genes Dev* 2001; Ghildiyal *et al. Science* 2008; Tam *et al. Nature* 2008). In complex with dsRNA binding proteins, such as TRBP and PACT in mammals or R2D2 in flies, Dicer loads the siRNA duplex into RISC (Liu *et al. Science* 2003, Tomari *et al. Science* 2004; Chendrimada *et al. Nature* 2005; Haase *et al. EMBO Rep.* 2005; Lee *et al. EMBO J* 2006; Kok *et al. J. Biol. Chem* 2007). Within RISC, one of the strands of the siRNA duplex – the 'guide' strand – directs RISC to its target, whereas the other strand, called the 'passenger' strand, is cleaved by the Argonaute and discarded (Matranga *et al. Cell* 2005; Rand *et al. Cell* 2005; Miyoshi *et al. Genes Dev.* 2005; Leuschner *et al. EMBO Rep.* 2006). Thermodynamic differences in stability of the two ends of the siRNA duplex determine which of the two strands becomes the guide, and which is discarded as passenger: the strand with less stable 5' end is preferentially chosen as the guide strand (Tomari *et al. Science* 2004; Schwarz *et al. Cell* 2003). siRNAs assembled into RISC predominantly act posttranscriptionally in the cytoplasm, guiding the RISC to their mRNA targets. They pair fully to their targets within the coding region, which leads to the cleavage of the target by the Argonaute protein, followed by degradation and hence the silencing of the target (Liu *et al. Science* 2004).

The canonical siRNA pathway is depicted schematically in Fig. 3a.

The main natural exogenous sources of siRNAs are viral genomes and replication intermediates. In plants and *Drosophila* RNA silencing induced by small RNAs of viral origin acts as the nucleic acid-based immune system (Hamilton and Baulcombe, *Science* 2001; Saleh *et al. Nature* 2009). Both long dsRNA and siRNAs can be experimentally introduced into organisms or cultured cells, resulting in silencing of their complementary mRNAs (Fire *et al. Nature* 1998; Kennerdell and Carthew *Cell*

a. siRNA pathway.



b. miRNA pathway.

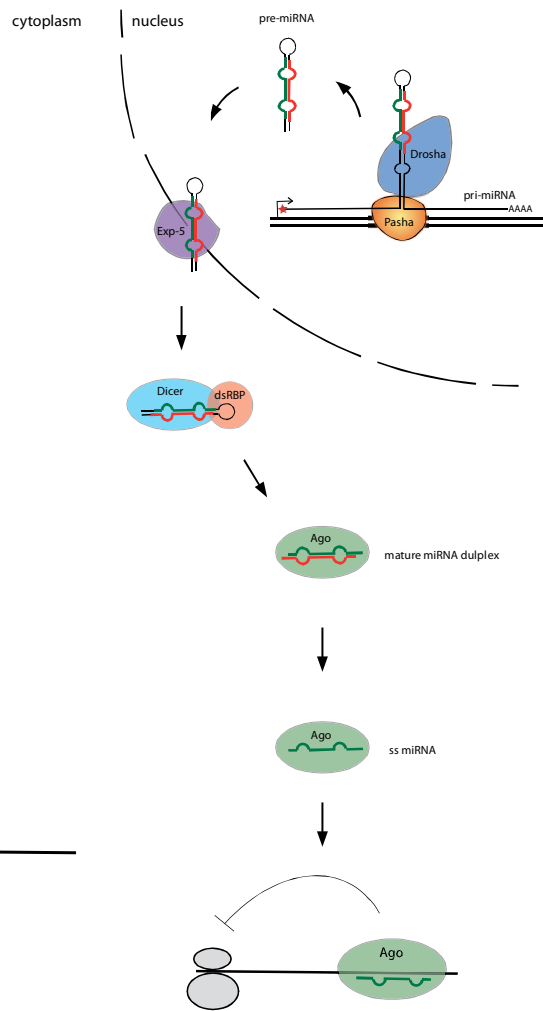


Figure 3. Canonical RNA silencing pathways.

- a.** The siRNA pathway is induced by a long perfect-duplex dsRNA, which is processed by the RNase III enzyme Dicer into siRNAs. siRNAs and proteins from the Argonaute family constitute the main effector complex, RISC. siRNAs pair fully to their targets and induce target cleavage by the endonuclease activity of the Argonaute protein.
- b.** miRNAs are encoded in the genome and transcribed by RNA polymerase II into a capped, polyadenylated precursor transcript, the co-called primary miRNA (pri-miRNA). Pri-miRNA is processed in the nucleus by the RNase III enzyme Drosha into the precursor miRNA (pre-miRNA) and exported to the cytoplasm by Exportin 5. Pre-miRNA is processed into the mature miRNA duplex by Dicer, and assembled into RISC. miRNAs pair partially to their targets, miRISC exerts translational repression of the target mRNA.

1998; Wargelius *et al. Biochem. Biophys. Res. Comm.* 1999; Wianny and Zernicka-Goetz, *Nat. Cell Biol.* 2000; Svoboda *et al. Development* 2000; Clemens *et al. PNAS* 2000; Elbashir *et al. Nature* 2001; Caplen *et al. PNAS* 2001; Harborth *et al. J. Cell Sci.* 2001). This discovery provided us with a powerful tool for studying functions of genes as it mimics the loss-of-function genetic mutations, and is nowadays a common laboratory technique. It also opened new possibilities for developing therapeutics (Bonetta, *Cell* 2009).

Until recently, the only organisms known to express endogenous siRNAs were plants and *C. elegans* (Hamilton and Baulcombe, *Science* 1999; Ambros *et al. Curr. Biol.* 2003). Plants, in particular, contain a variety of different endogenously produced siRNAs (Mallory and Vaucheret, *Nat. Genet.* 2006), and some of them, derived from promoter regions, operate in the nucleus and induce transcriptional silencing (Mette *et al. EMBO J.* 2000; Zilberman *et al. Science* 2003). Furthermore, both plants and *C. elegans* have primary siRNAs, produced from the original trigger, and secondary siRNAs, required for amplification and spreading of silencing (Voinnet, *et al. Cell* 1998; Sijen *et al. Cell* 2001; Deleris *et al. Science* 2006; Pak and Fire, *Science* 2007). Secondary siRNAs require RNA-dependent-RNA polymerase (RdRP) for their production (Vaistij *et al. Plant Cell* 2002; Pak and Fire, *Science* 2007). In plants, they are also Dicer-dependent (Bouche *et al. EMBO J.* 2006; Deleris *et al. Science* 2006), unlike in *C. elegans*, where they are thought to be products of RdRP (Sijen *et al. Science* 2007).

Endo-siRNAs were recently discovered in *Drosophila* and mouse (Ghildiyal *et al. Science* 2008; Czech *et al. Nature* 2008; Okamura *et al. Nature* 2008; Tam *et al. Nature* 2008; Watanabe *et al. Nature* 2008). They are derived from transcripts that have extensive secondary structure, such as pairs of genes and pseudogenes or regions of bidirectional transcription – thereby circumventing the lack of RdRP in the mouse and fly genomes. They originate from transposons, heterochromatic regions, and mRNAs, and associate with Ago2 in both *Drosophila* and mouse. Their exact function remains to be determined.

3. III. b. 2. miRNAs.

Several landmark papers in 2001 described the ubiquitous presence of miRNAs in animal genomes (Lagos-Quintana *et al. Science* 2001; Lee and Ambros, *Science* 2001, Lau *et al. Science* 2001), and revealed that many are conserved (Pasquinelli *et al. Nature* 2000).

miRNAs are derived from RNA polymerase II transcript (Pol II) called primary miRNAs (pri-miRNAs), which form stem-loop structures along its length (Lee *et al. EMBO J.* 2004). Some are encoded in intergenic regions and others within introns of protein-coding genes, often in clusters containing several miRNAs. The stem-loop structures of these transcripts are recognized by the RNase III enzyme Drosha and its dsRNA binding partners, Pasha in flies and DGCR8 in mammals (Lee *et al. Nature* 2003; Denli *et al. Nature* 2004; Gregory *et al. Nature* 2004; Landthaler *et al. Curr. Biol.* 2004). This complex, called the Microprocessor complex, cleaves the pri-miRNA in the nucleus to produce precursor microRNAs (pre-miRNAs) (Han *et al. Genes Dev.* 2004). Precursors of some intron-encoded miRNAs – called mirtrons – are liberated from primary transcripts through the action of splicing machinery, circumventing Drosha (Ruby JG *et al. Nature* 2007; Okamura K *et al. Cell* 2007). Pre-miRNAs are ~70 nt long, have a hairpin structure and bear signature 3' 2-nt overhang ends of RNase III cleavage reaction (Lee *et al. Cell* 1993; Lagos-Quintana *et al. Science* 2001; Lee and Ambros, *Science* 2001; Lau *et al. Science* 2001; Mourelatos *et al. Genes Dev.* 2002; Basyuk *et al. NAR* 2003). The precursor is exported from the nucleus to the cytoplasm by Exportin 5 (Yi *et al. Genes Dev.* 2003; Lund *et al. Science* 2004; Bohnsack *et al. RNA* 2004). In the cytoplasm, the precursor is recognized by Dicer and its dsRNA-binding partners, Loquacious in flies and TRBP in mammals (Hutvagner *et al. Science* 2001; Ketting *et al. Genes Dev.* 2001; Förstemann *et al. PLOS Biol.* 2005; Saito *et al. PLOS Biol.* 2005; Haase *et al. EMBO Rep.* 2005; Chendrimada *et al. Nature* 2005). Dicer cleaves off the loop, leaving ~21nt long mature miRNA duplex (Ketting *et al. Genes Dev.* 2001; Knight and Bass, *Science* 2001; Grishok *et al. Cell* 2001). In plants, but not in animals, the 3' ends of the mature duplex RNA are methylated by the methyltransferase Hen1, making the miRNA more stable (Yu *et al. Science* 2005; Li *et al. Curr. Biol.* 2005). Mature miRNA

duplex is channeled into RISC and bound to the proteins from the Argonaute subfamily (Hutvagner *et al. Science* 2002; Meister *et al. Mol. Cell* 2004). In some cases, as was shown in *Drosophila*, miRNAs are preferentially associated with some members of the Argonaute subfamily, distinct from the ones that bind siRNAs (Foerstemann *et al. Cell* 2007; Tomari *et al. Cell* 2007). Within RISC, miRNA is unwound and one of the strands (the 'miR*' strand) is discarded (Matranga *et al. Cell* 2005).

miRNA recognizes multiple sites in the 3' untranslated regions (3'UTRs) of mRNAs (Lee *et al. Cell* 1993; Wightman *et al. Cell* 1993; Reinhardt *et al. Nature* 2000; Lin *et al. Dev. Cell* 2003). The majority of animal miRNAs pair only partially to their targets; they are fully paired only within a limited region at the 5' end – the 'seed' region, which comprises nucleotides 2-8 (Lai *Nat. Genetics* 2002; Lewis *et al. Cell* 2003; Stark *et al. PLOS Biol.* 2003; Brennecke *et al. PLOS Biol.* 2005) – and induce translational repression, destabilization of the target, or both (Lee *et al. Cell* 1993; Wightman *et al. Cell* 1993; Lim *et al. Nature* 2005; Bartel *Cell* 2009).

miRNAs have emerged as essential regulators of gene expression, particularly in developmental context and in relation to diseases (Carthew and Sontheimer, *Cell* 2009; Ventura and Jacks, *Cell* 2009). Some predictions state that up to 60% of all mRNAs might be regulated by miRNAs. The current release of miRBase (13.0), the registry of miRNAs, lists 9169 mature miRNA products from 103 species.

Animal miRNA pathway is depicted schematically in Fig. 3b.

piRNAs are addressed in detail later in the text.

Many variations on these canonical pathways have been described in various organisms. Like the siRNAs produced from the promoter regions in plants, the siRNAs originating from centromeric repeats in fission yeast induce transcriptional silencing by inducing DNA and/or histone methylation (Mette *et al. EMBO J.* 2000; Zilberman *et al. Science* 2003; Verdel *et al. Science* 2004; Noma *et al. Nat. Genetics* 2004). Furthermore, microRNAs in plants, as well as some animal miRNAs, pair fully to their targets and induce their cleavage, as well as translational repression (Rhoades *et al. Cell* 2002; Tang *et al. Genes Dev.* 2003; Yekta *et al. Science* 2004; Voinnet, *Cell* 2009).

In the ciliate *Tetrahymena thermophila*, small RNAs called scan RNAs (scnRNAs) guide excision of internal eliminated sequences (IES) from the genome of the developing macronucleus during conjugation (Mochizuki *et al. Cell* 2002; Taverna *et al. Cell* 2002; Liu *et al. PNAS* 2004). In *Neurospora*, sequences not paired to their homologous sequences during meiotic prophase I, which is often the case for transposable elements, generate a signal that leads to their silencing, as well as silencing of all the sequences homologous to them, in the process called meiotic silencing by unpaired DNA (MSUD) (Shiu *et al. Cell* 2001). Transcripts from repetitive genomic regions, often comprising defective transposons, give rise to repeat-associated small RNAs (rasiRNAs) (Aravin *et al. Curr. Biol.* 2001; Aravin *et al. Dev. Cell* 2003). rasiRNAs were later described to be a subset of the third major class of small RNAs, the piRNAs, in flies, mice, and fish (Vagin *et al. Science* 2006; Lau *et al. Science* 2006; Houwing *et al. Cell* 2007; Aravin *et al. Science* 2007).

3. IV. Piwis and piRNAs.

The research of Piwis greatly intensified upon discovery of small RNAs associated with them – the Piwi-interacting RNAs, or piRNAs (Girard *et al. Nature* 2006; Aravin *et al. Nature* 2006; Vagin *et al. Science* 2006; Saito *et al. Genes&Dev* 2006; Watanabe *et al. Genes&Dev* 2006). The discovery of piRNAs and their deep sequencing and subsequent mapping onto the genome shed new light on the function of Piwis and provided insights into their mechanisms of action.

piRNAs, by and large, map onto large, repeat- and transposon-rich genomic clusters (with the exception of mouse pachytene piRNAs; discussed later in the text) (Vagin *et al. Science* 2006; Brennecke *et al. Cell* 2007; Aravin *et al. Science* 2007; Houwing *et al. Cell* 2007). They display profound strand bias, being generated preferentially from one strand at any given site within the locus. piRNAs are longer than siRNAs and miRNAs, ranging in size from 24 to 30 nucleotides. The biogenesis of piRNAs is Dicer-independent as they originate from single-stranded transcripts;

they display U1 bias and their 3' ends carry a 2'O-methyl modification (reviewed by Malone and Hannon, *Cell* 2009).

3. IV.a. Amplification of piRNA-mediated silencing of transposons: the ping-pong cycle.

One of the main characteristics of piRNAs is their active amplification in response to transposon activation. This is best described in *Drosophila*, where piRNAs are known to arise from long single-stranded transcripts of genomic loci containing a large number of defective transposon copies (Brennecke *et al.* *Cell* 2007). These loci act as master regulators of transposon expression. piRNAs processed from the primary transcripts are bound to two members of the Piwi subfamily, Aubergine and Piwi. These primary piRNAs are antisense to transposon transcripts, and that enables them to recognize these transcripts by base complementarity. Transposon transcripts are then cleaved and effectively silenced through the 'slicer' activity of Piwi proteins. In this process, secondary piRNAs are generated that are sense to transposons and bound to the third Piwi subfamily member, Ago3. These complexes can then again recognize the primary transcripts from genomic loci, cleave them and generate new piRNAs. This cycle of interrelated transposon transcript cleavage and production of new piRNAs has been termed the 'ping-pong' cycle (Fig. 4). It essentially makes the main weapon of transposons – their spreading through transcription – also their biggest weakness.

The 'ping-pong' cycle leaves certain signatures that point to regulation of transposons by Piwi proteins (Fig. 5). In most cases, primary piRNAs show the U1 bias. Argonautes cleave their targets at defined distance from the 5' end of the guide strand – in this case the primary piRNAs: it is always the phosphodiester bond between nucleotides that pair to nucleotides at positions 10 and 11 of the guide that is cleaved. This would mean that the secondary piRNAs should share a 10-nt complementarity to the primary piRNAs, and instead of a 5'U bias, display an A-bias at position 10.

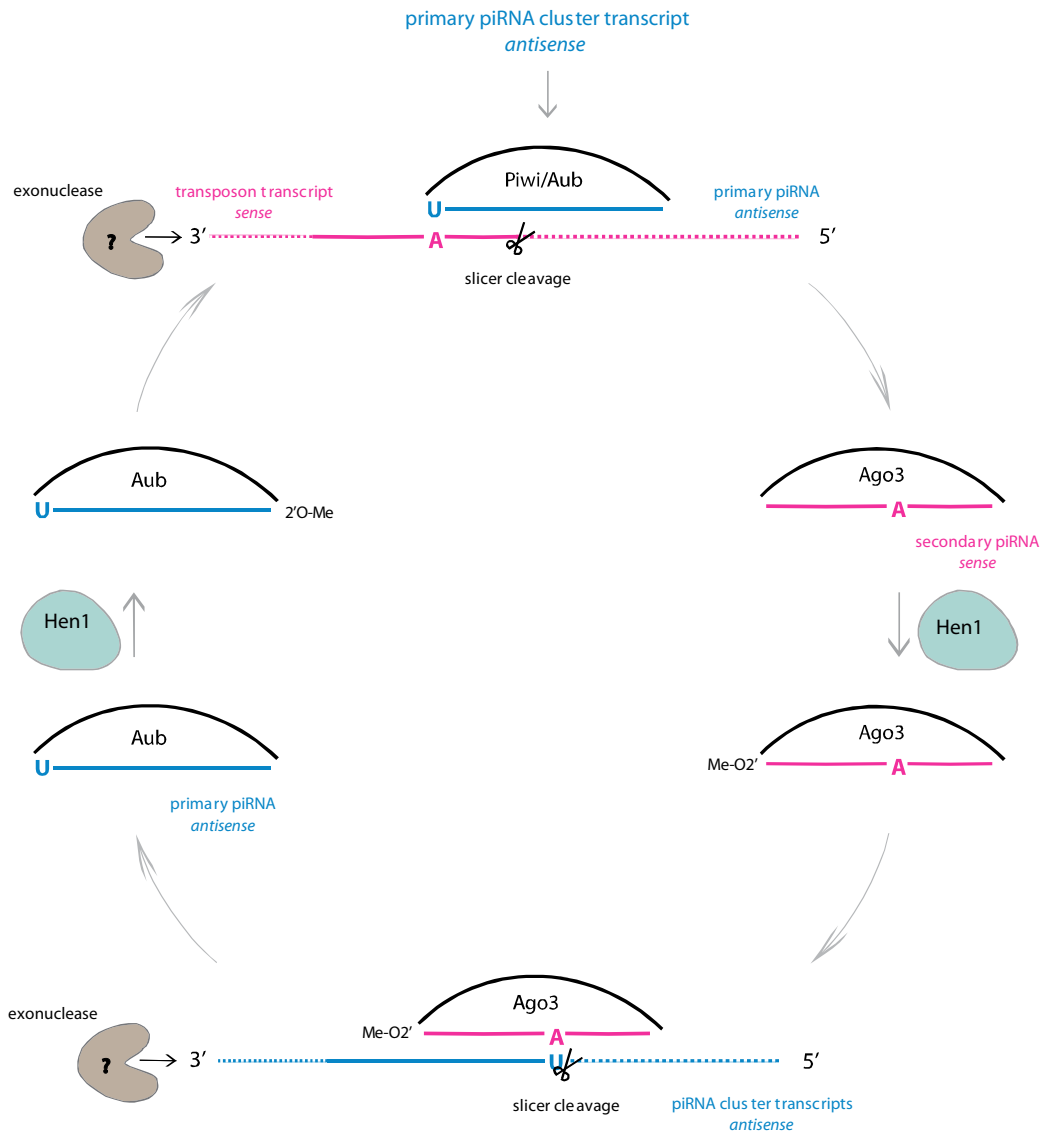


Figure 4. Canonical ping-pong cycle (*Drosophila*).

piRNA clusters are transcribed and processed by an unknown mechanisms into primary piRNAs, which are predominantly antisense to transposon transcripts and assemble into complexes with the two members of the Piwi subfamily of Argonaute proteins, Piwi and Aubergine. These piRISCs cleave the transcripts from the active transposons, thereby silencing their expression and creating new piRNAs. These piRNAs are sense in respect to transposon mRNAs and assemble into piRISC with the third member of the Piwi subfamily, Ago3. 'Sense' piRISC cleaves again the primary transcript from the piRNA clusters, leading to production of the new antisense piRNA.

Exactly this was observed when piRNA populations from complexes with Aubergine and Piwi were compared to the ones from Ago3 complexes. So, the 5' end of new piRNAs is generated through Piwi protein-catalyzed cleavage of transposon transcripts.

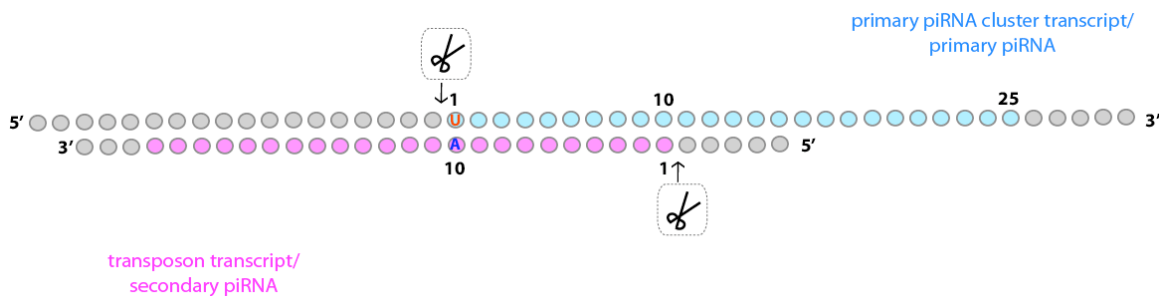


Figure 5. Signatures of the ping-pong cycle.

Primary piRNAs show a preference for uridine at position 1 ('U1 bias'). Secondary piRNAs are generated from target transcripts of the primary piRNA through a slicer activity of the Piwi proteins, which generates the 5' end of each new piRNA. Therefore, secondary piRNAs overlap with the primary piRNAs over 10 nucleotides, and predominantly have an adenine at position 10 ('A10 bias').

The mechanism that generates the 3' end is at present not known; the two putative nucleases, *zucchini* and *squash*, required for piRNA production and function, have been implicated in the process (Pane *et al. Dev. Cell* 2007). Populations of piRNAs bound to different Piwi proteins from pre-meiotic mouse testis (Aravin *et al. Science* 2007; Aravin *et al. Mol.Cell* 2008; Kuramochi-Miyagawa *et al. Genes Dev.* 2008), and zebrafish germ line (Houwing *et al. Cell* 2007) also display 'ping-pong' signatures. This makes the 'ping-pong' mechanisms one of the main ways of controlling transposon expression in the germ line of animals, guarding the fertility and fitness of the species.

Apart from acting in the 'ping-pong' cycle, Piwi-piRNA complexes were found to repress expression of transposons on the transcriptional level as well. In mouse, lack of Miwi2 and Mili causes demethylation of DNA at transposon loci, thereby

leading to their expression (addressed in further detail later in the text) (Carmell *et al. Dev. Cell* 2007; Kuramochi-Miyagawa *Genes Dev.* 2008; Aravin *et al. Mol. Cell* 2008). In *C. elegans*, piRNAs – termed 21U-RNAs – associate with Piwi protein PRG-1 and control expression of some transposon classes in male and female germ line. However, no evidence of ‘ping-pong’ cycle was found in *C. elegans*; the piRNA pathway seems to operate upstream of the endogenous siRNA pathway in the control of transposon expression (Das *et al. Mol. Cell* 2008).

In summary, Piwi proteins act in the germ line of animals, where they control the expression of mobile genetic elements on both transcriptional level - by inducing DNA methylation, and post-transcriptional level - by cleaving transposon transcripts.

The link between stem cell defects caused by the lack of Piwis and their mechanisms of action has not yet been fully explained. It remains to be seen if these are merely consequences of the damage caused in the genome upon transposon up-regulation, as one study suggests (Klattenhoff *et al. Dev. Cell* 2007), or if Piwis and their associated piRNAs have a separate function in controlling division and maintenance of germ line stem cells.

3. IV. b. Piwis in the mouse: Miwi, Mili, and Miwi2.

The mouse genome encodes three Piwi proteins, called the Miwis: Miwi, Miwi2 and Mili (Carmell *et al. Genes Dev.* 2002). Like other Piwis, the Miwis were reported to have germ line-restricted expression. Miwi is expressed only in testis, in spermatocytes and round spermatids (Kuramochi-Miyagawa *et al. Mech.Dev.* 2001; Deng and Lin, *Dev.Cell* 2002), and Mili is expressed throughout the development of the male germ line, from embryonic day 12.5 until the appearance of haploid cells (Kuramochi-Miyagawa *et al. Mech.Dev.* 2001). Absence of either Miwi1 or Mili results in the complete arrest of spermatogenesis, placing Miwis among master regulators of development of male germ cells. Expression of Mili has also been detected in the female germ line, being expressed throughout the embryonic development, peaking around the time of birth and decreasing in the adult ovary (Kuramochi-

Miyagawa *et al. Mech.Dev.* 2001). The absence of Mili, however, does not have any effect on the development of female germ cells. Both Miwi and Mili are expressed only in the germ cells and not in the somatic compartment of the gonads - unlike *Drosophila* Piwi, which is present in both somatic and germ cells in the fly gonads (Cox *et al. Development* 2002). Both proteins localize to the cytoplasm and are enriched in chromatoid bodies, RNA storage and processing centers reminiscent of somatic cells' P bodies (Grivna *et al. PNAS* 2006; Unhavaithaya *et al. J.Biol.Chem* 2009) – unlike *Drosophila* Piwi, which is a nuclear protein (Cox *et al. Development* 2002). IN contrast to Miwi, Mili is also present in the cytoplasm of spermatogonia, the germ line stem cells, in both embryonic and post-partum spermatogonia (Aravin *et al. Mol.Cel* 2008; Unhavaithaya *et al. J.Biol.Chem* 2009).

Miwi and Mili are present in complexes with so-called pachytene piRNAs, 24 - 30 nt in length, and are required for their biogenesis and/or stability (Girard *et al. Nature* 2006; Aravin *et al. Nature* 2006). Pachytene piRNAs appear approximately two weeks after birth, around the onset of the pachytene stage of meiosis, and are expressed throughout the adult life in spermatocytes and spermatids. They originate from large genomic clusters and show profound strand bias. These clusters are conserved between mouse, rat and human, despite the lack of the conservation of primary sequences of individual piRNAs (Girard *et al. Nature* 2006; Aravin *et al. Nature* 2006). Pachytene piRNAs are depleted of transposon-related sequences. The details of their biogenesis are not known; however, they are known to be Dicer-independent and are thought to arise by processing of a long single-stranded transcript. They show a strong U1 bias: a new piRNA starts with each uridine on the primary transcript. Pachytene piRNAs have no obvious targets and their exact function remains elusive.

Even before the discovery of piRNAs in 2006, Miwi and Mili were known to bind RNAs (Kuramochi-Miyagawa *et al. Mech.Dev.* 2001). Miwi was shown to bind the mRNAs of ACT (Activator of CREM in Testis) and CREM target genes (Deng and Lin, *Dev. Cell* 2002). CREM is considered to be the master regulator of spermiogenesis, as it regulates the transcription of spermiogenic factors such as protamines, transition proteins, and other germ line-specific genes abundantly expressed in early

spermatids (Fimia *et al. Mol. Cell. Endocrinology* 2001). Both ACT mRNA and the mRNAs of CREM target genes are downregulated in *miwi*^{-/-} testes. The fact that the expression of CREM itself does not depend on Miwi suggests that Miwi acts by stabilizing these mRNAs (Deng and Lin, *Dev. Cell* 2002). Furthermore, both Miwi (Grivna *et al. PNAS* 2006) and Mili (Unhavaithaya *et al. J.Biol.Chem* 2009) were found to be associated both with free ribonucleoproteins (RNPs), containing chromatoid body components, and polysomes in testicular extract, as well as with the cap-binding complex. This is consistent with their localization in both cytoplasm and chromatoid bodies. Association of Miwi with polysomes increased with the up-regulation of the overall translational activity in spermatids (Grivna *et al. PNAS* 2006), and the global translation rate is decreased in spermatogonia of *mili*^{-/-} mice (Unhavaithaya *et al. J.Biol.Chem* 2009). On the other hand, Miwi is also required for the integrity of chromatoid bodies (Kotaja *et al. J. Cell Sci.* 2006). These findings suggest a role for Miwi, Mili and their associated piRNAs in controlling translation – possibly both translational activation and repression – of mRNAs essential for successful completion of spermatogenesis and, in the case of Mili, differentiation of spermatogonia.

Consistent with its expression prior to the onset of meiosis in the post-birth testis, Mili is additionally associated with a separate population of piRNAs, termed pre-pachytene piRNAs (Aravin *et al. Science* 2007). This population of piRNAs originates from genomic clusters distinct from the ones giving rise to pachytene piRNAs and, in contrast to pachytene piRNAs, is enriched for repeats and transposon sequences. Transcripts of certain transposon classes are upregulated in *mili*^{-/-} male germ line, implicating Mili and its associated pre-pachytene piRNAs in transposon control. In *Drosophila*, Piwi proteins in complexes with piRNAs regulate transposon expression primarily on the post-transcriptional level, by catalyzing cleavage of transposon transcripts and simultaneous generation of new piRNAs through the ‘ping-pong’ cycle (Brennecke *et al. Cell* 2007). A substantial fraction of pre-pachytene piRNAs in the mouse was found to bear ‘ping-pong’ signatures, indicating that Miwis are also involved in the post-transcriptional silencing of transposons (Aravin *et al. Science* 2007).

A scheme of mouse spermatogenesis with some hallmark processes and the timing of expression of the Miwis and different piRNA populations is shown in Figure 6.

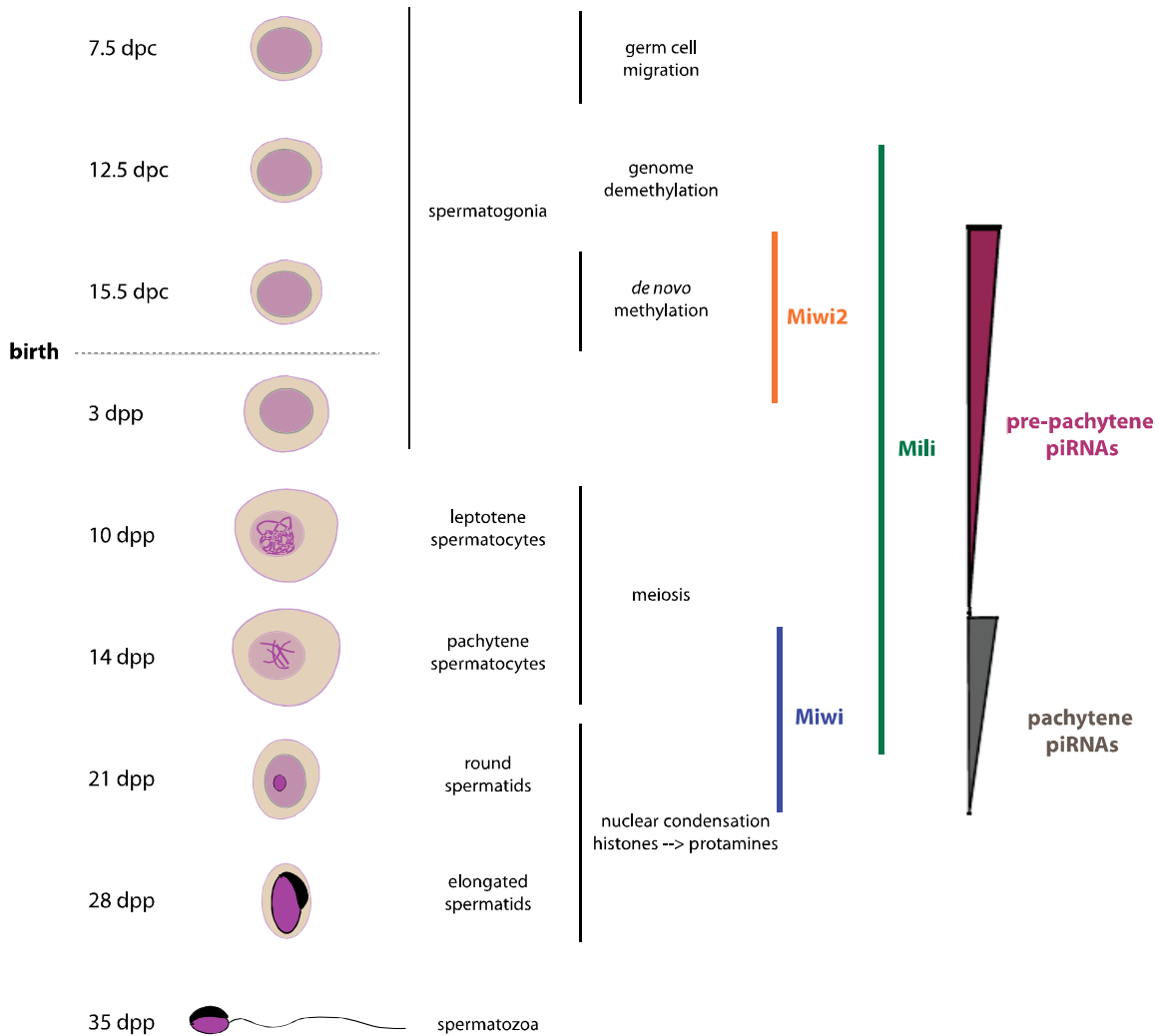


Figure 6. Schematic representation of mouse spermatogenesis.

Main cell types, approximate time of their emergence and hallmark processes are depicted, as well as time windows of expression of the Miwi proteins and their associated piRNA populations.

dpc; days *post coitum*, dpp; days *post partum*.

3. IV. c. Miwi2 and Mili in male fetal gonads.

Silencing of transposons takes place in mammals primarily through the establishment of CpG methylation at transposon regulatory regions in the embryo. Some transposon loci were found to be extensively demethylated in mutants of two Piwi subfamily members, accompanied with up-regulation of transposon transcripts: Mili and Miwi2 (Carmell *et al. Dev.Cell* 2007; Kuramochi-Miyagawa *et al. Genes Dev.* 2008; Aravin *et al. Mol. Cell* 2008).

Both Mili and Miwi2 are expressed in the developmental window when *de novo* methylation of transposon sequences takes place in the male germ line – between embryonic day 15.5 and 18.5 (Aravin *et al. Mol Cell.* 2008). Mili localizes to cytoplasmic granules reminiscent of *Drosophila* nuage. Miwi2 localizes predominantly to the nucleus and is also present in cytoplasmic granules, which partially co-localize with the Mili-containing ones (Aravin *et al. Mol. Cell* 2008). Immunoprecipitation of Mili and Miwi2 from fetal testis revealed that they are present in complexes with diverse population of piRNAs. Mili binds ~26nt piRNAs, around 55% of which correspond to transposons and repeats. Mili-bound piRNAs are predominantly sense to transposon transcripts. Miwi2-bound population of piRNAs is even more enriched for transposon transcripts, containing about 80% transposon- and repeats-derived transcripts. These piRNAs are predominantly antisense to transposons, and their ‘ping-pong’ partners are detected among Mili-bound piRNAs. Mili-associated piRNAs are the primary piRNAs of the mouse ‘ping-pong’ cycle. Some of them originate from genomic clusters (Kuramochi-Miyagawa *et al. Genes Dev* 2008; Aravin *et al. Mol Cell.* 2008), as is the case in *Drosophila* and neonate mouse, but mostly they’re derived directly from primary transcripts of individual transposons. Primary piRNAs display U1 bias and guide the cleavage of antisense transcripts, giving rise to Miwi2-associated secondary piRNAs.

This process likely takes place in the cytoplasmic granules, where partial co-localization of Miwi2 and Mili is observed (Aravin *et al. Mol Cell.* 2008). As mentioned, mutations in either *miwi2* or *mili* lead to reduction in DNA methylation at transposon loci (Kuramochi-Miyagawa *et al. Genes Dev* 2008; Aravin *et al. Mol Cell.*

2008). piRNAs are still present in the *dnmt3l* mutant, suggesting that Miwi2 and Mili likely act upstream of DNA-methyltransferases responsible for establishment of germ line methylation patterns. Furthermore, in the *mili* mutant, Miwi2 does not associate with secondary piRNAs nor does it translocate to the nucleus. These results suggest that Miwi2, associated with secondary piRNAs produced in cooperation with Mili-piRNA complexes through ping-pong mechanism in the cytoplasm, shuttles to the nucleus where it guides the establishment of DNA methylation on transposon sequences. The exact mechanism is elusive to date.

The schematic representation of the 'ping-pong' cycle in the mouse embryonic gonad is shown in Fig. 7.

The expression of Miwi2 decreases after birth, and only Mili seems to contribute to the 'ping-pong' cycle in the neonate mouse. Profiles of piRNAs change dynamically through development, with Mili-associated piRNAs changing accordingly: different types of transposons are targeted, strand bias changes, as does the ratio between primary and secondary piRNAs populating the pathway (Aravin *et al. Mol.Cell* 2008). These changes likely reflect the dynamics in expression of certain transposon classes and the preferred mechanism of transposon defense.

The phenotypic changes observed in the *mili* and *miwi2* mutant testis after birth of the mouse are quite similar, and resemble those observed in the *dnmt3l* mutant. Spermatogenesis in the *miwi2* and *mili* mutant mice is arrested in the meiotic prophase I, at early pachytene stage, leading to complete sterility of male mice. Spermatocytes show highly abnormal nuclear morphologies, and the level of apoptosis is elevated. Mutant mice also lose germ cells progressively with age and seem to have impaired germ cell differentiation. There are also differences between the two mutants: *mili* mutant mice can still form the sex body (Kuramochi-Miyagawa *et al. Development* 2004), a nuclear sub-domain containing transcriptionally silenced sex chromosomes, and this is not the case in the *miwi2* mutant testis (Carmell *et al. Dev. Cell* 2007).

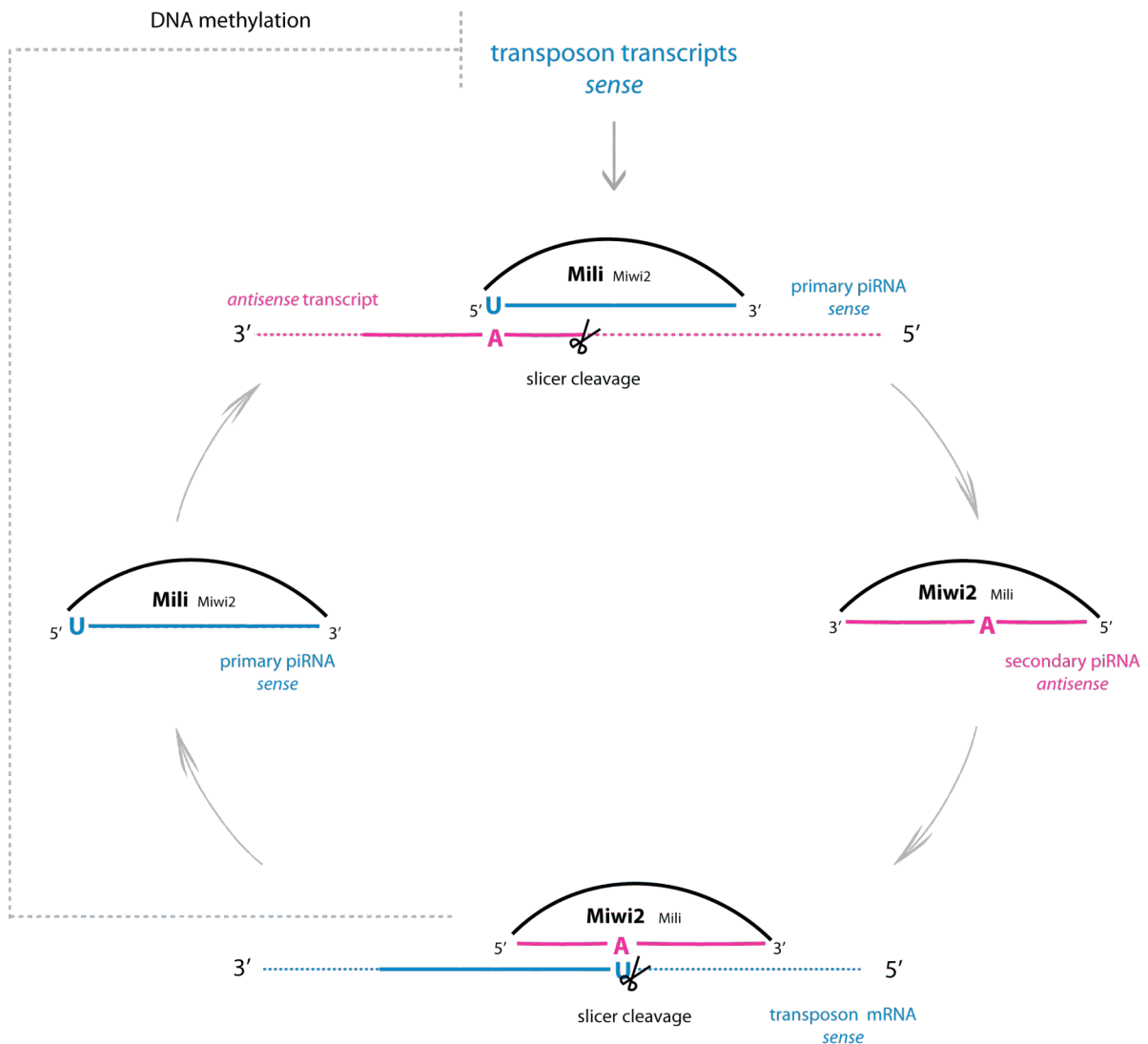


Figure 7. The ping-pong cycle in the mouse male embryonic gonad.

Individual transposons and some piRNA clusters are transcribed and processed by an unknown mechanism into primary piRNA, which are predominantly sense to transposons and assemble predominantly into complexes with one of the members of the Piwi subfamily in mouse, Mili. These piRISCs cleave the antisense transposon transcripts, thereby generating new piRNAs. Secondary piRNAs are antisense in respect to transposon mRNAs and assemble into piRISC predominantly with the second member of the Piwi subfamily, Miwi2. 'Antisense' piRISC cleaves again the primary transposon transcripts, leading to their silencing. In addition, Miwi2 in complex with secondary piRNA migrates into the nucleus, where it is involved in DNA methylation of the regulatory regions of transposons, providing thereby an additional level of silencing.

Unlike Mili, Miwi2 is expressed in both somatic and germ cells of the testis (Carmell *et al. Dev. Cell* 2007); it is the only Piwi family member in the mouse whose expression is not restricted to germ line cells.

However, the function of Miwi2 in survival of germ line cells seems to be germ-line autonomous, as wild type germ cells transplanted into a mutant recipients can completely reconstitute spermatogenesis.

In summary, Miwi2 associates with piRNAs derived from transposable elements in the embryonic male germ line and, together with Mili and its associated piRNAs, plays an essential role in transposon silencing through DNA methylation. It is required for continued production of germ line stem cells in the adult, progression of cells through prophase I of meiosis, and consequently male fertility. It is the only Miwi expressed both in germ line and somatic cells of testis.

3. V. DNA methylation in control of transposition and imprinting.

During development of a multicellular organism, cells differentiate from the pluripotent state in the early embryo into a wide variety of specialized cell types of adult tissues and organs. Development is epigenetic: as all cells in an organism have the same genetic content, it is the epigenetic information that orchestrates differentiation and functional specialization of cells by allowing expression of some genes while keeping others silent. Tissue-specific gene expression patterns are maintained by flexible histone modifications, with the aid of the proteins from the Polycomb and Tritorax families. The only sequences in the genome that are stably silenced on the long term by DNA methylation are pluripotency-associated genes, imprinted genes and transposable elements (reviewed by Reik *Nature* 2007).

Genomic imprinting is an epigenetic process that leads to specific expression of certain genes from only one allele, depending on its parental origin. Some 90 genes, usually clustered in the genome, are known to be imprinted in mammals, and loss of imprinting is associated with a number of cancers and developmental disorders. Expression of imprinted genes is regulated mainly by DNA methylation,

as well as by histone modifications, of the so-called imprinting control regions (ICRs), also known as differentially methylated regions (DMRs). ICRs can act either as insulators, or as promoters for the non-coding RNAs, which then regulate the expression of the neighboring genes. Examples of the non-coding RNAs involved in the regulation of imprinting include Xist, involved in somatic X inactivation, and Air, involved in regulation of *Igf2r* (reviewed by Ideraabdullah *et al. Mut. Res.* 2008; and Royo and Cavaillé *Biol. Cell* 2008).

Transposable elements, or transposons, have an important role in shaping the genomic landscape as well as controlling expression of their neighboring genes (reviewed by Slotkin and Martienssen, *Nature Rev. Genet.* 2007). However, their ability to move from one genomic location to another represents a potential threat, as excision of a transposon from one site in the genome and insertion into another causes chromosome breakage, illegitimate recombination, and disruption of coding or regulatory regions of genes, leading to misregulation of their expression. Activation of transposons poses a particular problem when it occurs in the germ line. It then influences not only the fitness of a particular single organism, but also of the species, as mutations and genomic rearrangements that occur in the germ line get transmitted to the next generation. In addition, transposable elements have been proposed to be involved in regulation of imprinted genes – some transposon classes are hypomethylated and expressed specifically in the male or female germ line, e.g. IAP and LINE retrotransposons are hypomethylated in the female germ line, their mRNA is highly abundant in the oocytes and is deposited in the embryo (Peaston *et al. Dev. Cell* 2004). The opposite is true for non-autonomous SINE elements, which are hypomethylated in the sperm (Rubin *et al. Nuc. Acid Res.* 1994). This led to the proposal that transposons may contribute to differential expression of maternally or paternally imprinted genes.

In mice, DNA methylation patterns are inherited through gametes and are maintained for one cell division in the zygote by a maternally inherited form of DNA methyltransferase 1, DNMT1o (Howell *et al. Cell* 2001). Methylation is first erased, and then re-established by *de-novo* DNA methyltransferases from the DNMT3 family (Okano *et al. Cell* 1999). Throughout development of somatic tissues,

DNA methylation patterns are maintained by somatic maintenance methyltransferase, DNMT1s (Li *et al. Nature* 1993). A second wave of de-methylation of DNA takes place during development of the germ line in the embryo, in progenitor germ cells (PGCs) in the period between embryonic days E10.5 and E12.5. Methylation patterns are then re-established depending on the gender of the embryo (Hajkova *et al. Mech. Dev.* 2001), in the period between E15.5 and E18.5.

Stable silencing through DNA methylation of transposable elements, imprinted genes and genes involved in pluripotency is particularly important, due to detrimental effects of transposon activity on integrity of the genome, importance of correct imprinting for development of the embryo after fertilization, and higher susceptibility to cancer resulting from expression of pluripotency-associated genes in differentiated tissues.

Proteins from the DNMT3 family, Dnmt3a, Dnmt3b, and Dnmt3L have known roles in establishment of DNA methylation of transposable elements and imprinted genes (Kaneda *et al. Nature* 2004; Hata *et al. Development* 2002; Bourc'his and Bestor, *Nature* 2004; Kato *et al. Hum. Mol. Genet.* 2007). As described above, Piwi family members Miwi2 and Mili are expressed at the high levels in the period of *de novo* DNA methylation, and together with their associated piRNAs have a role in establishing DNA methylation of transposable elements in the male embryonic germ line. Mutations in Dnmt3 family members result in loss of maternal imprinting and consequent death of progeny early in development, as well as loss of transposon silencing in the male germ line resulting in male sterility. Similarly, as described above, mutations in either Miwi2 or Mili lead to expression of transposons in the male germ line and male sterility.

The function of Miwis in the regulation of imprinting has so far not been addressed.

4. Materials and Methods.

RNA isolation.

Total RNA was isolated from ~100 µg of mouse tissue, or one 15 cm tissue culture dish in case of cultured cells, using the standard phenol/chloroform extraction method. The tissue was homogenized, and the cells harvested, in 3 ml of extraction solution [1 volume Solution A (4M guanidium isothiocyanate, 25 mM sodium citrate, 0.5% lauroylsarcosinate, 50 mM β-mercaptoethanol) and 1 volume acidic phenol, Ph 4.0, (AppliChem)]. Chloroform:isoamyl alcohol 24:1 (AppliChem) and 2M sodium acetate, pH 4.2, were added to the homogenate (1/10 and 1/20 of total volume, respectively), which was then vigorously mixed. After centrifuging at 10000 rpm for 10 min at 4°C, aqueous phase was separated from organic phase. RNA was further extracted by adding ½ of total volume of phenol:chloroform:isoamyl alcohol 25:24:1 (PCI; AppliChem) to the aqueous phase, mixing and centrifuging at 10000 rpm for 1 min at 4°C. Aqueous phase was ultimately mixed with 3 volumes of cold absolute ethanol (Merck), and precipitated over night at -20°C. After centrifugation at 10000 rpm for 30 min at 4°C, RNA pellet was resuspended in 1 ml Solution A, mixed with 1 volume of isopropanol (Fluka) and incubated 15 min at room temperature. After one more round of centrifugation, RNA pellet was washed twice in 70% ethanol, dried and resuspended in 100 µl of RNase-free water. Quality and concentration of isolated RNA were determined using the Agilent 2100 BioAnalyzer.

Reverse transcription.

Starting from 5 µg of total RNA, cDNA was synthesized using either First Strand cDNA Synthesis Kit (GE Healthcare, 27-9261-01) or SuperScript First Strand Synthesis System (Invitrogen, 12371-019).

As a primer for reverse transcription, either oligo d(T) or the gene-specific primers, listed in Table 1., were used.

Table 1. Gene- specific primers used for reverse transcription.

PRIMER NAME	SEQUENCE
3'UTR_1	CCTGAGCGGTAGGCAGATCTGAACAACG
3'UTR_2	GAGTCTCAAGCCGTCCTCTGGCTCACTTG
e21r	GGTTCCTTGTGGACACTTTGTGCCACC
e15r	CCACAGCTCTCCGCCAAGTTTGC
e12r	CCATCTATTCAAAGGCTGAG

Polymerase chain reaction (PCR).

cDNA was amplified using Taq polymerase (0.5 µl per 50 µl reaction; Eppendorf/5 PRIME, 2900167), with 1 µl of cDNA as template. Primers used for PCR are listed in Table 2.

Rapid Amplification of cDNA Ends (RACE).

Total RNA (5 µg) from various mouse tissues (testis, different age; adult ovary, spleen, kidney), embryonic fibroblasts or embryonic stem cells was analyzed by 5' RACE, using Invitrogen 5' Rapid Amplification of cDNA Ends System (cat# 18374-041). First strand cDNA synthesis was performed using e21r gene-specific primer (GSP1). Following purification, cDNA was tailed and amplified by nested PCR. In the first round of PCR, AAP primer matching the tail was used in combination with e20r primer. The primers were annealed at 55°C for 30 sec and 35 cycles of amplification were carried out using Taq DNA polymerase (Eppendorf) at 68°C for 2:30 min. The PCR product was purified using Qiagen PCR Purification Kit; the DNA was eluted in 30 µl of water and 2 µl were used as template for second round of PCR. In the second round, primers AUAP and nested primer e15r were annealed at 55°C for 30 sec. Amplification conditions were the same; the number of cycles was 30. The products were finally gel-purified and analyzed by direct sequencing using primers e15r, and e12r and e3r were applicable.

Table 2. PCR primers.

PRIMER NAME	SEQUENCE
e1alt	ATGCGCCTGCGCATTCTGGGGGTCC
e1f	GGGCGGTCCCTCAAGAGACCTCATGG
e2f	GGTTACATCTGCCAGTCCTGGGGACAG
1320	GCAGCCAAGCAACCTCAGATTTCC
1590	GCGATCTTGCAAAGTTTTGAGTTCTCAGCC
1711	GGTTCCATGGGATTTAACGTGGGCTACC
e15f	GGAACGATGCTGAGTGTG
β tubulin_F	GCTCTCTACGACATTTGCTTCAGAACCC
Miwi1_f	ATGACTGGCCGAGCCCCGAGCTCG
Mili	ATGGATCCTGTCAGGCCGTTG
e2r	CCCACCTCCGGCCTCGCTGTCC
e3r	CTGGTACAGACACCCAACTCT
e8r	CCACATATGTGACCTCTGAGCCGTCC
e12r	CCATCTATTCAAAGGCTGAG
e15r	CCACAGCTCTCCGCCAAGTTTGC
e16r	GCAGACTTTGAGGCAATCTGC
e20r	GCCCTGCCAGTTGTAGTAG
e21r	GGTTCCTTGTGGACACTTTGTGCCACC
β -tubulin_R	CAGACAGCCGTTTTTCACATTGTTGG
Miwi1_r	CCATCGAAAGCATGACACCTTCC
Mili_r	CCATTCCTCTGCCTATACTACTGG

Quantitative PCR (qPCR).

qPCR was performed using Fermentas Maxima SYBR Green 2x Master Mix (cat# K0222) according to the manufacturer's instructions, on the BIO-RAD MyiQ Single Color Real-Time PCR Detection System and using the accompanying iQ5 software.

Total reaction volume was 25µl, and 7.5 pmol of each primer was used. 0.8 µl of cDNA reverse transcribed using oligo (dT) primer from 5 µg of total RNA was used as a template. Primers used for qPCR are listed in Table 3.

Table 3. Primers used for qPCR.

PRIMER NAME	SEQUENCE
βtubulin_f	TCAATGCTGACCTGCGGAAG
βtubulin_r	TCAAACATCTGCTGGGTGAGC
H19_f1	AATCAGAACCACTACACTAC
H19_r1	ATTTATTGATGGACCCAGGAC
H19_f2	CATCCAGCCTTCTTGAACACC
H19_r2	GGGAAAAGTGAAAGAACAGACGG
Igf2_f1	GGCTTCTACTTCAGCAGGCCT
Igf2_r1	ACTCTTCCACGATGCCACG
Igf2_f2	GAACCACATTCCCCAAAACAAGG
Igf2_r2	GAGCCAATAAGCAGCAGCAGAAAC
Miwi2_q5'f	GTTACATCTGCCAGTCCTGG
Miwi2_e3r	CTGGTACAGACACCCAACTCT
e15f	GGAACGATGCTGAGTGTG
Miwi2_q3'r	CCACATTCTTGTTGAGGGCATCTC
Oct4_f	TCACTCACATCGCCAATCAGC
Oct4_r	CCCTGTAGCCTCATACTCTTCTCG

Antibodies.

Synthetic peptides GITTGHSAREVGRSSRD (N-terminal) and TDHTVQNPPLGTVVDSEATRPEW (C-terminal) were injected into rabbits. The fifth bleed, after the heart puncture, was used. The antibodies were additionally affinity-purified using the two peptides conjugated to the column resin, and eluted from

the column using 1,5M $MgCl_2$ + 50mM NaAc pH 5.2, and 0,1M Glycine pH 2,45 + 0,1M NaCl solutions. The two fractions were dialyzed against 1.5 l of PBS at 4°C over night with constant agitation, and concentrated using Milipore Centricon (Plus-20, cat# UFC2LTK24) system. Additionally, a commercially available C-terminal antibody was used (Abcam, ab21869). The N-terminal antibody was used at 1:2000 dilution for Western blotting, 1:5000 for immunofluorescence on frozen tissues and cells, and 1:500 for immunohistochemistry on FF-PE tissue. The C-terminal antibody was used at 1:6000 for Western blotting, and 1:5000 for immunofluorescence on cells. The Abcam Miwi2 antibody was used at 1:500 for Western blotting and 1:100 for immunofluorescence on cells. BRCA1 antibody was from Santa Cruz (cat# sc-1553, M-20). It was used at 1:500 for IF on frozen tissue sections. γ H2AX antibody was from Upstate (cat# 05-636). It was used at 1:5000 for IF on frozen tissue sections. Actin antibody was from Sigma (cat# A-2066). It was used at 1:1000 for Western blot. Secondary antibodies were from Molecular Probes/Invitrogen. HRP-conjugated secondary antibodies were used for Western blot at 1:5000. AlexaFluor-conjugated secondary antibodies were used at 1:2000 for immunofluorescence on frozen tissue sections and cultured cells.

Polysome gradient separation of cytoplasmic tissue extracts and subsequent RNA isolation.

Tissues were dissected from the mouse and grinded in a lysis buffer (10 mM Tris-HCl, pH 8; 150 mM NaCl; 5 mM $MgCl_2$; 1% NP40; 40 mM DTT; 50 U/ml RNAsin (Promega, cat# N2511); 40 mM vanadyl-ribonucleoside complex (New England Biolabs, cat# S1402S), supplemented with 1% deoxycholate. 1 ml of lysis buffer was added per approximately 10 mg of tissue. Nuclei were pelleted by centrifuging at 13000 rpm for 15 seconds at 4°C, and supernatant was transferred to a new tube containing 500 μ l of 2x extraction buffer (0.2 M Tris-HCl, pH 7.5, 0.3 M NaCl, supplemented with 150 μ g/ml of cycloheximide, 650 μ g/ml of heparin, and 0.1 mM AEBSF). After centrifuging for 5 min at 13000 rpm, the supernatant was loaded onto a 10 ml 15%-40% [w/v] sucrose gradient, supplemented with 10 mM Tris-HCl, pH 7.5,

140 mM NaCl, 1.5 mM NaCl, 10 mM DTT, 100 µg/ml of cycloheximide, and 0.5 mg/ml of heparin. The gradient was spun at 38000xg for 2 hours at 4°C with brake off. 600 µl fractions were collected and digested with 100 µg of proteinase K in 1% SDS and 10 mM EDTA for 30 minutes at 37°C. RNA was extracted with equal volume of phenol:chlorophorm:isoamyl alcohol 25:24:1, followed by precipitation in 100% ethanol. RNA was additionally precipitated in 2M LiCl at 4°C over night, pelleted by centrifugation at 13000 rpm for 20 minutes, washed in 70% ethanol, and resuspended in RNase-free water. The quality of RNA was measured using Agilent 2100 BioAnalyzer. Fractions 1-2, 3-4, 5-10, and 10-20 were pooled and reverse transcribed, followed by PCR.

Testicular single-cell suspension.

Testes were dissected from the mice, and tunica albuginea was removed. The tubules were mechanically macerated and digested in 100 U/ml of collagenase type I (Gibco, cat# 17100-017) and 50 U/ml DNase I (Roche, cat# 04536282001) in HBSS (Gibco, cat# 14025-092), supplemented with 20 mM Hepes, pH 7.2; 6.6 mM Na-pyruvate; 0.05% w/v lactate; 1.2 mM MgSO₄ x 7H₂O; 1.3 mM CaCl₂ x 2H₂O, for 25 min at 32°C. Tubules were pelleted by centrifuging once for 2 min at 1000 rpm, washed once in HBSS, and incubated in Cell Dissociation Buffer (enzyme-free, Hanks'-based; Gibco, cat# 13150-016) for 25 min at 32°C. Cells were flushed out by pipetting 10 times and filtered through 70 µm and 40 µm nylon meshes (BD Falcon, cat# 352350 and , respectively). The cell suspension in the flowthrough was centrifuged for 2 min at 1000 rpm. The pelleted cells were washed once in HBSS and resuspended in the incubation buffer (HBSS, supplemented with 20 mM Hepes, pH 7.2; 6.6 mM Na-pyruvate; 0.05% w/v lactate; 1.2 mM MgSO₄ x 7H₂O; 1.3 mM CaCl₂ x 2H₂O; 3mM glutamine; 1%FCS). After 10 min incubation at 32°C, the cells were pelleted by centrifugation.

Preparation of extracts from testicular single-cell suspension.

The cells isolated from the testis were resuspended in 1 volume of Buffer A (10 mM Hepes, pH 8; 1.5 mM MgCl₂; 10 mM KCl; 1 mM DTT) and incubated on ice for 15 min. A 1 ml syringe was washed once in Buffer A, and the cells were resuspended by sucking them through the needle and flushing out five times. After centrifuging for 20 sec at 13000 rpm, the supernatant was kept as the cytoplasmic extract, and the pelleted nuclei were resuspended in 2/3 volume of Buffer C (20 mM Hepes, pH 8; 1.5 mM MgCl₂; 25% glycerol; 420 mM NaCl; 0.1 mM AEBSF; 1 mM DTT). After adding 50 U/ml DNase I and stirring on ice for 30 min, the nuclear extract was sonicated twice using a microtip sonicator at 30% power and 30% output. The extract was centrifuged for 5 min at 12000 rpm and the supernatant was recovered as nuclear extract.

Preparation of protein extracts from MEFs and organs.

Cells from a confluent Ø 15 cm plate were harvested in 500 µl of lysis buffer (50 mM Hepes, pH 7.3; 100 mM KCl; 5 mM MgCl₂; 10% glycerol; 0.1 mM AEBSF; 0.5 mM DTT), incubated on ice for 30 min and sonicated twice using a microtip sonicator at 30% power and 30% output.

Organs were dissected from mice, grinded in lysis buffer (~1 ml of buffer/ 10 mg tissue), incubated on ice for 30 min and sonicated twice using a microtip sonicator at 30% power and 30% output.

The total protein concentration was determined using Bradford protein assay (BioRad, cat# 500-0006).

Western blotting.

Fifty µg of total protein was loaded onto a 10% SDS-PAGE (ProtoGel, National Diagnostics, EC-890) and separated by running for 1 hour at 5 mA/cm. The proteins were transferred onto a nitrocellulose membrane (Whatmann, 0.2 µm, cat# 10401396) using the BioRad semidry blotting system for 2 hours in a buffer containing 39 mM glycine; 48 mM Tris; 0.037 % SDS, w/v; and freshly added 20%

methanol. The membrane was blocked in 5% milk, 1% FCS in 1x PBS over night at 4°C, or for 1 hour at room temperature. Primary antibodies were diluted in 3% milk in 1x PBS and incubated for 1 hour at room temperature. The membranes were washed twice in 1xPBS, once in PBS containing 0.1% Tween-20 (Fluka, cat# 93773). The HRP-conjugated secondary antibodies were diluted 1:5000 in 3% milk in 1x PBS and incubated for 1 hour at room temperature. The ECL detection system (Amersham/General Electric, cat# RPN2106V1/V2) was used for detection.

siRNA transfections.

MEFs were split 1:20 from a confluent plate, grown for 24 hours, and transfected with 100 nM or 150 nM siRNA duplexes using Dharmafect Transfection Reagent 1 (#T-2001-02) according to the manufacturer's instructions. After three day in culture, the cells were split 1:10, re-transfected after 24 hours, and harvested four days later. siRNAs against PML were used as a control and the knockdown efficiency was estimated to be ~35%. The sequences of the siRNAs are listed in Table 4.

Table 4. siRNAs.

Miwi2 siRNA #1	UUAUGAUUUUGGGGUAGCC
Miwi2 siRNA #2	AUUAAGAGAUUUGUAACC
Miwi2 siRNA #3	AAGAUUAUUGGCUAAUUC
PML siRNA #1	UUUUUUUCAGGGACGGUGC
PML siRNA #2	UUUCUGGUUCUGGGGCUC
PML siRNA #3	AUUGUUGUUGGGCAGGAGG

Immunofluorescence on frozen tissues.

Organs were isolated from mice, immediately frozen on dry ice in the embedding medium, and stored at -80°C for at least one day. Six-micrometer sections were cut, fixed in cold acetone at -20°C, and stored at -80°C.

For staining, sections were melted on room temperature for 10 minutes, blocked in 10% FCS/1x PBS for 30 minutes, and washed in 0.05% Tween-20/1x PBS three times for 15 minutes. Primary antibodies were diluted in 0.05% Tween-20/1x PBS and incubated over night at 4°C. The sections were then washed at room temperature in 0.05% Tween-20/1x PBS three times for 15 minutes. AlexaFluor-conjugated secondary antibodies were diluted 1:2000 in 0.05% Tween-20/1x PBS and incubated in dark for one hour at room temperature. Slides were mounted in ProLong-GOLD mounting medium with DAPI (Molecular Probes/Invitrogen, cat# P36931).

Immunocytochemistry.

The cells were grown on poly-L-lysine coated slides (Sigma, cat# P8920; coated according to the manufacturer's instructions). After removing the medium and washing once in PBS, the cells were fixed in 2% formaldehyde/1x PBS for 5 min at room temperature, washed twice for 5 min in PBS, and permeabilized in 0.1% TritonX-100 (Fluka, cat# 93420) for 10 min at room temperature. After washing once in PBS for 5 min, the cells were blocked in 10% FCS in PBS for 20 minutes. Primary antibodies were diluted in 0.1% saponin/10% FCS in PBS, and incubated over night at 4°C, and the cells were subsequently washed three times in 10% FCS in PBS. AlexaFluor-conjugated secondary antibodies were diluted in 0.1% saponin/10% FCS in PBS and incubated for one hour at room temperature. After washing three times in 10% FCS in PBS and once in PBS, the cells were mounted in ProLong GOLD.

Fluorescent *in situ* hybridization (FISH).

The frozen tissue sections were fixed in 4% paraformaldehyde and washed twice in 2x SSC for 5 min. After permeabilization in 1M NaSCN for 5min at 70°C, the slides were washed again twice in 2x SSC for 10 min, dehydrated through a series of ethanol dilutions (5 min in 30% EtOH; 5 min in 50% EtOH; 5 min in 70% EtOH; 5 min in 96% EtOH; 2x 5 min in abs. EtOH) and air-dried for 10 min. The probe (Cambio, ready-to-use Chromosome X Cy-3-labeled probe, cat# 1200-XMCy3-01) was equilibrated to room temperature, boiled for 5 min at 70°C, immediately transferred

to ice and chilled for 10 min. Five microliters were put on the tissue section, covered, and denatured on-slide for 10 min at 70°C in the water bath. The probe was then incubated over night at 37°C in a humid chamber, and subsequently washed 10 min 4x SSC with 50% formamide; 2x 10 min 4x SSC; 2x 10 min 2x SSC; 5 min 1x PBS, and mounted in ProLong GOLD.

Immunohistochemistry (IHC).

Tissues were dissected from mice, fixed in 4% paraformaldehyde with agitation over night in dark, dehydrated in ethanol series and embedded in paraffin. Slides were de-paraffinized in xylene, passed through the ethanol series, and stained using the automated *Ventana* system.

5. Results and Discussion.

Piwi proteins have been reported to be expressed and have functions exclusively in the germ line of animals (Malone and Hannon, *Cell* 2009, and references therein). In humans, however, both Hiwi1/hPiwil1 and Hiwi2/hPiwil4 can be detected on the RNA level outside of the germ line (Sasaki *et al. Genomics* 2003, Sharma *et al. Blood* 2001). Hiwi1 mRNA was detected in a number of fetal and adult tissues, and transient expression of Hiwi1 in human leukemia cells resulted in reduced proliferation rate of these cells (Sharma *et al. Blood* 2001). Miwi, the mouse homolog of Hiwi1, could also be detected on the RNA level by Northern blot outside of the germ line, in adult spleen and thymus (Kuramochi-Miyagawa *et al. Mech Dev.* 2001). These results raise a question if Piwis could have functions distinct from the ones in the germ line.

I used mouse as a model organism to study the function of the Piwi proteins. Since both expression and function of Miwi2, the closest mouse homolog of hPiwil4 (Carmell *et al. Genes Dev.* 2002), were unknown, I conducted a comprehensive study of Miwi2 expression both on RNA and protein levels in various adult mouse tissues and mouse embryonic fibroblasts (MEFs).

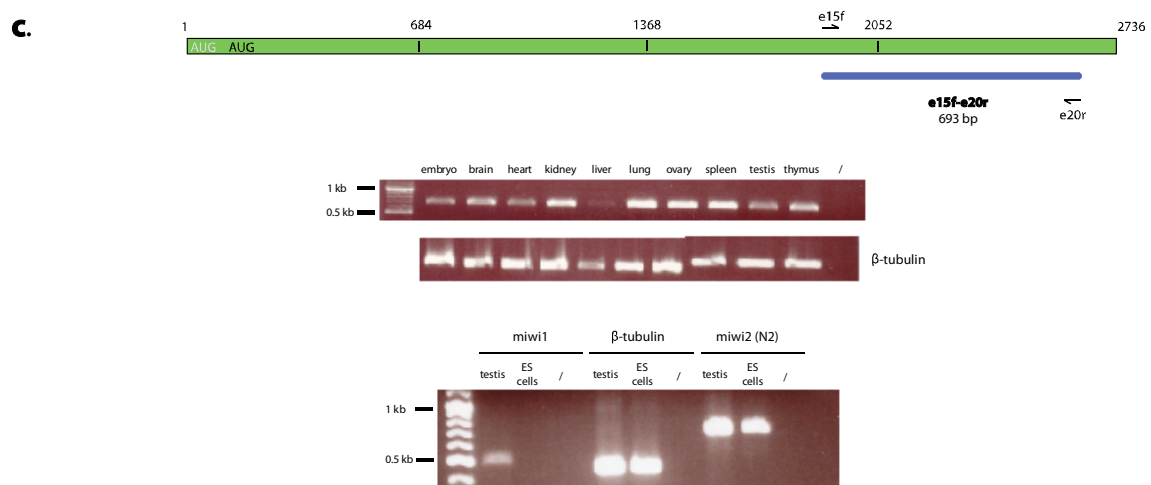
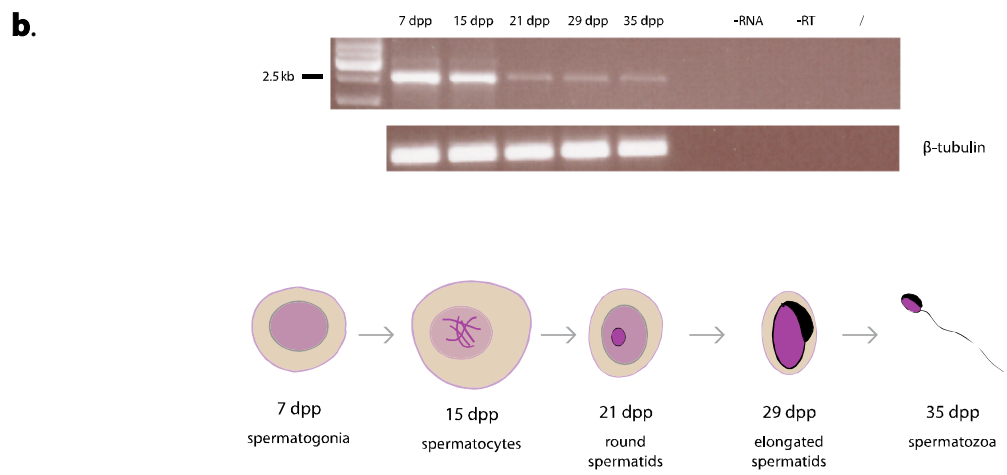
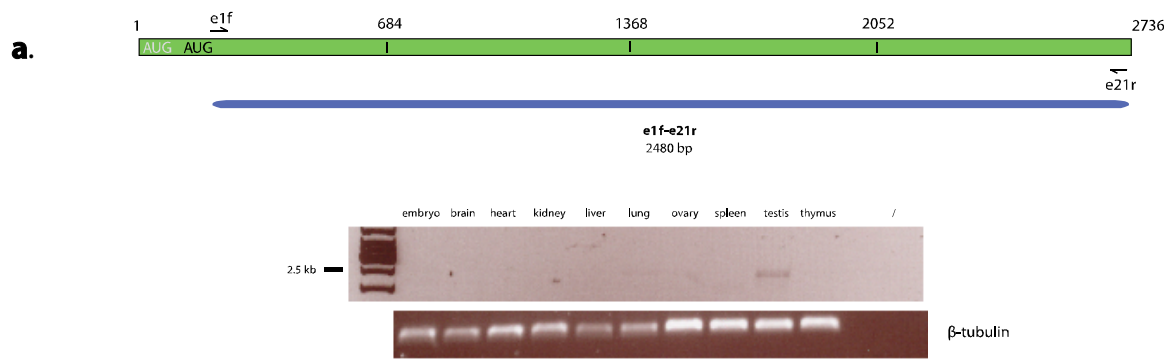
The main characterized function of Miwi2 is the control of expression of the transposable elements in the male germ line. Miwi2 and its associated piRNAs seem to play a role in the initiation of DNA methylation of the regulatory regions of transposons, resulting in their silencing (Aravin *et al. Mol. Cell* 2008; Kuramochi-Miyagawa *et al. Genes Dev.* 2008). The only sequences in the mammalian genome stably silenced through DNA methylation, apart from transposons, are imprinted genes and pluripotency-associated genes (Reik, *Nature* 2007). I therefore investigated if Miwi2 is involved in regulation of imprinted genes in both germ line and somatic tissues.

5. I. Analysis of miwi2 mRNA expression.

5. I. 1. The ubiquitous expression of miwi2 mRNA.

First, I analyzed the expression of miwi2 in the full mouse embryo and various adult tissues. Reverse transcription (RT) using oligo (dT) primer followed by the polymerase chain reaction (PCR) spanning the entire coding sequence of miwi2 confirmed that the full-length transcript is present only in testis (Fig. 7a.) This transcript can be detected throughout development of male germ cells, with decreased expression in the post-meiotic cells (Fig. 7b). However, PCR analysis using the primers covering a region close to the 3' end of miwi2 mRNA showed the expression of the transcript in the full embryo, embryonic stem cells, and majority of adult mouse tissues (Fig. 7c).

In order to analyze the expression of miwi2 mRNA in more detail, I designed different primer sets covering various regions of miwi2 mRNA (Fig. 7d). Reverse transcription of the total RNA isolated from adult mouse tissues or mouse embryonic fibroblasts using either random hexamer primers (RH) or a gene-specific primer (e21r), followed by PCR amplification of the indicated regions, confirmed that all parts of the full-length miwi2 mRNA could be detected only in testis (Fig. 7e). The product corresponding to the 5' part of the transcript (E1f-e8r) could only be detected in testis and, at the much lower levels, in ovary. Somatic tissues seemed to express only the transcripts from the 3' region of the gene (Fig. 7e). In the ovary, the e4f-e12r PCR product, spanning the 5' and 3' regions of the transcript, was not detected, despite the fact that parts of the transcript containing the binding sites for primers e4f and e12r (E1f-e8r product and 1320-e15r product, respectively) could be readily detected. In thymus and kidney, like in spleen, only the products corresponding to the 3' region were detectable (data not shown). Sequencing of the PCR products confirmed that they correspond to correctly spliced miwi2 mRNA. To confirm the different levels of expression of 5' and 3' regions of the miwi2 mRNA, as I will refer to them throughout the Thesis, I performed quantitative PCR (qPCR) (Fig. 8).



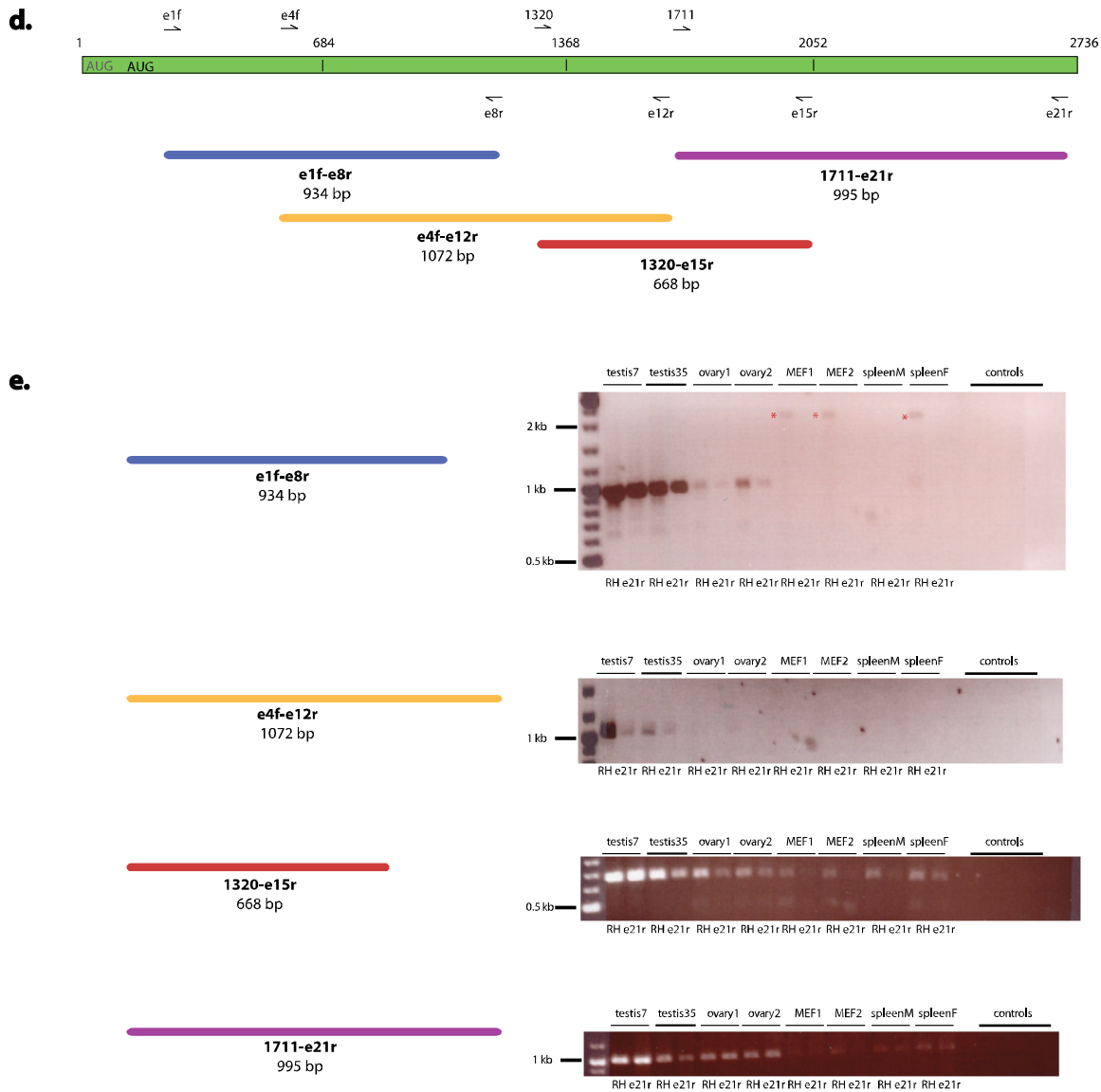


Figure 7. RT-PCR analysis of miwi2 expression in mouse tissues and cell lines.

- a.** RT-PCR analysis of expression of the full-length miwi2 mRNA, in adult mouse tissues and the full embryo. The entire transcript can only be detected in the testis. The mRNA is depicted by the green rod; this depiction is used throughout the thesis. The negative control where the cDNA template was not added is always represented by '/'.
- b.** Full-length miwi2 transcript was expressed throughout development of male germ cells. Its amount decreased after meiosis.

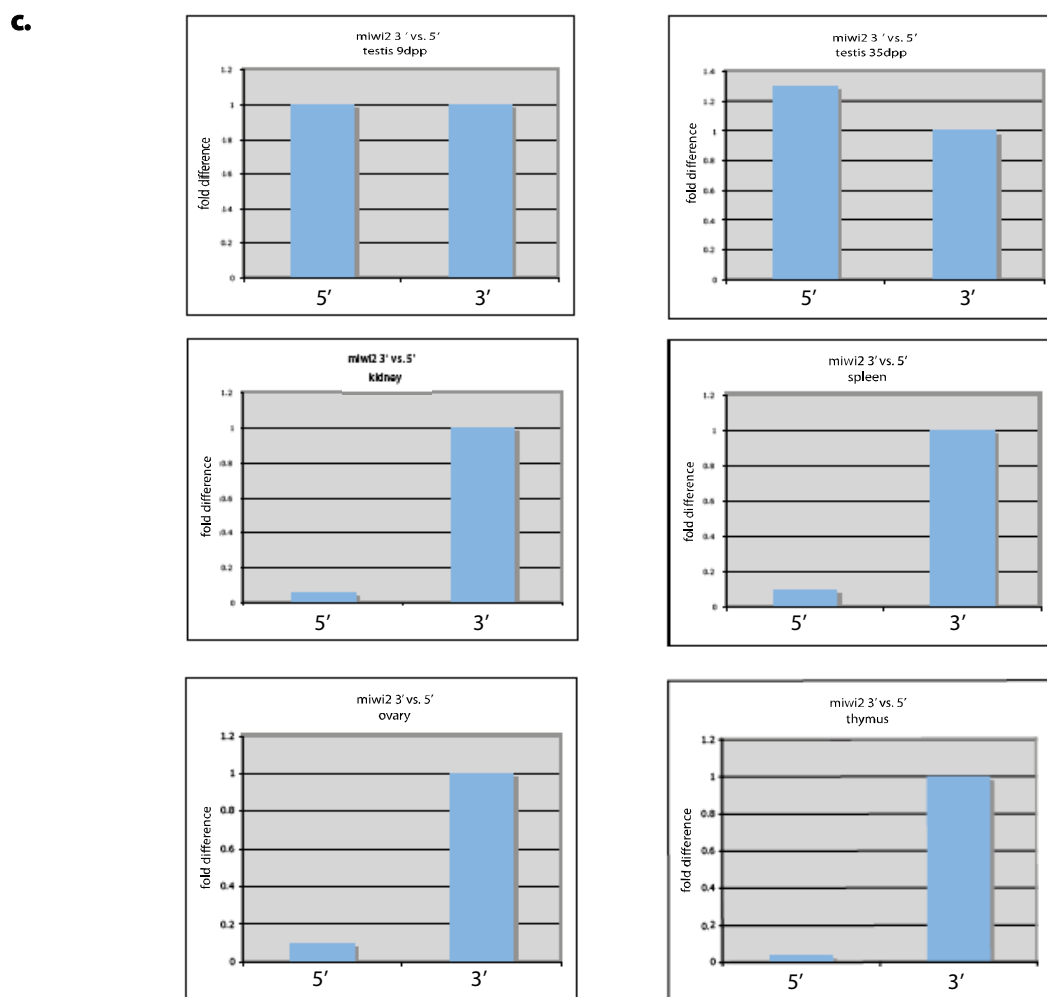
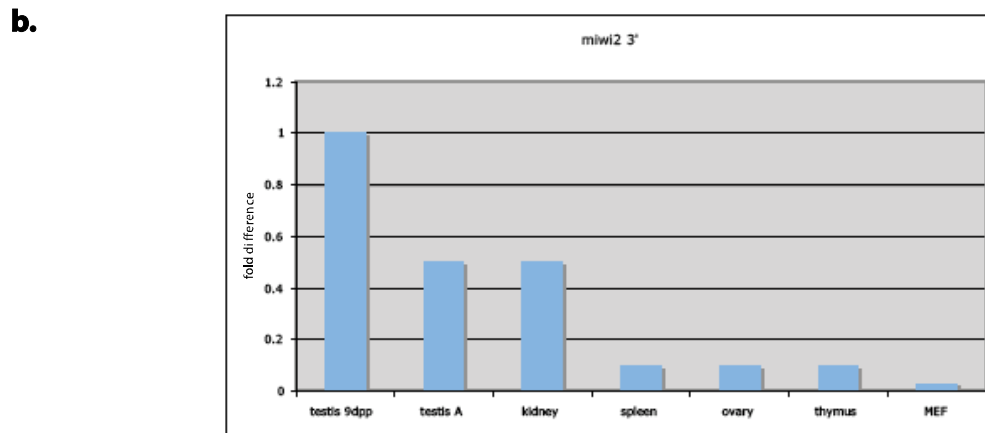
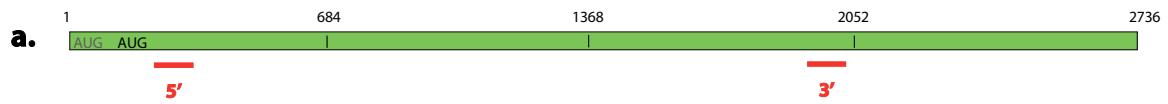
Lower panel: Schematic depiction of developmental stages of male germ cells, with indication of time after birth (days *post partum*; dpp) when the certain population of cells first appears. All stages are present throughout male adulthood.

- c. RT-PCR analysis of the indicated 3' fragment of the miwi2 transcript. **Middle panel:** the 3' region is expressed in the full embryo and majority of adult tissues. **Lower panel:** unlike miwi1, the 3' region of miwi2 can be detected in the ES cells. Controls: -RNA; -RT; -cDNA.
- d. Schematic representation of the full-length miwi2 mRNA. Shown are the positions of primers used for PCR, as well as position and length of the expected PCR products.
- e. RT-PCR analysis of expression of different regions of the miwi2 mRNA in adult testis, ovary and spleen, as well as in MEFs. Asterisks mark the unspecific bands. Testis 7: RNA isolated from 7 days-old mice; testis35: RNA isolated from 35 days-old mice; ovary1/2: two samples of RNA isolated from adult ovaries, 100 µg; MEF1/2: two samples of RNA isolated from 15 cm dish; spleenM/F: RNA isolated from adult spleen, 100 µg; M=male F=female.

Using two different primer pairs, one on the 5' end and the other roughly in the middle of the 3' region. The positions of their products are depicted in Figure 8a. First, I compared the levels of the ubiquitously expressed 3' transcripts in different tissues, using testicular tissue of young mice (9 dpp) as a reference. As observed in RT-PCR experiments, the levels of the transcript in the testis decreased with the age of mice (Fig. 7b and 8b). Also, the levels in the adult testis were higher than in the somatic tissues, with the exception of adult kidney. The levels of the transcript were approximately the same in adult spleen, ovary and thymus, and were ~3.5-fold lower than in the adult testis and kidney. A heterogeneous population of MEFs expressed the 3' transcript at a very low level, ~10-fold lower than adult testis (Fig. 8b).

Next, I compared the levels of 5' transcripts to the levels of the 3' transcripts separately for each organ and MEFs. In testis, 5' and 3' products were present at equal or slightly different levels throughout development, consistent with the presence of the full-length transcript. In somatic tissues and MEFs, the 5' transcript was virtually undetectable (Fig. 8c). The levels of the 5' product were low even in the ovary, where it could be detected by RT-PCR (Fig. 7e).

A Northern blot using the probes against the 5' and the 3' half of the miwi2 mRNA failed to reveal the sizes of the expected transcripts and therefore was not useful to evaluate the observed differences in expression of different regions of miwi2 (data not shown).



C.

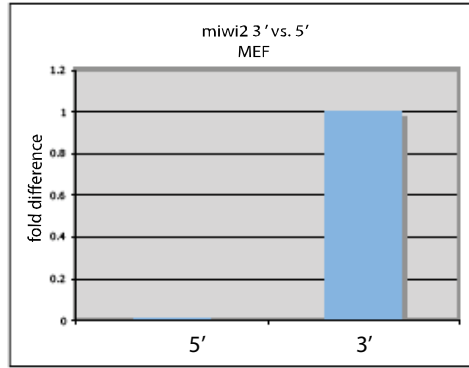


Figure 8. Quantitative PCR analysis of the miwi2 mRNA expression. The levels of the transcript are highest in the testis. 3' region is expressed at higher levels in all tissues, as well as in MEFs, compared to the 5' half, which is virtually absent in all tissues with the exception of testis.

- a. Positions of the qPCR products corresponding to the two halves of the miwi2 mRNA.
- b. Expression levels of the 3' half of miwi2 mRNA in various adult tissues and MEFs. The expression is highest in testis. The values were normalized to actin mRNA. The expression of miwi2 in 9dpp testis was taken as a reference. Testis A: adult testis sample. All somatic organs' sample are from adult mice.
- c. Comparison of difference in expression levels between 5' and 3' half of miwi2 in various adult tissues and MEFs. The levels in various tissues were compared separately, using the expression level of the 3' half as a reference, and normalized to actin.

In summary, these results show that, contrary to the previous reports, the expression of miwi2 seems to not be strictly limited to the male germ line. Even though the full-length, wild type transcript could be detected only in testis, transcripts corresponding to the 3' region of the gene seem to be ubiquitously expressed. In the ovary, the full-length transcript was not present, but PCR products from both 5' and 3' parts could be detected. The two products were present at drastically different levels, with the 3' product being ~10-fold more abundant than the 5' product, allowing for the possibility that they correspond to two distinct transcripts. In the majority of somatic tissues, only the products corresponding to the 3' part of the full-length transcript could be detected, with the products from the 5' region being virtually absent.

Is it possible that two promoters exist within the *miwi2* locus? The promoter at the 5' end of the gene seems to be active only in the germ line tissues. In testis,

transcription from this promoter gives rise to the full-length transcript. In ovary, it seems to have very low activity. As revealed by sequencing of the PCR products, this transcript is correctly spliced. The position of the 3' end of this transcript is at present not known; it could be determined by performing 3' Rapid Amplification of cDNA Ends (RACE) experiments.

Insertion of a GFP-tag directly behind the conserved START codon enables detection of Miwi2 in the male, but not female, embryonic germ line through GFP fluorescence (Aravin *et al. Mol. Cell* 2008). This suggests that the transcript corresponding to the 5' end of miwi2 that can be detected in the ovary is likely non-coding. Alternatively, the protein might be expressed in some other developmental stage, which was not analyzed to date.

On the other hand, the products corresponding to the 3' region of the full-length transcript can be detected in the majority of adult tissues, suggesting the potential existence of a second promoter within the *miwi2* locus. Additionally, the databases contain the expressed sequence tags (ESTs) corresponding only to the 3' region of miwi2 (see Fig. 13a).

This issue is addressed in further detail later in the text.

5. I. 2. Analysis of the 5' end of miwi2: detection of two differentially expressed full-length transcript variants in testis and retention of intron 1.

Closer inspection of exon 1 revealed that it contains two in-frame START codons (Fig. 9a). The first START codon, at position 1-3 of the predicted full-length mRNA, is not conserved among species; only the protein translated from the second AUG (position 190-192) aligns with the closest homologues of Miwi2 (Fig. 9b). The mRNA containing the first AUG was present in testis at very low levels: using the forward primer positioned at the beginning of the predicted full-length mRNA ('e1alt', position 1-25) and the reverse primers spanning either the first 440 nucleotides (nt) (e3r) or the full (e21r) transcript, the mRNA was not detectable after 35 cycles of PCR.

a.

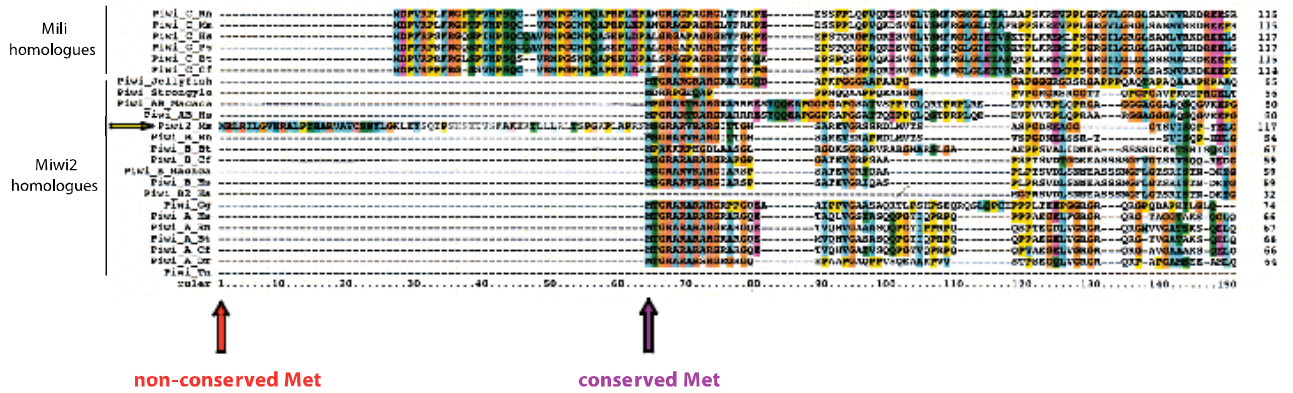
5'UTR EXON 1

--CCCGCCTTGCCGCGCCCATC **ATG**CGCCTGCGCA TTCTGGGGGTCCACGGGGCCCTGCCACCCACGCTC
 GGGTT GCTGTTTGTGTA GTAAGTA TCTTGGGAAGCTTGAAT ACTCCCAAATCCGA GTCACAGCCACACCGT
 CTCITTTT GCCAAAGAGAA ACACCTCCTGTT GAGGCTCACCAGCCCTGGAAGCCGCTGGCACCACAGGAAC
ATGAGTGGACGGGCCCCGT GTGAGA GCGCGAGGCATC ACCACTGCCCCACAGT GCTAGGGAAGTGGGGCGGT
 CCTCAAGAGA CCTCATG GTGAGTA CTCTGTCTAT CCTTCTTTGGCTA TCCTGGGCAGT GTTTTCTTCT--

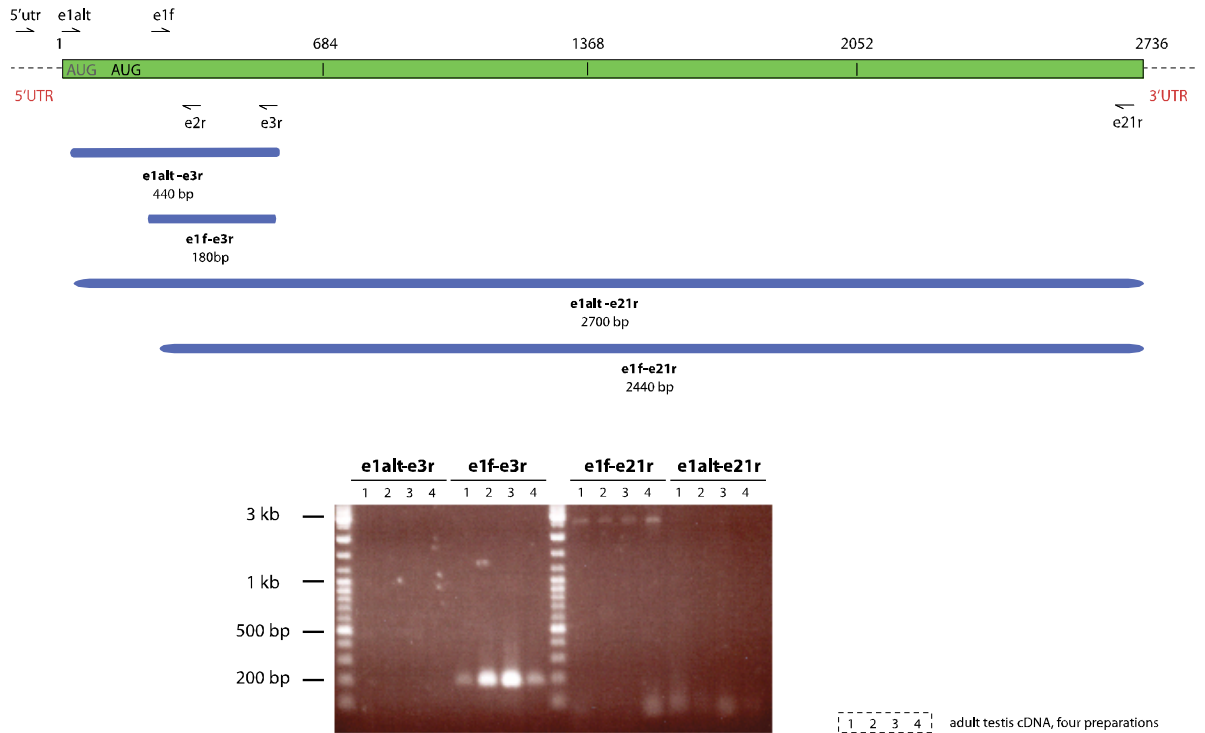
INTRON 1

non-conserved ATG/mRNA
 conserved ATG/mRNA

b.



c.



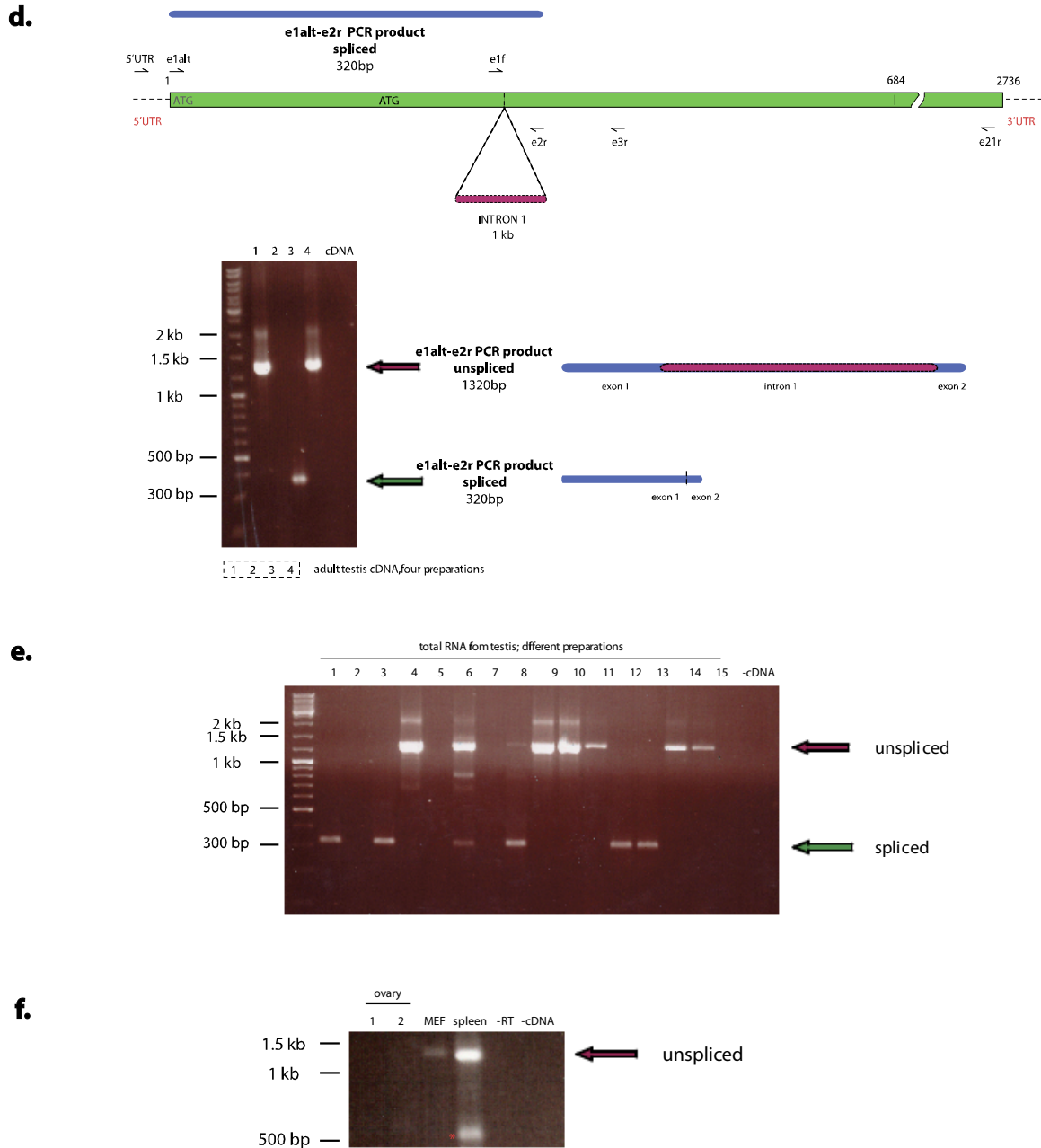


Figure 9. Analysis of the 5' end of the miwi2 mRNA.

- Genomic sequence of exon1 of *miwi2*, showing it contains two in-frame START codons.
- Alignment of the N-termini of the Piwi proteins from various species. The putative Miwi2 protein starting from the upstream START codon is not conserved among species.
- Detection of the non-conserved and conserved 5' *miwi2* variant in testis. The non-conserved transcript, unlike the conserved one, cannot be detected in the testicular tissue.

- d. Detection of the non-conserved 5' miwi2 variant using nested PCR. Both spliced and unspliced transcript variants can be detected.
PCR 1°: 5'UTR/e3r, 25 cycles; PCR 2°: e1alt/e2r, 35 cycles.
- e. Expanded analysis of the expression of the 5' end of miwi2 in testicular tissue. Intron 1 is often retained in the mature transcript. 1-15: different preparations of testicular RNA/cDNA.
- f. MEFs and female spleen, but not ovary, express the transcript containing intron 1 of *miwi2*. Asterisk marks the unspecific band. PCR: e1alt/e2r, 35 cycles.

However, if a forward primer (e1f), lying 250 nt downstream from the beginning of the mRNA, was used with the same reverse primers, the mRNA containing the conserved AUG was readily detectable (Fig. 9c). The mRNA containing the non-conserved AUG codon could be detected only upon performing the nested PCR, using the 5' UTR-e3r primer pair for the first PCR, and e1alt-e2r nested primer pair for the second PCR (Fig. 9d). Additionally, as revealed by sequencing of the PCR products, the unspliced version of the mRNA, containing intron 1, could also be detected (Fig. 9d). The same transcript, retaining intron 1, could be detected when a reverse primer in exon 3 is used, suggesting it is a part of otherwise correctly spliced transcript (data not shown). Expanded analysis of 15 different preparations of testis mRNA by RT and nested PCR revealed the presence of both spliced and unspliced mRNAs in the tissue (Fig. 9e). The unspliced variant could also be detected in MEFs and female spleen, but not in ovary, upon reverse transcription using a gene-specific primer (e21r) followed by 35 cycles of PCR using primers e1f and e2r (Fig. 9f). In female spleen, the spliced version could also be detected (see also Fig. 10g). This is particularly striking as the transcripts from the conserved 5' region of the mRNA can be detected in the ovary, but not in MEFs and spleen, after RT using the same gene-specific primer (Fig. 7e, upper panel).

The very low levels of the transcript containing the non-conserved AUG could be explained by the very weak activity of the promoter giving rise to it, or by its expression in a limited number of cells, e.g. in a defined population of spermatogenic cells. An mRNA *in situ* detection experiment with probes designed to distinguish the conserved and non-conserved transcript could give an answer to this. This transcript retains the coding potential, and the expression of the

alternative, non-conserved, longer Miwi2 protein in a small population of spermatogenic cells, or in a strictly defined time period, cannot be fully excluded. The non-coding transcript containing intron 1 is present both in testis, where the full-length transcript is expressed, and in spleen and MEFs, where the products corresponding to the conserved 5' region cannot be detected. PCR using a forward primer lying at the end of exon 2, after the retained intron, and a reverse primer in exon 15, does not yield a product in MEFs and spleen (data not shown), suggesting it is terminated earlier. Together, this data allow for the possibility that either transcription through the 5' end of the gene, potentially particularly through intron 1, or the mRNA arising from it, could have a regulatory role. However, as only a limited number of samples was analyzed, this cannot be concluded with certainty. As shown in Fig. 9e, in most cases, apart from samples six and eight, either spliced or unspliced variant is detected in the tissue. However, this could be the artifact of the nested PCR, and from this result it is not clear if the two variants are mutually exclusive.

Taken together, these data give rise to the possibility that more than one promoter upstream of the conserved START codon is utilized for transcription of the *miwi2* gene. The promoter that gives rise to the transcript containing the non-conserved START codon seems to be active in the male germ line, MEFs, and female spleen, and not in the ovary. This promoter seems to generate the potentially coding transcript in testis, and the intron 1-containing non-coding transcript in testis, spleen and MEFs.

On the other hand, transcripts corresponding to the 3' region of *miwi2* seem to be present in the majority of adult tissues. The existence of these transcripts was not detected before, and I therefore decided to analyze them in more detail.

As described in the next section, two *miwi2* knockout models have been generated (Carmell *et al.* *Dev. Cell* 2007; Kuramochi-Miyagawa *et al.* *Genes Dev.* 2008). The activity of the putative in-locus promoter, giving rise to the transcripts corresponding the 3' region of the full-length *miwi2* transcript, should not be affected in either of the

two knockouts. Therefore, it should still be possible to detect these transcripts in knockout tissues. Similarly, the activity of the promoter(s) upstream of the conserved START codon should also not be affected.

On the other hand, the deletion of exon 1, intron 1, and exon 2 in the Kuramochi-Miyagawa *et al.* knockout model provides potentially a good platform for examining the impact of the absence of the wild type 5' end of the *miwi2* gene, and intron 1, on the transcription of its different regions.

5. I. 3. Analysis of transcription from the *miwi2* locus in the *Miwi2* knockout mice.

Carmell *et al.* used insertional mutagenesis to disrupt the *miwi2* gene: 10kb of vector sequence and a part of the *miwi2* locus containing exons 9-12 were inserted into the wild type locus downstream of exon 12. This strategy left the 5' half of the gene – exons 1-12 – intact. Exons 9-12 became duplicated and are followed by exons 13-21 of the wild type locus (Fig. 10a). The full coding capacity of *miwi2* is still present in these mice; however, the wild type full length mRNA cannot be detected and, consequently, the full length protein is absent, resulting in the up-regulation of transposons in the male germ line, abnormal spermatocyte morphology, apoptosis of germ line cells, and male sterility (Carmell *et al. Dev. Cell* 2007; Fig. 10b). However, if the hypothesis of an alternative promoter within the *miwi2* locus is correct, the 3' half of the transcript should be detectable. Similarly, the promoter at the 5' end of the gene should not be affected in this knockout and could also still produce a transcript.

To check if this is the case, I reverse transcribed total RNA from testes of heterozygous and knockout mice using either oligo d(T) primer, a primer specific for the 3' UTR of *miwi2*, located between the STOP codon in exon 21 and the polyadenylation signal, and a primer binding within the coding sequence, at the beginning of exon 10, followed by PCR with primers spanning the full-length transcript, the 5' region alone, and the 3' region alone (Fig. 10a and 10b). As

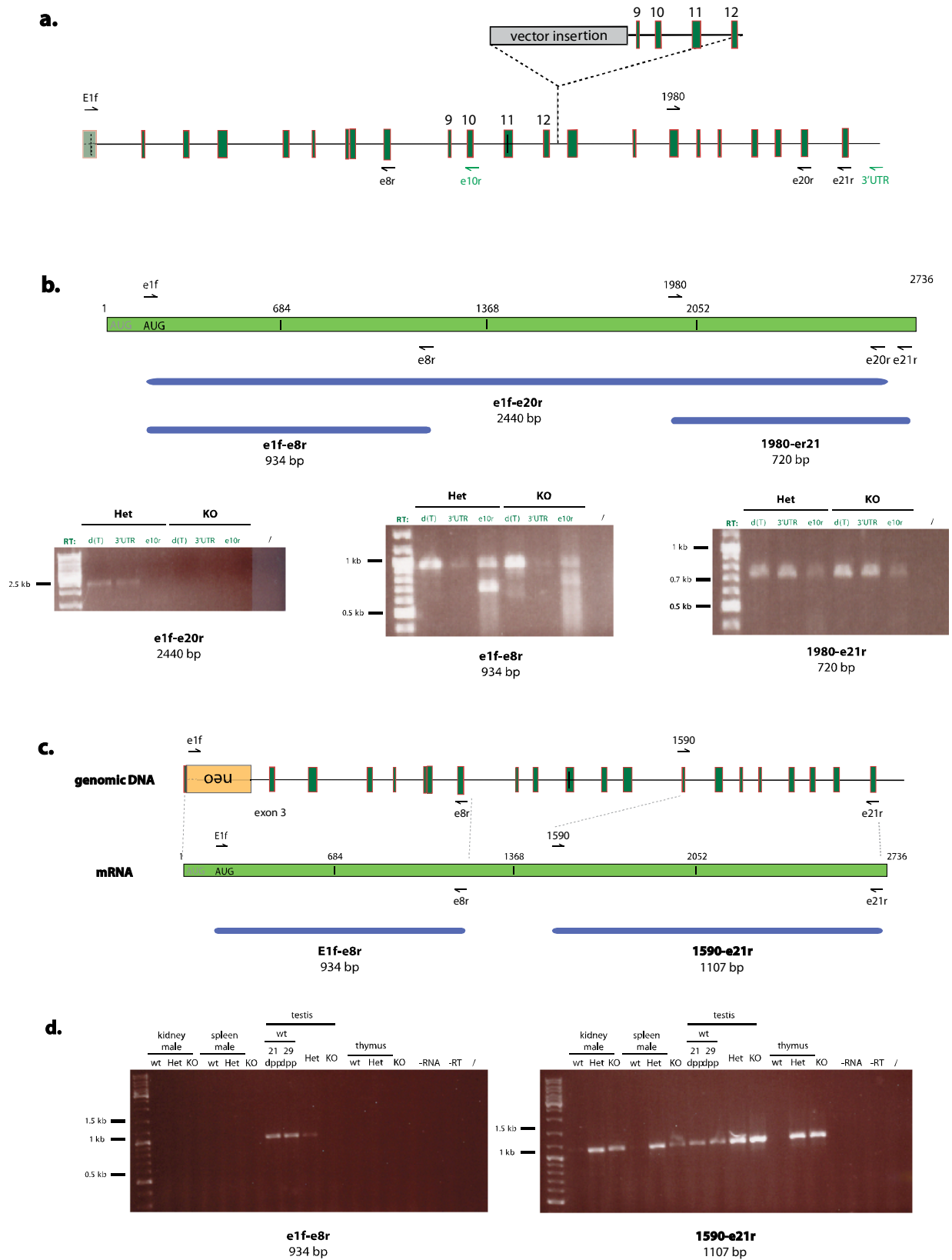
expected, the full-length transcript was present in the testicular tissue from heterozygous mice, but could not be detected in the knockout (Fig. 10b, left panel). However, as predicted, the transcripts corresponding to both 5' and 3' regions of the gene were still present in the knockout (Fig. 10b).

Higher levels of the e1f-e8r PCR product, corresponding to the 5' region, were detected after reverse transcription using oligo d(T) primer in comparison to their levels after reverse transcription using the primer in the 3' UTR in both heterozygous and knockout tissue (Fig. 10b, middle panel). This difference did not exist for transcripts from the 3' half of the gene (1980-e21r PCR product; Fig. 14b, right panel).

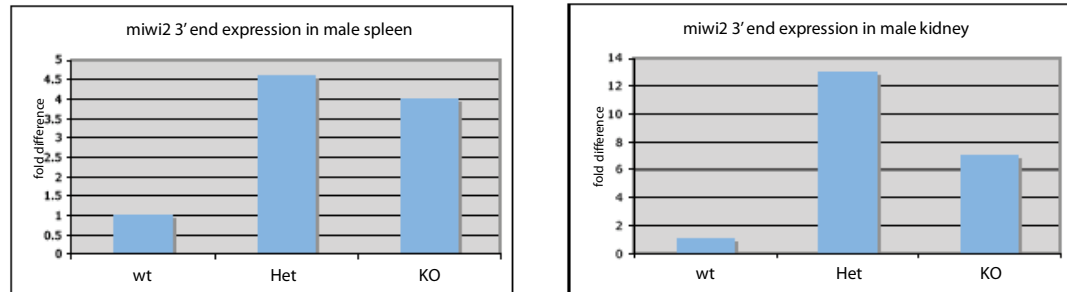
This observation suggests the possibility of the independent polyadenylation of the transcripts from the 5' region, which would not be surprising due to the presence of the vector insertion, likely containing the transcriptional termination region(s).

Surprisingly, the PCR fragment corresponding to the 3' half of the gene can be detected after reverse transcription using a primer in exon 10 (e10r), which is located upstream of the indicated PCR fragment (Fig. 10a). The most plausible explanation is the unspecific binding of the e10r primer to a site downstream from the indicated PCR fragment.

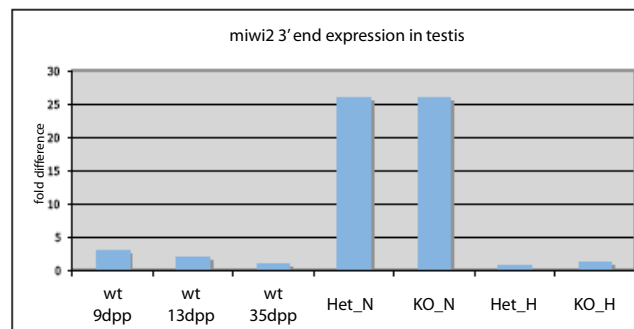
In the knockout model created by Kuramochi-Miyagawa *et al.*, the deletion of the 3' part of exon 1, intron 1 and exon 2 (Fig. 10c) should not affect the activity of the upstream promoter(s), but is predicted to give rise to a non-coding transcript with multiple STOP codons, which should be degraded through nonsense-mediated decay. Phenotypically, these mice are the same as the knockout mice created by Carmell *et al.* The putative in-locus alternative promoter should not be affected in these mice either, and the transcripts from the 3' half of the gene should be detectable.



e.



f.



testis Het_N; testis KO_N: heterozygous and knockout mice created by Kuramochi-Miyagawa *et al.*

testis Het_H; testis KO_H: heterozygous and knockout mice created by Carmell *et al.*

g.

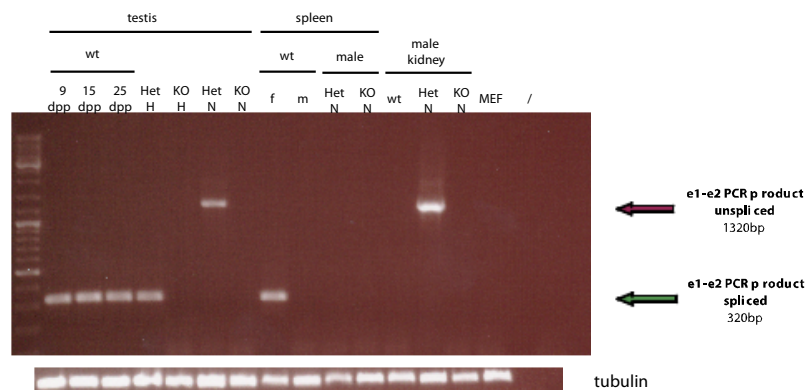


Figure 10. Analysis of transcription from the miwi2 locus in the knockout mice.

- Schematic representation of miwi2 locus in the knockout model generated by Carmell. *et al.* Primers used for RT are depicted in green. Primers used for PCR are depicted in black.
- RT-PCR analysis of miwi2 transcripts in the testis of the knockout mice (Carmell *et al.* 2007). The full-length transcript was absent from the knockout tissue, but the products corresponding to both 5' and 3' region could be detected. Schematic representation of the wild type miwi2 mRNA, with positions of primers used for PCR and expected products are depicted.

- c. Schematic representation of miwi2 locus in the knockout mouse generated by Kuramochi-Miyagawa *et al.*, with depiction of the corresponding mRNA regions and analyzed PCR products.
- d. RT-PCR analysis of expression of 3' region of miwi2 in the heterozygous and knockout mice created by Kuramochi-Miyagawa *et al.*
- e. Quantitative PCR analysis of expression of 3' region of miwi2 in somatic tissues. Tissues from mice created by Kuramochi-Miyagawa *et al.* were used for this analysis.
- f. Quantitative PCR analysis of expression of 3' region of miwi2 in testis from both models. All heterozygous and knockout samples are from adult mice.
- g. The 5'-end transcript containing intron 1 is upregulated in heterozygous kidney in mice created by Kuramochi-Miyagawa *et al.*

I performed reverse transcription of total RNA from heterozygous and knockout mice using oligo d(T) primer, followed by PCR using primers that should give rise to products corresponding to either 5' or 3' region of the full length transcript (Fig. 10c). As expected, the transcript containing exon 1 could not be detected by PCR in either somatic tissues (here spleen and kidney) or the KO testis, and was readily detectable in wild type and heterozygous testis (Fig. 10d). Surprisingly, the 3' transcript was not only detectable in the knockout tissues, but it also seemed to be significantly upregulated (Fig. 10d, right panel). The transcript also seemed to be upregulated in the heterozygous mice, and, strikingly, in the somatic tissues its levels even seemed to be slightly higher than in the tissues from the knockout mice (Fig. 10d).

qPCR using primer pairs giving rise to products depicted in Fig. 8a confirmed the up-regulation of the 3' transcript levels in both heterozygous and knockout tissues, as well as the slight difference in its levels between the heterozygous and knockout somatic tissues (Fig. 10e). The difference was more pronounced in kidney, almost 2-fold, and was only slight in spleen. In testis, the transcript seems to be upregulated to roughly the same levels in both heterozygous and knockout mice from this model. However, its levels are unaffected in the testicular tissue from the model created by Carmell *et al.* (Fig. 10f).

Transcripts containing intron 1 could be detected in the wild type testis, female spleen, and MEFs (Fig. 9. d, e and f). Performing the same nested PCR on a limited number of samples from the testicular tissue of both miwi2 models and the

somatic tissues' samples from the Kuramochi-Miyagawa *et al.* model showed that the unspliced transcript can be detected in the heterozygous male kidney and testis. The transcript could not be detected in male spleen from the mice created by Kuramochi-Miyagawa *et al.*, or in the wild type kidney (Fig. 10g).

The absence of product in the KO tissues from this model is not surprising, as the binding site for the nested forward primer is deleted. PCR products could not be detected either in the KO testis from the mice created by Carmell *et al.*, or in MEFs, where they could be detected in a separate experiment (see Fig. 9f). Only the spliced version is detected in the wild type testicular tissue.

As shown by these results, the transcripts corresponding to the 3' region of *miwi2* can be detected in heterozygous and knockout testicular tissue from both *miwi2* models, as well as in heterozygous and knockout somatic tissues from the model created by Kuramochi-Miyagawa *et al.* (transcription in the somatic tissues from the model created by Carmell *et al.* was not analyzed due to lack of material). These data speak in favor of the possibility of existence of an alternative promoter within the *miwi2* locus.

This putative promoter seems to have low activity in the wild type tissues, giving rise to transcripts detected by RT-PCR. However, upon the insertion of the *neo* cassette and consequent partial deletion of exon 1 and complete deletion of intron 1 and exon 2, the levels of the 3' transcripts are elevated in both germ line and somatic tissues, possibly suggesting the enhancement of the promoter activity.

The observed up-regulation of the 3'-half transcript could be a consequence of the insertion of the *neo* cassette into the locus, the absence or changed levels of either the full-length transcript or the full-length protein, of the deletion of the 5' region of the gene, particularly intron 1, and the putative regulatory elements it contains, as well as consequential absence of the transcript it gives rise to. Finally, the up-regulation could also result from a combination of any of these changes.

The *neo* cassette had been described to have 'neighboring effects' and affect the activity of the surrounding genes, particularly in the multi-gene clusters (Pham *et al.* PNAS 1996). However, the transcript seems to be upregulated to the higher levels in the heterozygous compared to the knockout somatic tissues, and to the same

levels in the heterozygous and knockout testis (Fig. 10e and f). If the observed transcript up-regulation would be merely the consequence of the presence of the *neo* cassette, then it could be expected that the levels would increase linearly from heterozygote to the knockout tissues. As this is not the case, the observed up-regulation is likely not the simple consequence of the presence of the *neo* cassette. Adult somatic tissues do not express the full-length transcript and protein, and it is therefore unlikely that their absence would cause the observed up-regulation of the 3' transcript in the knockout mice. In the testis, the absence or changed levels of either the full-length transcript or the full-length protein could theoretically cause the up-regulation of transcription from the in-locus promoter. However, as revealed by the qPCR analysis of the 3' transcript in testis, there is no difference in its levels between heterozygous and knockout tissues in both *miwi2* models (Fig. 10f). This result suggests that the full-length transcript and protein are likely not involved in the regulation of transcription from the in-locus promoter, or are not the only factors affecting it.

Rather, the observed ~25-fold higher transcript levels in the testicular tissue from the Kuramochi-Miyagawa *et al.* model (Fig. 10f) suggest that the absence of the 5' region of the gene, containing intron 1, the mRNA transcribed from it, or the combination of the two, could play a role in regulation of transcription from the putative downstream promoter. The surprising observation that the levels of the 3' transcript seem to be highest in the heterozygote somatic tissues, compared to the wild type and knockout tissues, suggest the involvement of both *in-cis* and *in-trans* mechanisms in the regulation of the putative in-locus promoter activity. However, a more thorough analysis is required to address this issue in detail.

An additional factor that can be considered here, if the 5' region of the gene is thought to be involved in the regulation of the putative downstream promoter, is the distance between the two regions on the chromosome. However, and assuming that the in-locus promoter is located downstream from the vector insertion in the mice created by Carmell *et al.*, the 10 kb change in distance does not lead to the change in the 3' transcript levels in testicular tissue (Fig. 10f).

Additionally, as shown in Fig. 10g, analysis of the limited number of samples suggested that the absence of this 5' genomic region might have an impact on the activity of the upstream promoter, judged by the up-regulation of the unspliced transcript in the male heterozygous kidney. Interestingly, this unspliced transcript, also present in the heterozygous testis from the Kuramochi-Miyagawa *et al.* model, has the same size as the unspliced transcript in the wild type testis and spleen, suggesting it is derived from the wild type and not the targeted allele. If the observed up-regulation is true, this could potentially suggest that this 5' genomic region might contain elements that *in-trans* regulate the activity of the promoter used for its transcription on the homologous chromosome in some tissues. It is at present not clear why the transcript would be upregulated in kidney, and not in spleen. In the female, but not male, wild type spleen, both spliced and unspliced version can be detected (Fig. 9f and 10g). This could also suggest the possible existence of sex-specific regulatory mechanisms, which were not investigated in further detail here. An additional important question is whether the unspliced 5' transcript is also upregulated in the knockout tissues from the mice created by Kuramochi-Miyagawa *et al.*; the product could not be detected using a reverse primer in exon 3 and a forward primer at position 58 – 82 of exon 1 (data not shown). However, it has to be stressed here again that this is purely speculative, and a much broader analysis is necessary to confirm or disprove these assumptions.

In summary, analysis of *miwi2* transcription in the two knockout models showed that the transcripts corresponding to the 3' region of the gene are still present, both in germ line and in somatic tissues. This further suggests the existence of a separate promoter within the *miwi2* locus, giving rise to these transcripts. In addition, the observed up-regulation of these transcripts upon deletion of the regions on the 5' end of the gene could imply that this region is involved in the regulation of the downstream promoter.

The involvement of the 5' end in the regulation of the upstream promoter, the existence of different regulatory mechanisms in different tissues, and potentially a

regulatory network involving different regions of the gene, can also not be completely excluded. Exact identification of promoter regions and their subsequent deletion could help address these questions.

5.1.4. 5'Rapid Amplification of cDNA Ends (RACE) analysis of *miwi2*.

In order to determine the 5' ends of transcripts from the 3' region of the *miwi2* gene, I performed 5' RACE experiments (Fig 11). Total RNA was reverse transcribed using a gene-specific primer (e21r), and the cDNA was 5' tagged and amplified by nested PCR. Sequencing of the 5' RACE products marked with an asterisk in Fig. 11a revealed transcripts with 5' ends roughly in the middle of the full length mRNA (Fig. 11b), but a unique start was not evident. In testis, both the full-length transcript containing the conserved START codon and the alternative short transcripts corresponding to the 3' region could be detected. The transcript containing the non-conserved START codon was not detected by 5' RACE, consistent with its low levels. As revealed by sequencing, nearly full-length mRNAs detected in some tissues (e.g. spleen) lacked some exons and are non-coding due to the frameshifts introduced by such splicing.

The majority of 5' ends mapped onto exons, primarily exon 10, and there were several examples (ovary, spleen, testis; see Fig. 11b) of mRNAs beginning in introns 9 and 10. This is mostly consistent with the PCR analysis, which showed that using a forward primer in exon 10 ('1320') with the reverse primer in the ubiquitously expressed 3' region yields a product, unlike the forward primer lying in exon 4 (Fig. 7e). Importantly, the transcript from the 3' region of *miwi2*, with the 5' end in exon 10, can be detected in the total mRNA isolated from the *miwi2* knockout (KO) mice created by Carmell *et al.* (Carmell *et al. Dev.Cell* 2007), suggesting again that it is expressed independently of the full-length wild-type form not present in these mice (Fig. 11a, right panel). If the position of the 5' end is correctly mapped, the putative promoter would be positioned downstream of the vector insertion in

these mice, as assumed earlier. This result is therefore also consistent with the PCR analysis (Fig. 10b).

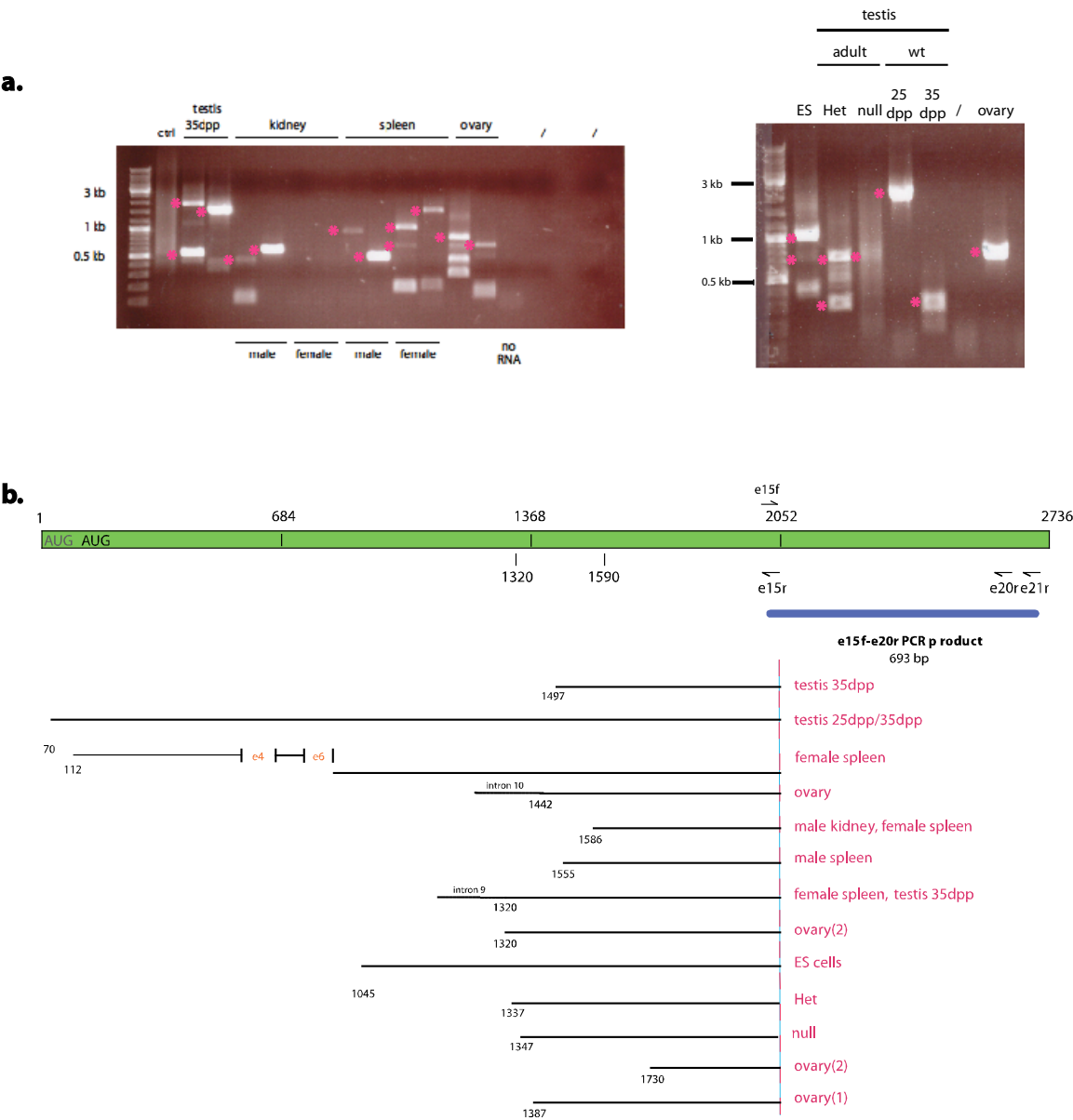


Figure 11. 5' RACE analysis of miwi2 mRNA in various tissues and ES cells.

- a.** Analysis of 5'RACE products. Bands marked with an asterisk were analyzed by sequencing.
- b.** Comparison of full-length miwi2 mRNA and 5'RACE products.
e15f-e20r PCR product can be stably detected in all samples analyzed.
e21r primer was used for reverse transcription. e20r primer was used for the 1st PCR; e15r primer was used for the 2nd PCR.

The short transcripts arising from the 3' region of the gene were correctly spliced and therefore could potentially still code for a protein product. Sequence inspection revealed the presence of several in-frame AUG codons in the identified transcripts (Fig. 12a). Inspection of their surrounding sequences revealed that none of the AUGs has a canonical Kozak sequence context required for efficient initiation of translation, so it is at present not clear which AUG, if any, is utilized for protein production (Kozak *Mamm Genome* 1996). The expected molecular weights of putative proteins arising from these transcripts are depicted in Fig. 12a and correspond to the unmodified protein. Post-translational modifications and/or differential splicing at the 3' end of the gene, which were at present not analyzed, could also influence the size of the protein. This protein forms would lack the PAZ domain, but would still have the Piwi domain containing the active site responsible for the nuclease activity of the Argonautes, and could still bind RNAs through motifs in the mid-domain on the N-terminus of the Piwi domain, which usually anchor the 5' end of the small RNA (Fig. 2 and 12b).

The most plausible explanation for the observed discrepancy in the positions of the identified 5' ends are the technical issues during the 5' RACE procedure, such as inefficient reverse transcription, faulty annealing of the primers during PCR, or partial degradation of the transcripts. Another possible explanation is the existence of several alternative transcriptional starting sites in the close proximity to each other, which are differently utilized in different tissues (see Fig. 13a).

I next analyzed if the regions of the *miwi2* locus in the vicinity of the identified 5' ends contain putative promoters.

5.1.5. Analysis of the *miwi2* locus: presence of alternative transcriptional starting sites.

As described above, transcriptional analysis of *miwi2* by RT-PCR and 5' RACE suggested the existence of alternative transcriptional starting sites (TSSs) within the locus.

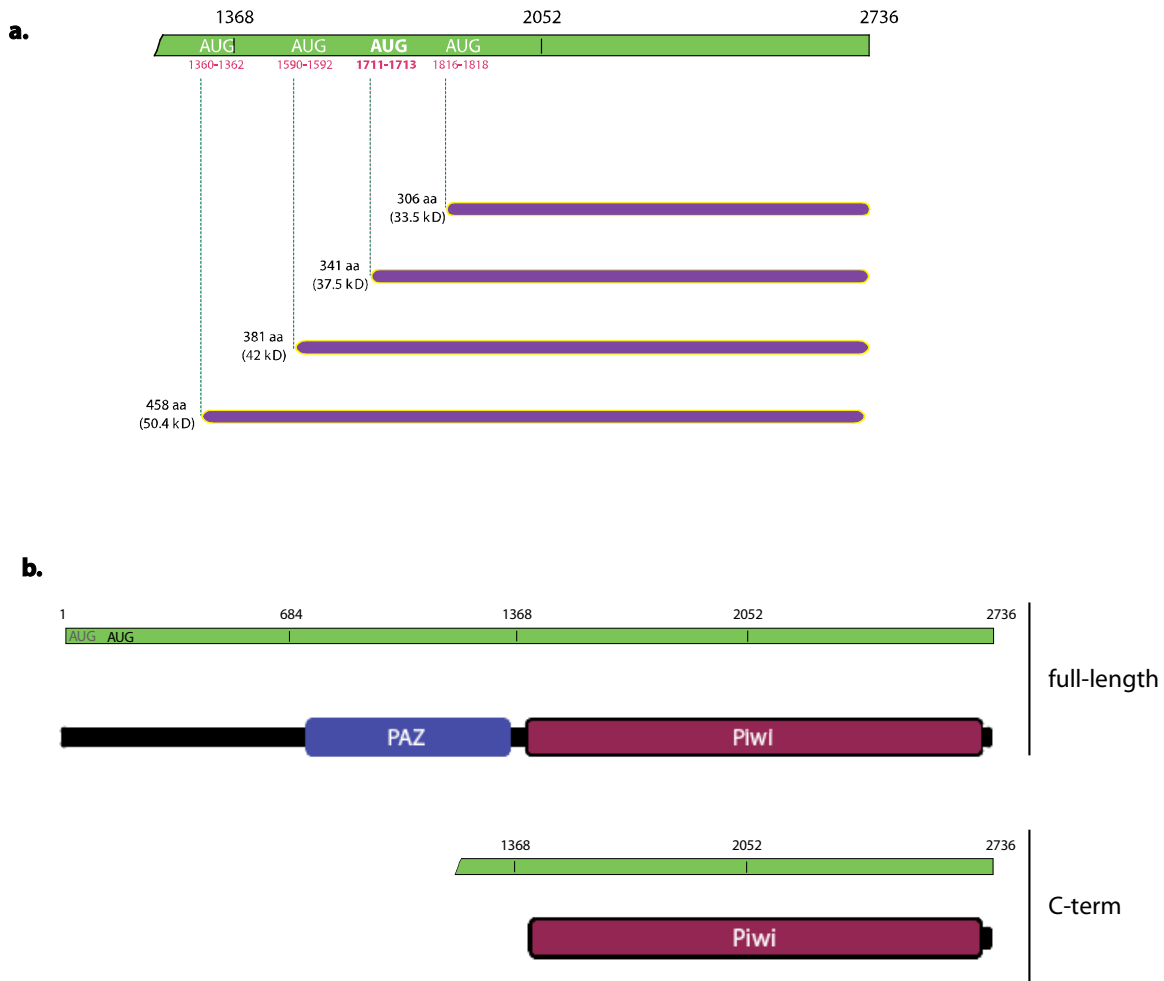
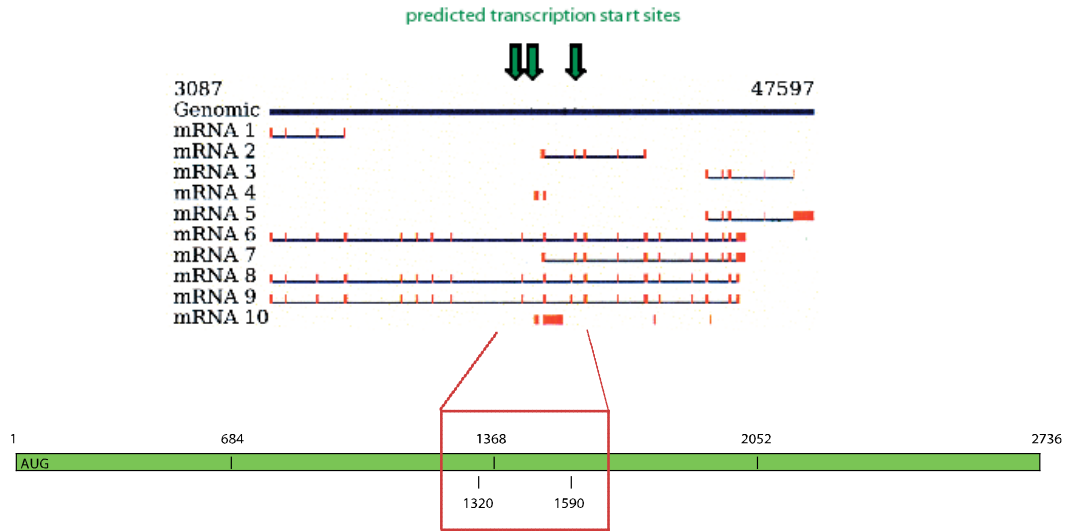


Figure 12. Transcripts from the 3' region contain several in-frame AUG codons.

- a.** The positions of in-frame AUG codons (in white) on the 3' transcripts, which could give rise to depicted protein variants. The indicated nucleotide positions of AUGs correspond to their positions on the full length mRNA.
- b.** Comparison of the domains of the full-length Miwi2 and the putative C-terminal variant.

Analysis of the locus by the Spidey TSS-searching web interface (Wheelan *et al. Genome Res.* 2001) confirmed that there are several putative TSSs in the region of the locus that could give rise to the detected mRNAs (Fig. 13a). The presence of several TSSs in the region where the 5' ends of RACE products map could explain the observed heterogeneity of identified 5' ends of mRNAs from different tissues (Fig. 13b).

a.



b.

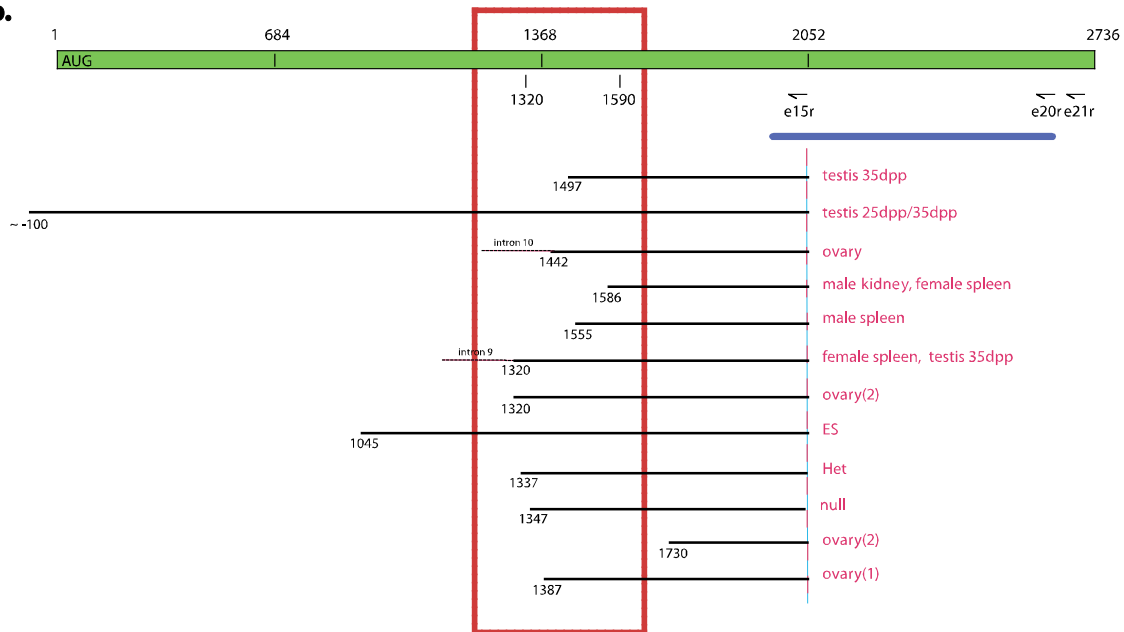


Figure 13. *miwi2* locus contains several potential transcriptional starting sites (TSSs). 5' ends of RACE products coincide with the presence of an alternative TSSs in the *miwi2* locus.

- a.** Schematic representation of *miwi2* locus with relative positions of alternative TSSs is depicted. The red square marks the region on the full-length transcript corresponding to the genomic region where the identified TSSs are located.
- b.** The 5' ends of RACE products coincide with the putative TSSs.

Cloning the computationally identified putative promoters in front of a reporter gene, inserting a reporter gene into the *miwi2* locus downstream of the predicted starting site(s), and/or knocking out the genomic regions comprising the putative promoters could further confirm if these regions are in fact real transcriptional starting sites.

5.1.6. Identified *miwi2* transcriptional and splicing variants.

As revealed by sequencing of RT-PCR and 5' RACE products, several different transcriptional and splicing variants can be detected in different tissues and cell lines (Fig. 7, 9 and 11).

The most prominent transcriptional variants are the ubiquitously expressed transcripts from the 3' region of the gene. The existence of an independent transcript arising from the 5' region in the ovary can also not be excluded (Fig. 7).

Comparing the sequence of the *miwi2* mRNA I have detected to the sequence in the NCBI Nucleotide database (NM_177905.3) revealed inclusion of some exons and exclusion of others (Fig. 14). Namely, I have never detected exon 11, included in the NCBI mRNA, as a part of the mature *miwi2* mRNA. On the other hand, exons 12 and 19, not included in the NCBI mRNA, are always present in my analysis, as is the 5' extension of exon 7. These changes do not affect the open reading frame and the transcript remains coding.

Nearly full-length mRNAs detected in some tissues (Fig. 7 and 11) lack certain exons and are non-coding due to the frameshifts introduced by such splicing. The most prominent ones are the one in which exon 3 is spliced directly to exon 5 and exon 4 is excluded, the one in which exon 5 is spliced directly to exon 7 and exon 6 is excluded, and the one in which only the first 31 nucleotides on the 5' end of exon 5 are included and spliced directly to exon 6, leaving out the remaining 90 nucleotides of exon 5. Transcripts missing both exons 4 and 6 can also be detected. Additionally, a non-coding transcript retaining intron 1 can also be detected in certain tissues (Fig. 8e, 8f and 10g).

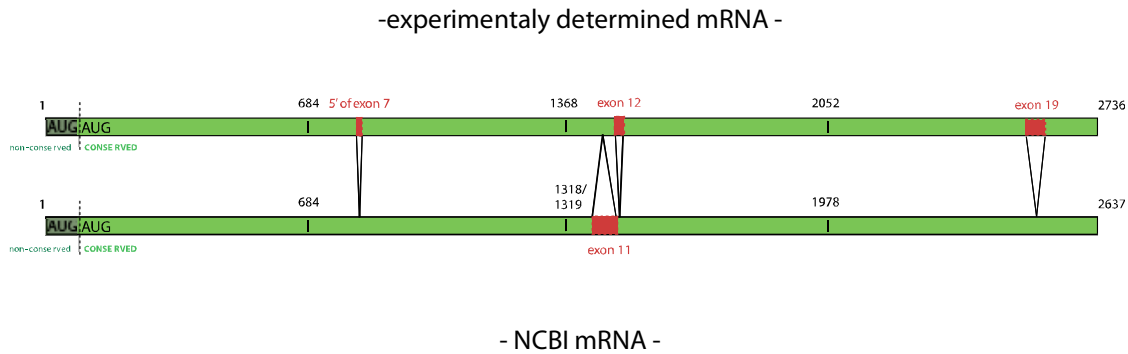


Figure 14. Comparison between NCBI miwi2 mRNA and experimentally determined mRNA.

Therefore, both transcription and splicing of miwi2 are dynamically regulated in tissues, possibly related to developmental status or, in the case of testis, spermatogenic cycle.

Despite the fact that exon skipping and intron retention render the splicing variants non-coding, they seem to be stable in tissues, and not degraded through nonsense-mediated decay. As the transcripts are non-coding, their role is at present not clear. One possibility is that transcription from one of the promoters affects the activity of other(s), as the analysis of transcription in the heterozygous and knockout tissues suggested (Fig. 10), and alternative splicing prevents the expression of the protein. Furthermore, transcripts themselves could have a regulatory role.

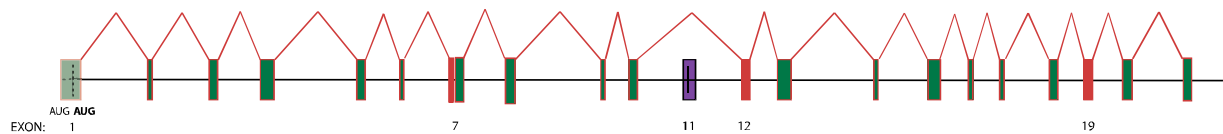
All the identified transcriptional and splicing variants are summarized in Fig. 15a. An overview of the more prominent transcriptional variants and tissues and cells they were identified in is given in Fig. 15.b

An example of genes giving rise to transcripts from alternative promoters during development in both somatic and germ line tissues are DNA methyltransferases (Trasler *et al. NAR* 1992; Mertineit *et al. Development* 1998; Chen *et al. J Biol. Chem.* 2002; Shovlin *et al. Hum. Reprod.* 2007). Dnmt1, the methyltransferase responsible for maintenance of DNA methylation patterns in proliferating cells, and members of the Dnmt3 family, *de novo* methyltransferases Dnmt3a and Dnmt3b and catalytically inactive Dnmt3L, required for establishment of methylation of imprinted genes and

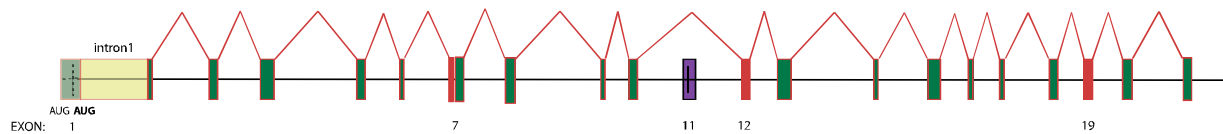
transposons during embryonic germ line development (Bourc'his and Bestor, *Nature* 2004; Chen *et al. J Biol. Chem.* 2002), have all been shown to employ alternative splicing and promoter switching in different tissues and at different developmental stages for production of alternative transcripts.

a.

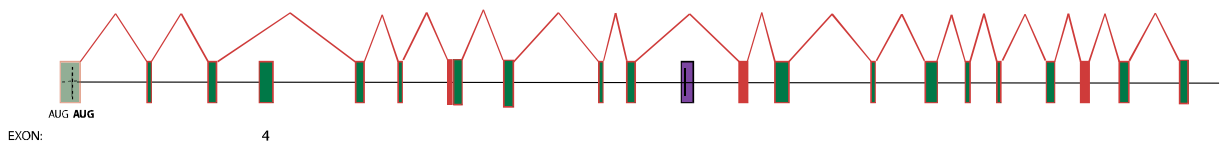
wild type, full-length; coding; testis-specific



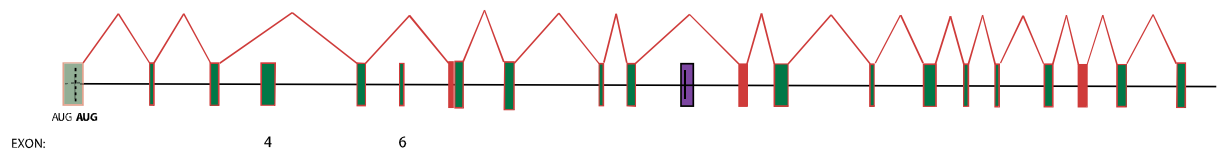
retention of intron 1; non-coding; testis, female spleen, MEFs



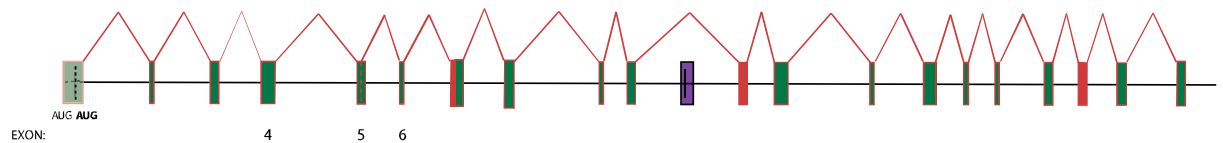
skipping of exon 4; non-coding; ubiquitous?



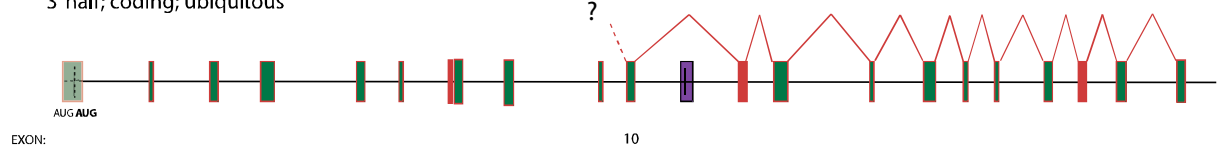
skipping of exons 4 and 6; non-coding, ubiquitous?



partial exon 5 inclusion; non-coding; somatic tissues



3' half; coding; ubiquitous



b.

	MEF	testis	ovary	spleen	kidney	thymus
full-length transcript	—	+	—	—	—	—
5' transcript	—	?	+	—	—	—
3' transcript	+	+	+	+	+	+
unspliced 5' end w/ intron 1	+	+ wt + Het	?	+ female wt — male wt — male Het	— male wt + male Het	?

Figure 15. Transcriptional and splicing variants of miwi2: a summary

- a.** Exons are depicted in green. Exons or parts of exons not present in the NCBI mRNA are depicted in red. Exon 11, in purple, is not detected in the mRNA but is contained in the NCBI mRNA. Exon 1, in pale green, contains two in-frame ATGs. Second ATG in exon 1, in bold, is conserved. Retained intron 1 is depicted in yellow.
- b.** An overview of the more prominent miwi2 transcriptional variants and tissues or cells they were identified in.

In some instances, as is the case for Dnmt3L, the short transcripts arising from an in-locus, intronal promoter during spermatogenesis seem to be non-coding and were suggested to have a role in regulation of the full length form (Shovlin *et al. Hum. Reprod.* 2007). On the other hand, a shorter transcript of Dnmt3a, Dnmt3a2, also produced from an intronal promoter, gives rise to a protein with different sub-cellular localization and developmental expression pattern. Namely, Dnmt3a is expressed ubiquitously and localizes to heterochromatin, whereas Dnmt3a2 is expressed only in tissues in which *de novo* DNA methylation is believed to occur during cellular differentiation, such as testis, ovary, spleen and thymus, and localizes to euchromatin (Chen *et al. J Biol. Chem.* 2002). Similar observation has been made about Dnmt3b (Chen *et al. J Biol. Chem.* 2002).

Miwi2 is involved in repression of transposable elements through DNA methylation in the male germ line. It likely acts upstream of DNA methyltransferases involved in this process, Dnmt3a, Dnmt3b and Dnmt3L, as the piRNA pathway is largely unaffected in the *dnmt3L* knockout mice (Aravin *et al. Mol. Cell* 2008).

My analysis of miwi2 expression on the mRNA level suggests that regulatory strategies, similar to ones regulating the expression of DNA methyltransferases, could act upon the *miwi2* locus, potentially with different regulatory strategies being employed in different tissues. This could be analyzed by interfering with the activity of the promoters by deleting them or by introducing insulators.

5. II. Analysis of putative different forms of Miwi2, their expression outside of the germ line, sub-cellular localization, and potential functions.

The presence of the alternative, potentially coding miwi2 transcripts in tissues and cells suggested that a shorter protein form of Miwi2 could be expressed in some tissues. Additionally, the full-length transcript is present in the post-birth spermatogenic cells, therefore the possibility of the expression of the full-length protein cannot be excluded.

I used antibodies directed against different parts of the Miwi2 protein to detect its potential different forms by Western blot analysis, immunohistochemistry and immunocytochemistry.

5. II. 1. The Miwi2 antibodies.

We generated two antibodies in order to investigate the expression and localization of various potential forms of the Miwi2 protein. Additionally, I used a commercially available antibody directed against the C-terminus of the Miwi2 protein.

The two custom-made antibodies are directed against peptides in either the N-terminal or the C-terminal part of the full-length protein (Fig. 16a). Both antibodies could recognize full-length recombinant Miwi2 protein on the Western blot (Fig. 16b). The N-terminal antibody detects Miwi2 in the embryonic male gonads, and was successfully used for immunoprecipitation of Miwi2 in the preparation of the embryonic piRNA library (Aravin *et al. Mol. Cell* 2008).

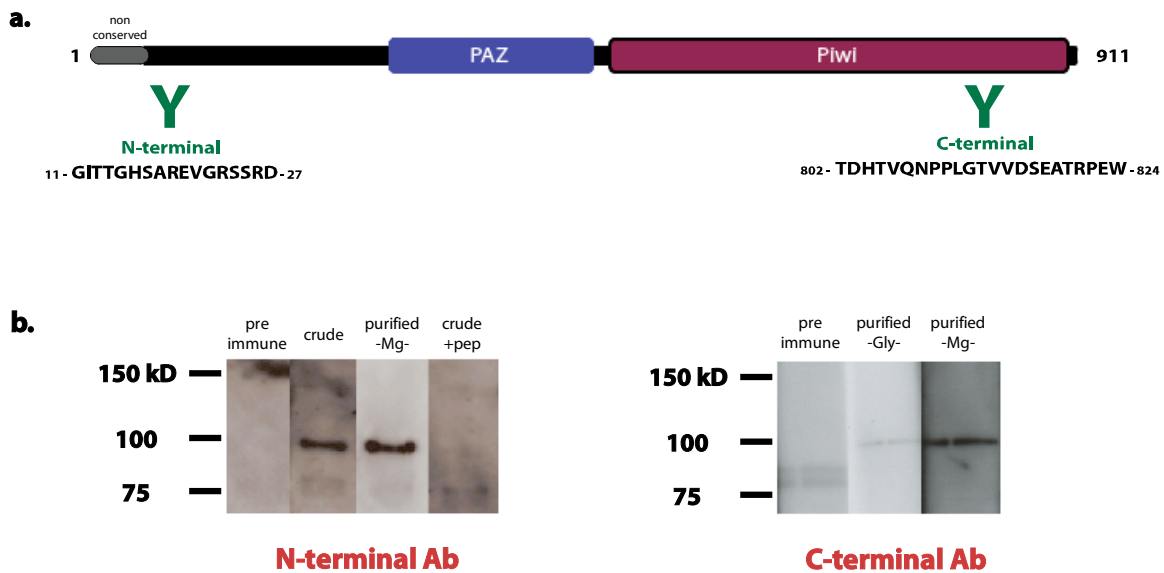


Figure 16. Miwi2 antibodies.

- a.** Sequence and position of the two peptides used for antibody production.
- b.** Both N-terminal and C-terminal antibody recognize recombinant Miwi2 protein on the Western blot.

5. II. 2. Detection of Miwi2 using the N-terminal antibody.

Using the N-terminal antibody, it should only be possible to detect Miwi2 in testis, providing that the mRNA is translated. Therefore, I first investigated if miwi2 mRNA is associated with polysomes. Testicular cytoplasmic extract from wild type 16 days old (16 days-post-partum; dpp) and adult mouse were separated on the sucrose

gradient, and total RNA was isolated from mono-ribonucleoprotein (mRNP) and polysome fractions. RT-PCR analysis using primer pairs 60-e3r and e15f-e20r showed that miwi2 mRNA can be detected in polysomal fractions (Fig. 17a). Therefore, even though the transcript levels are much lower than in the embryonic gonads, this experiment suggest it could still be translated into a protein.

Using the N-terminal antibody, I detected the ~100 kD band in the nuclear fraction of testicular single-cell suspension, prepared from testis of ~2 weeks old mice (Fig. 17b).

The band is lost upon incubation of the antibody with the epitope-specific synthetic peptide, suggesting specificity (Fig. 17b). The levels of the protein are very low, as it can only be detected in the nuclear fraction of testicular single-cell suspension, and not in the total protein extract of the full organ (Fig. 17b and c). Very low protein levels in post-birth testis could also be the reason why the full-length Miwi2 protein fused to GFP cannot be detected in adult testis through fluorescence.

The ultimate system to test if any protein form containing the epitope is still present in the tissues is the miwi2 knockout model created by Kuramochi-Miyagawa *et al.* where the epitope of the N-terminal antibody at the end of exon 1 has been deleted. Consistent with the lack of full-length transcripts in the majority of adult tissues and the apparently very low protein levels in post-birth testis, no specific band could be detected with the N-terminal antibody in the wild type tissues as compared to the knockout (Fig. 17c). Additionally, a transcript corresponding to the 5' half of the gene is expressed in ovary, and it could potentially be translated into a protein product. However, I could not detect any shorter protein forms using the N-terminal antibody, suggesting that this transcript is not translated or is present at the very low levels (Fig. 17c).

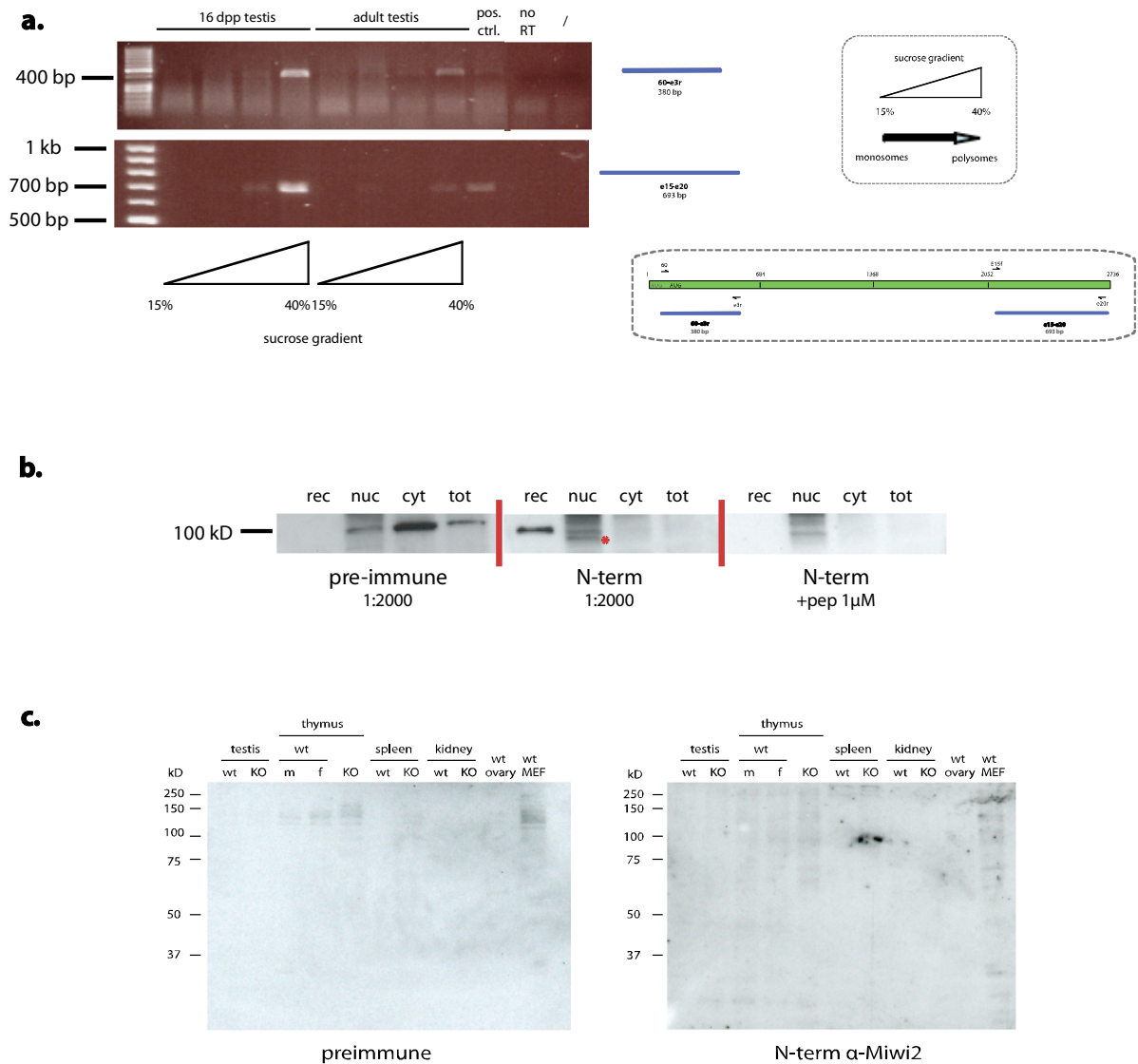


Figure 17. Detection of Miwi2 in adult mouse tissues.

- miwi2 mRNA can be detected in the polysomal fractions, suggesting it is translated.
- Detection of Miwi2 in the testicular single cell suspension using N-terminal antibody.
A band marked by the asterisk is neither recognized by the pre-immune serum nor after blocking the antibody with the N-terminal peptide used for antibody production.
rec: recombinant protein; nuc: nuclear fraction; cyt: cytoplasmic fraction; tot: total single-cell suspension protein extract; pep: N-terminal peptide used for antibody production.
- Detection of Miwi2 in adult mouse tissues using N-terminal antibody. In the KO mice from which the tissues were taken, the epitope for the antibody is deleted.

5. II. 3. siRNA-mediated knockdown of Miwi2 in MEFs and protein detection using the C-terminal antibody.

Our second custom-made antibody and the Abcam antibody both have epitopes in the C-terminal region of the protein. Therefore, if the transcripts from the 3' region of the gene give rise to a shorter protein, these antibodies could be used to detect the alternative protein form(s).

Mouse embryonic fibroblasts are a suitable system to test this, as they express the 3' transcripts and can be transfected with siRNAs for knockdown experiments. If the 3' transcripts give rise to protein, it should be possible to detect a decrease in protein levels on the Western blot upon knockdown. I therefore designed two siRNAs specific for the 3' half of miwi2 (siRNAs #1 and #3), and one specific for the 5' half as a control (siRNA #2), using the RNAs tool (Tafer *et al. Nat Biotech* 2008) (Fig. 18a).

Transfecting siRNAs into MEFs followed by RNA isolation and RT-PCR showed that siRNA #3, but not #1 and #2, causes reduction of miwi2 mRNA levels in MEFs (Fig. 18b and c).

I next analyzed protein extracts from the same cells on the Western blot using both antibodies with C-terminal epitopes. The custom-made C-terminal antibody detects a band of ~45 kD in MEFs and adult mouse tissues (Fig. 19a). This size of protein is approximately what can be expected if the in-frame START codons present on the transcripts detected in MEFs would be translated (Fig. 12a). However, no decrease in the protein levels were detected on the Western blot upon the knockdown (Fig. 19b).

The C-terminal antibody Abcam antibody detects a band of ~75kD on the Western blot of MEF protein extracts (Fig. 19c), and a moderate decrease in the protein levels can be observed upon knockdown of miwi2 using siRNA #3 (Fig. 19d). This, however, is not the expected size of the protein product that should be translated from the 3'-half transcripts (Fig. 12a). The same ~75kD band can be detected in some of the adult mouse tissues, such as spleen and thymus, and not in testis and kidney. In the ovary, a band of ~100kD is detected (Fig. 19c).

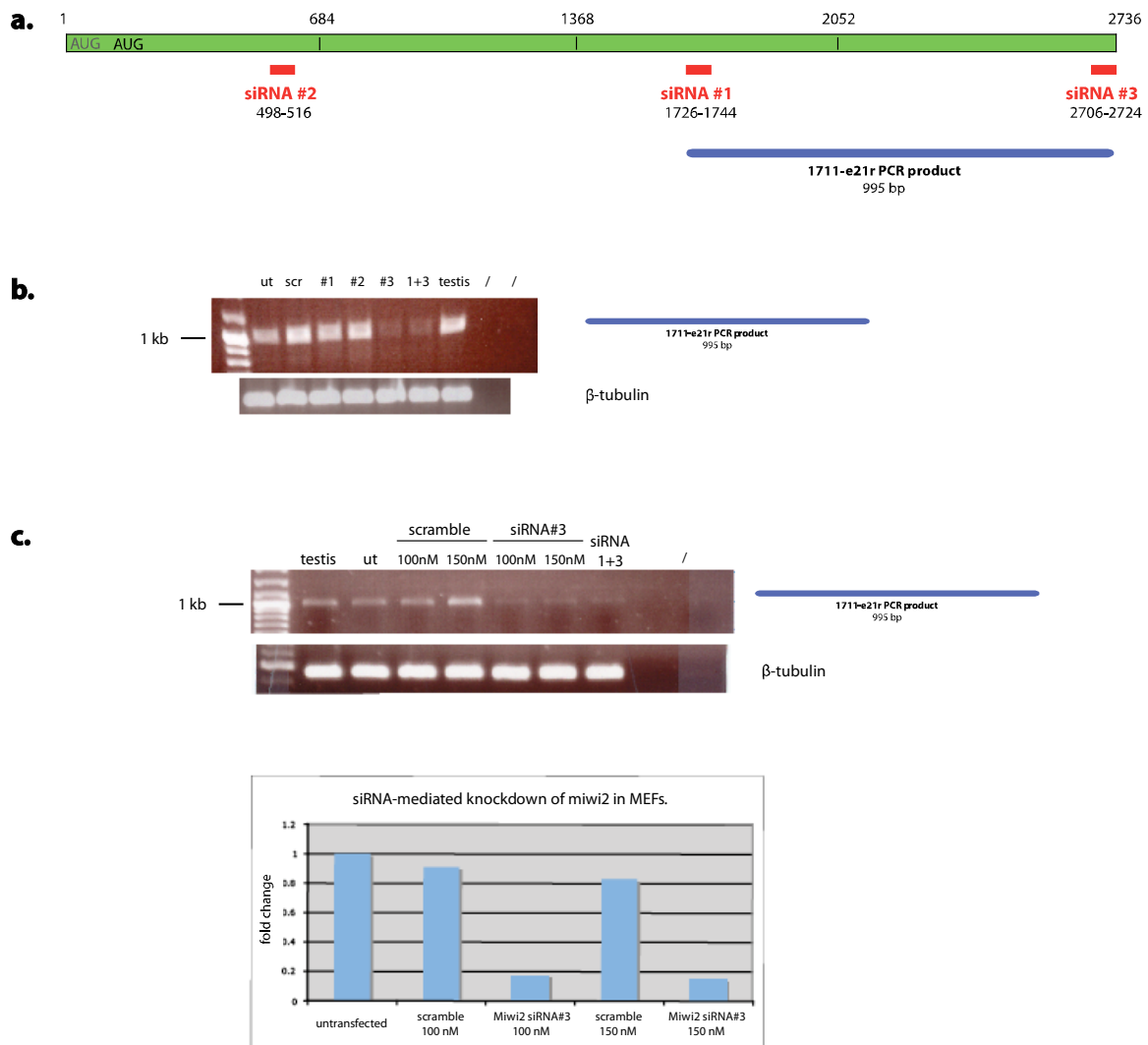
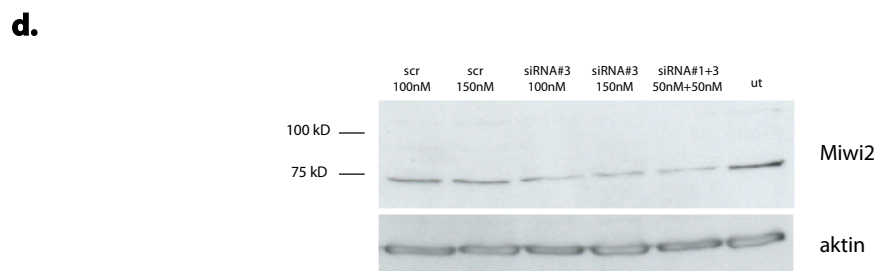
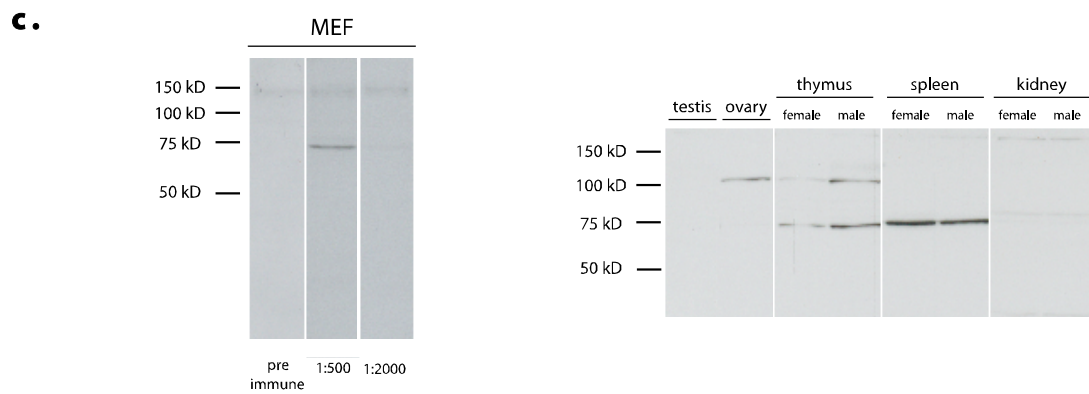
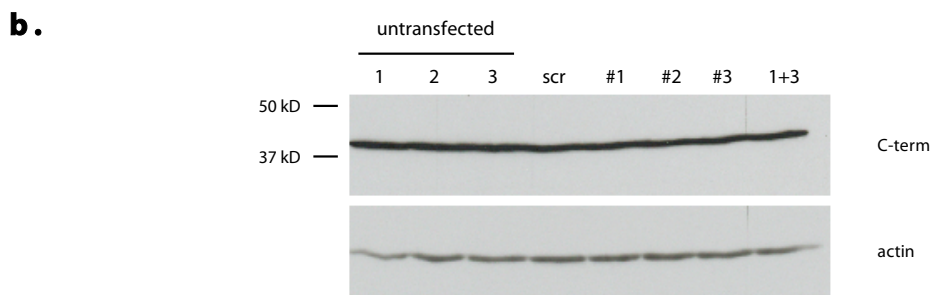
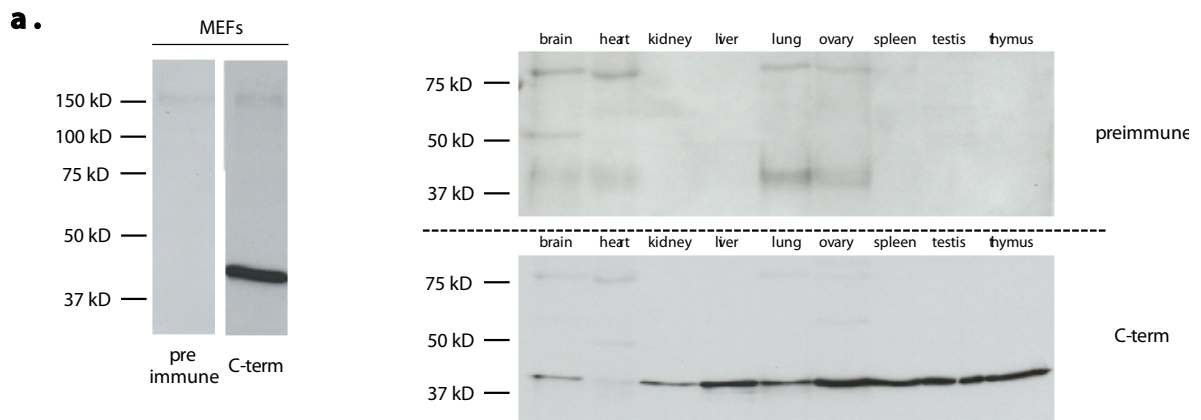


Figure 18. siRNA-mediated knockdown of miwi2 in MEFs. Assessment of the knockdown on the RNA level.

- Positions of siRNAs used for knockdown of miwi2 in MEFs. Knockdown efficiency on the RNA level was assessed through PCR detection of the 1711-e21r product.
- Testing the siRNA-mediated knockdown efficacy on the RNA level. ut: untransfected MEF; scr: scrambled siRNA control.
- RT-PCR and qRT-PCR analysis of the siRNA#3-mediated knockdown of miwi2 in MEFs, which results in the decrease of miwi2 mRNA levels in MEFs.



e.

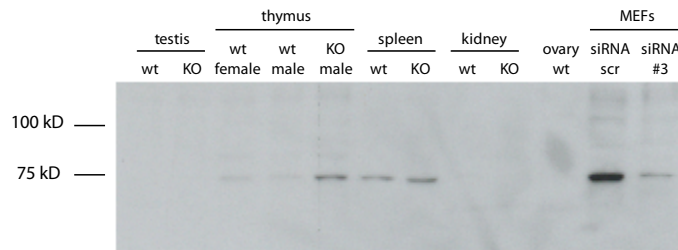


Figure 19. Detection of Miwi2 in adult tissues. The assessment of siRNA-mediated knockdown of Miwi2 in MEFs on the protein level.

- Western blot detection of Miwi2 in MEFs and adult mouse tissues using custom-made C-terminal antibody.
- The levels of the protein, detected by the custom-made C-terminal antibody, are not decreased upon siRNA-mediated knockdown.
- Western blot detection of Miwi2 in MEFs and adult mouse tissues (Abcam antibody).
- siRNA-mediated knockdown results in a moderate decrease of Miwi2 protein levels, as detected by the Abcam antibody.
- Up-regulation of 3'-half transcripts in the knockout tissues is accompanied by slight protein up-regulation, but only in tissues where it seems to be expressed in the wt. Abcam antibody was used. Spleen and kidney are from male mice.

In testis, the full-length protein was also not detected, consistent with its absence, or low-level expression. The nuclear and cytoplasmic fraction from the testicular single-cell suspension were not analyzed using the C-terminal antibodies.

Since the 3'-half transcripts have elevated levels in the heterozygous and knockout tissues of the mice created by Kuramochi-Miyagawa *et al.* (Fig. 14), I checked if a similar change could be observed on the protein level. The up-regulation could be observed in thymus and less in spleen, but not in kidney or testis. If the ~75 kD band truly corresponds to Miwi2, then the tissues where the protein cannot be detected seem to not translate the mRNA, independently of its levels (Fig. 19e).

Using the N-terminal antibody for the Western blot analysis confirmed the absence of the full-length protein from the somatic tissues. In the testis, where the full-length transcript is expressed and associates with the polysomes, a ~100 kD band

can be detected in the nuclear fraction of the total protein extract. This allows the possibility that the protein is still expressed, either in the limited cell population or at the very low levels, despite the fact that it cannot be detected by fluorescence in the post-birth tissues upon insertion of the GFP tag. Similarly, while this antibody was successfully used for immunoprecipitation (IP) of Miwi2 from the embryonic gonadal extracts (Aravin *et al. Mol. Cell* 2008), the protein was not detected by IP from the adult testis extracts. The presence and localization of the full Miwi2 protein in the testicular tissue was further analyzed by immunohistochemistry, as described later.

Detection of the putative alternative, C-terminal Miwi2 variant in the tissues using two antibodies with epitopes in the C-terminal region of the protein gave contradictory results. The custom-made C-terminal antibody detected a protein of the expected size, but its levels did not change upon siRNA-mediated knockdown in MEFs. This result could be explained only by the very high stability of the protein. Upon IP, a protein of ~45 kD can be detected on the silver gel, but the mass spectrometry analysis showed it corresponds to an hnRNP protein, and Miwi2 peptides were not detected (data not shown). Together, this data suggest that the detected ~45 kD protein does not correspond to Miwi2.

Using the commercially available antibody, a ~75 kD protein was detected in MEFs and some tissues, a size not corresponding to the expected size of the protein arising from the 3' transcripts. The difference of ~30 kD could potentially be explained by a post-transcriptional modification of Miwi2 by ubiquitination or SUMOylation. This has been previously described for proliferating cell nuclear antigen (PCNA), and is required for its role in translesion DNA synthesis (Hoegge *et al. Nature* 2002). Miwi2 could not be detected upon IP using this antibody. However, surprisingly, a moderate decrease in the protein levels can be observed upon siRNA-mediated knockdown of miwi2. Additional siRNA-mediated knockdowns leading to the down-regulation of the detected protein, and/or an additional antibody specific for the C-terminus of Miwi2, would be essential to confirm that the detected ~75 kD band truly corresponds to Miwi2 protein.

From these results, it is not clear if the ubiquitously expressed 3' transcripts give rise to a shorter Miwi2 protein variant. If the ~75 kD protein, the levels of which are down-regulated upon the knockdown and that can be detected in some adult tissues, truly corresponds to Miwi2, this would suggest that an additional level of regulation of miwi2 expression exists in certain tissues (e.g. testis and kidney), and not in others (e.g. spleen and thymus), preventing the translation of the transcript. A quick way to address this issue could be the analysis of the association of the 3' transcripts with polysomes in different tissues. A translation analysis method based on sequencing the mRNA fragments protected by the ribosome, which enables not only identifying translated mRNAs but also the identification of the codon used for the initiation of translation, has recently been described (Ingolia *et al. Science* 2009). This method could be suitable for analysis of miwi2 expression, as it could not only address the question in which, if any, tissues does the transcript associate with the polysomes, but also which, if any, AUG is used for the initiation of translation.

The most straightforward way to address the expression and function of the putative alternative form of Miwi2 would be to create a knockout mouse model affecting the 3' region of the gene. Alternatively, a knockout affecting only the genomic regions containing the putative promoter that gives rise to the 3' transcript could also address this issue, and it would potentially not affect the expression of the full-length, wild type Miwi2.

It is worth noting here that, if the observed expression of Miwi2 in spleen and thymus is correct, it would resemble the expression patterns described for alternative forms of DNA methyltransferases (Chen *et al. J Biol Chem* 2002). As mentioned earlier, Miwi1 could also be detected on the RNA level in spleen and thymus (Kuramochi-Miyagawa *et al. Mech Dev* 2001).

5. II. 5. Immuno-detection of Miwi2 in tissues.

In order to further analyze the expression patterns of Miwi2, I performed immunohistochemistry (IHC) and immunofluorescence (IF) experiments using antibodies directed against N-terminal and C-terminal parts of Miwi2.

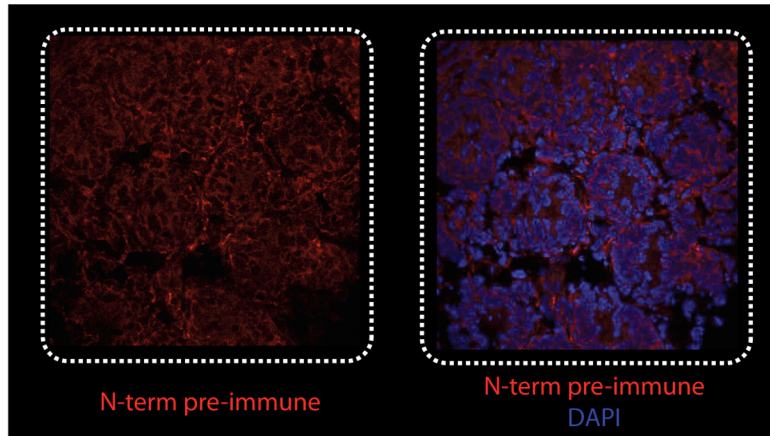
First, in order to check if the full-length Miwi2 can still be detected in testicular cells despite its low levels, I stained testis from different developmental time points with the N-terminal antibody. In early stages of postnatal development the protein can be detected in the nuclei of spermatocytes (Fig. 19a). In the adult organ, the protein is detected in the nuclei of spermatogonia, as well as in a discrete peripheral nuclear sub-domain in spermatocytes resembling the sex body (Fig. 19b and c). Such staining pattern might be consistent with the Western blot analysis on the adult tissue, as the protein is only present in the nuclei of a limited number of cells, possibly explaining its low levels (see Fig. 17b).

Sex body, or XY body, is a nuclear territory formed in spermatocytes during meiotic prophase I (Solari, *Rev. Cytol.* 1974; Turner, *Development* 2007). It contains sex chromosomes, which get sequestered to the nuclear periphery at the transition between zygotene and pachytene stages of prophase I as a result of a big difference in size between chromosomes X and Y and their consequent inability to synapse over the entire length of the chromosome arms. Partial synapsis of chromosomal regions is often a consequence of non-homologous pairing, which is potentially deleterious as it can lead to unequal partitioning of genetic material into the mature gametes. Sex chromosomes in male gonads are therefore recognized at the beginning of the pachytene stage of prophase I by DNA damage repair proteins, such as BRCA1 (Turner *et al. Curr. Biol.* 2004). BRCA1 recruits the ATR protein kinase, which phosphorylates histone H2AX at serine 139, leading to accumulation of this histone variant, referred to as γ H2AX, along the unsynapsed arms of sex chromosomes (Turner *et al. Curr. Biol.* 2004; Bellani *et al. J. Cell Sci.* 2005; Turner *et al. Nat. Genet.* 2005). This leads to further accumulation of silencing histone marks, such as methylation of histone 3 on lysine 9 (H3K9), and transcriptional silencing of chromosomes X and Y (Hoyer-Fender, *Cytogenet. Genome Res.* 2003).

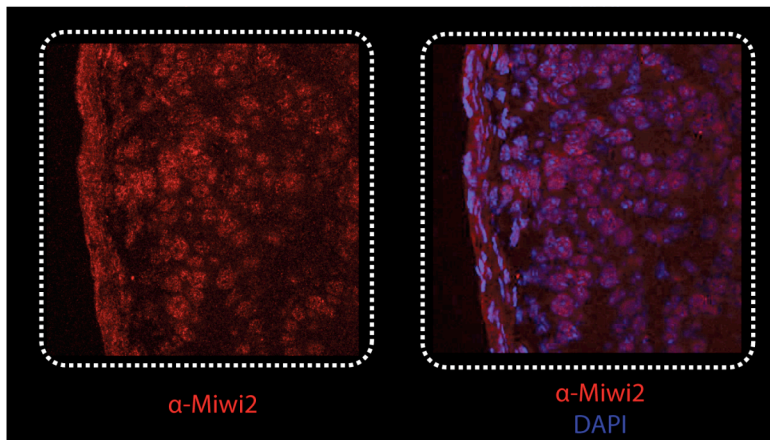
This process, called meiotic sex chromosome inactivation (MSCI), is a manifestation of a more general process, meiotic silencing of unsynapsed chromatin (MSUC) (Turner *et al. Nat. Genet.* 2005; Baarends *et al. Mol. Cell. Biol.* 2005; Schimenti. *Nat. Genet.* 2005). MSUC is responsible for recognizing and transcriptionally silencing any

a.

i.



ii.



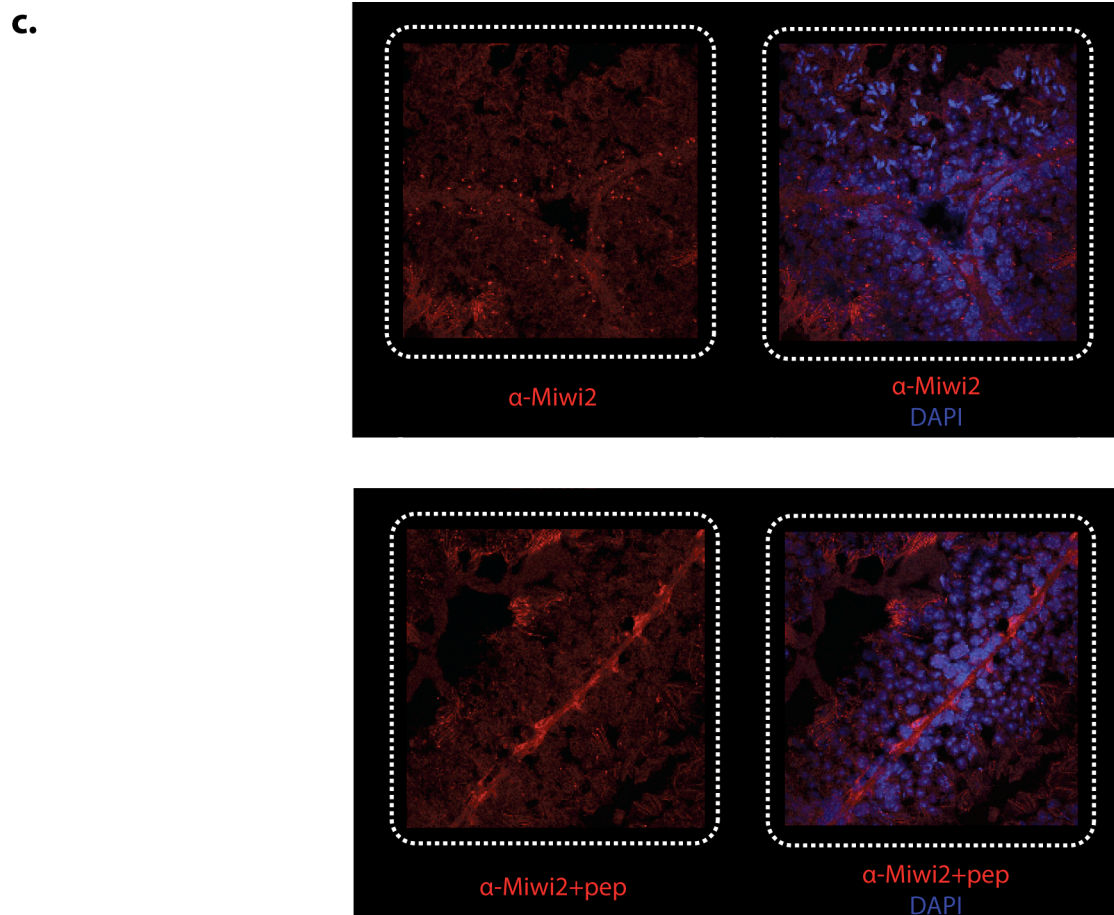
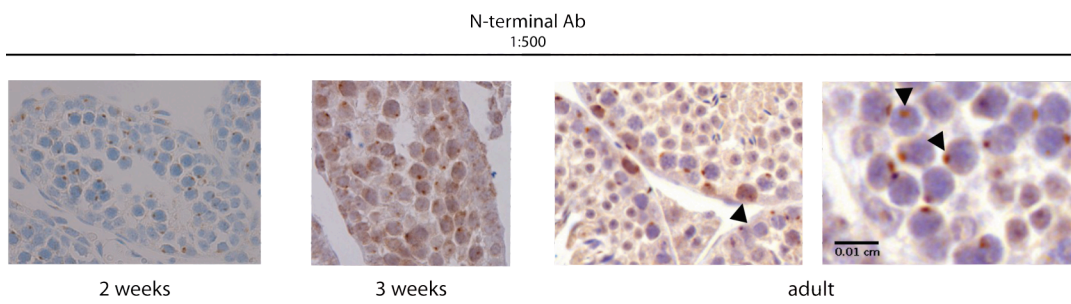
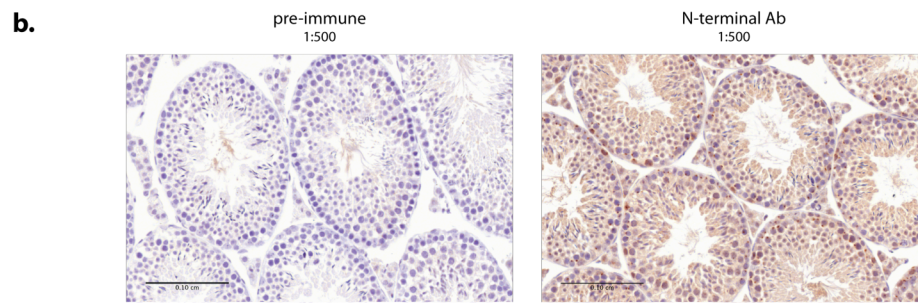


Figure 19. Immunohistochemical detection of Miwi2 in post-birth testicular tissue using the custom-made N-terminal antibody.

- a.** Miwi2 is detected in the nucleoplasm of premeiotic cells. Shown here is the staining of the tissue from 10 days old mice.
- b.** In adult tissues, Miwi2 is detected in nuclei of spermatogonia, and a discrete nuclear sub-domain of meiotic spermatocytes resembling sex body.
- c.** Blocking the N-terminal antibody with the epitope-specific peptide leads to the disappearance of the signal in the sex body-like foci.

region of partial chromosomal synapsis in both sexes, which can arise as a result of events such as translocation. The two processes share the mechanistic core, and have as a consequence sequestration of all unsynapsed and transcriptionally silenced regions to nuclear periphery, where they form the structure called the sex body.

Failure to transcriptionally silence unsynapsed chromosomal regions leads to detection of unrepaired double-stranded breaks, induced at the beginning of meiosis, by checkpoints at the pachytene stage and arrest of spermatogenesis, leading to male sterility.

A similar process, meiotic silencing by unpaired DNA (MSUD), occurs in the fungus *Neurospora crassa* (Shiu *et al. Cell* 2001). This mechanism is capable of examining the entire genome in search of regions of chromosomes that cannot fully pair. Unsynapsed chromosomal regions, and all the regions homologous to them, are then silenced through a mechanism that requires RNA-dependent RNA polymerase *sad-1*, which converts the transcript from the unpaired DNA into dsRNA (Shiu *et al. Cell* 2001), and the Argonaute-like protein Sms-2 (Lee *et al. Genetics* 2003). This mechanism is therefore thought to resemble RNA silencing mechanisms, even though the small RNAs arising from these dsRNAs have not been identified. MSUD is thought to be responsible for protecting the genome from the deleterious consequences of transposon activity, as a transposon that jumps into a new location in the genome often does not have a complementary sequence on the homologous chromosome to synapse with. MSUD would therefore lead to

silencing of the activated transposon, as well as all the other transposons in the genome homologous to it.

Based on the apparent localization of Miwi2 in the sex body, and RNA silencing mechanisms being involved in similar processes of transcriptional silencing in other organisms, I hypothesized that Miwi2 could be involved in the process of meiotic sex chromosome inactivation.

5.II. 6. Miwi2 in the sex body: co-localization with markers.

In order to confirm the sex body localization of Miwi2, I co-stained the testicular tissue using the N-terminal antibody, and antibodies against known sex body markers – BRCA1 and γ H2AX (Fig. 20a and b). I also performed co-immuno-FISH (fluorescent *in situ* hybridization) experiments, in order to check for the co-localization of Miwi2 with sex chromosomes (Fig. 20c).

Co-immunohistochemistry showed co-localization of Miwi2 with BRCA1, the early sex body marker (Fig. 20a). Miwi2, however, did not co-localize with γ H2AX, the marker of the fully formed sex body at the stage where the chromosomes are transcriptionally silenced (Fig. 20b). Co-immuno-FISH experiments showed co-localization of Miwi2 and chromosome X (Fig. 20c).

As mentioned, MSUC is a manifestation of a general mechanism of silencing of unsynapsed chromatin. I therefore wanted to check if Miwi2 could also be detected on unsynapsed chromosomal arms in a mouse model with an autosomal translocation. Such a model, in which the mice bear a reciprocal translocation between chromosomes 16 and 17, was described by Homolka *et al.* (Homolka *et al. Genome Res.* 2007), and is referred to as the T43 model. This model was used in studies of MSUC, and it was shown that unsynapsed regions of chromosomes 16 and 17 reside in the nucleus in the close vicinity to the sex body and undergo transcriptional silencing (Homolka *et al. Genome Res.* 2007). BRCA1 and γ H2AX could be detected on unsynapsed regions.

Staining the testis from the T43 model mice with the N-terminal antibody showed structures in close proximity to the sex body resembling unsynapsed chromosomal arms (Fig. 21a). Co-localization of Miwi2 and BRCA1 on these structures could also be observed (Fig. 21b). I tried to confirm that the observed structures truly correspond to chromosomes 16 and 17, but the *in situ* detection did not work.

These results suggest the possible involvement of Miwi2 in transcriptional silencing of unsynapsed chromosomal regions in meiotic prophase I.

Co-localization of Miwi2 with BRCA1, but not with γ H2AX, suggest that it could be involved in the early stages of silencing. Miwi2 could be detected in the subnuclear foci prior to onset of meiosis and MSCI/MSUC. I therefore hypothesized that Miwi2

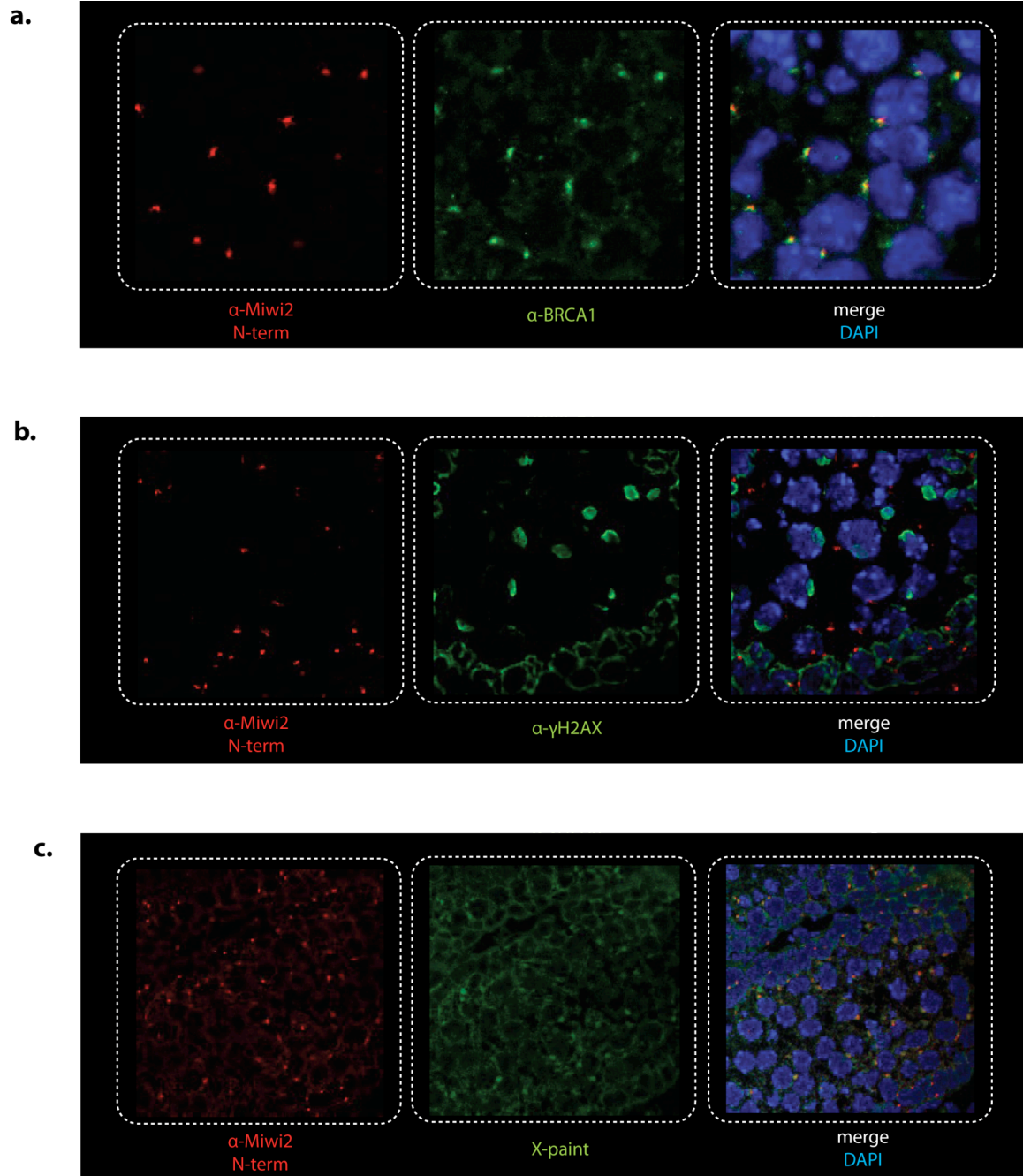


Figure 20. Localization of Miwi2 in the sex body of meiotic spermatocytes. Miwi2 co-localizes with BRCA1 and chromosome X, but not with γ H2AX.

- a.** Co-staining of testicular tissue for Miwi2 and BRCA1.
- b.** Co-staining of testicular tissue for Miwi2 and γ H2AX.
- c.** Co-staining of testicular tissue for Miwi2 and chromosome X.

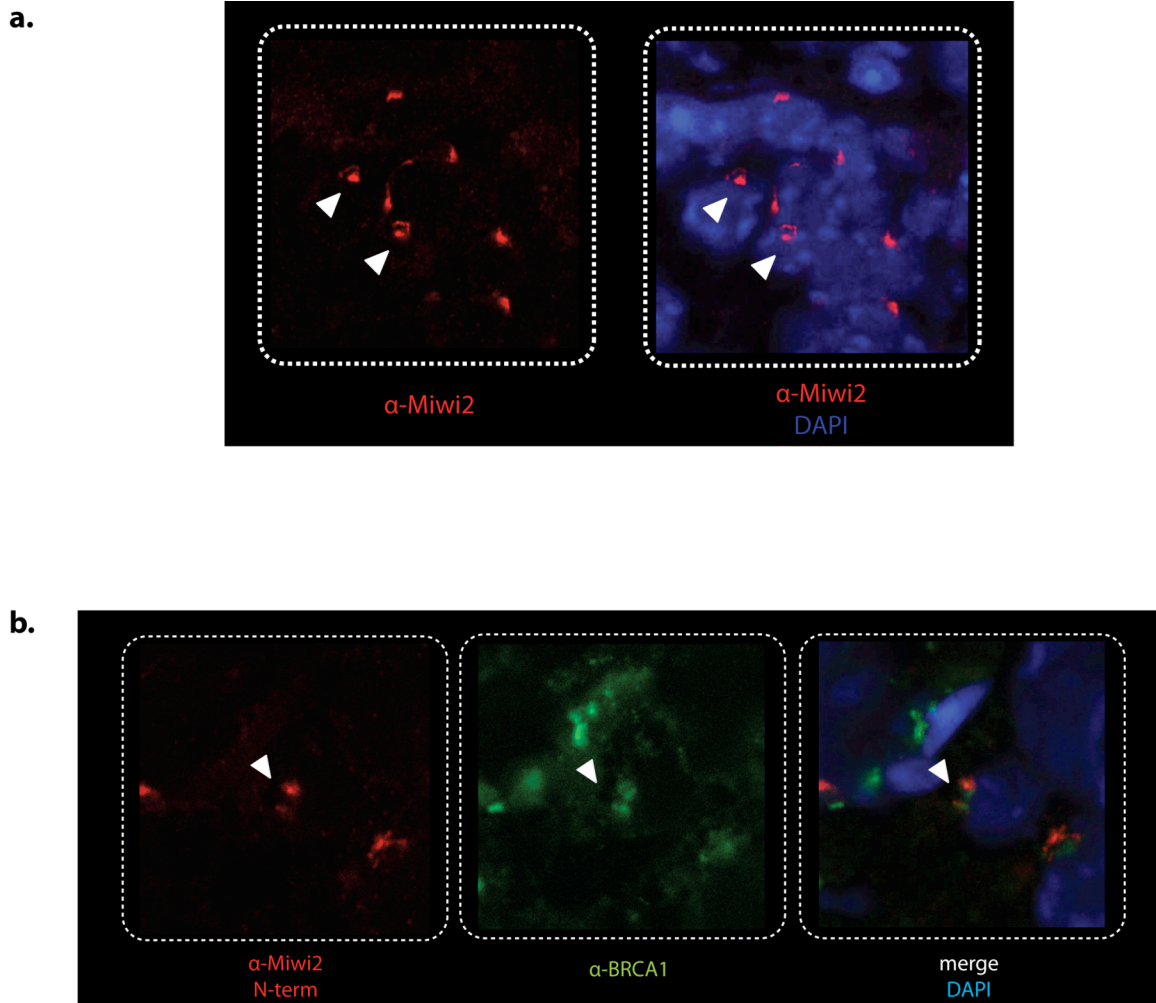


Figure 21. Immunofluorescent staining on T43 autosomal translocation model using N-term antibody.

- a.** Localization of Miwi2 in T43+ autosomal translocation model. Structures resembling unsynapsed chromosomal arms can be observed in close proximity of the sex body.
- b.** Miwi2 co-localizes with BRCA1 in sex body and on unsynapsed arms of autosomes with translocations.

might be involved in the initiation of MSCI/MSUC, possibly through the recruitment of BRCA1. A way to test this is to analyze the localization of BRCA1 in the Miwi2 knockout mice. I only analyzed the localization of BRCA1 in the knockout

model created by Carmell *et al.*, but these experiments also never gave a conclusive result. Interestingly, Carmell *et al.* observed a nuclear territory in the *miwi2* knockout spermatocytes that resembles sex body, but does not stain positive for γ H2AX (Carmell *et al. Dev. Cell* 2007).

The localization patterns of γ H2AX are abnormal in the early meiotic cells from these mice, reflecting the defects in spermatocytes' morphology (Carmell *et al. Dev. Cell* 2007). This is likely the consequence of the transposon activity (Carmell *et al. Dev. Cell* 2007), but similar phenotype and the arrest of spermatogenesis at the same stage of prophase I have been observed for mutants of proteins involved in MSCI/MSUC.

Another way to functionally address the failure of transcriptional silencing of unsynapsed regions is to analyze the expression of genes that should undergo silencing, such as majority of genes on the sex chromosomes, on the transcriptional level. This question can also be addressed by staining the testicular tissue for RNA polymerase II, as it is excluded from the territory of the sex body once the silencing becomes effective, from the late spermatocyte stage onwards. Unfortunately, the phenotype of the *miwi2* knockout mice prevents the assessment of these issues, as the spermatogenesis only progresses to the early spermatocyte stage, and the late spermatocytes are absent from the tissues.

The mechanism of initiation of meiotic silencing of sex chromosome is elusive to date. The post-meiotic fate of sex chromosomes and the possible marks left upon the sex chromosomes during MSCI, presumptively for the recognition in the next generation, have been extensively discussed. Additionally, there is speculation of the involvement of the marks left on the X chromosome in recognition of paternal X during initiation of somatic X inactivation in the early embryos, even though Xist does not play a role in MSCI/MSUC (Namekawa *et al. Curr. Biol.* 2006). It is worth noting here that one-third of human chromosome X consists of LINE1 retroelements, and they have been proposed to be involved in the process of somatic X chromosome inactivation (XCI) in females (Bailey *et al. PNAS* 2000; Ross *et al. Nature* 2005). In mouse, Miwi2 associated with the piRNAs from the LINE1 element (Aravin *et al. Mol Cell* 2008). It is therefore a tempting speculation that Miwi2

and its associated piRNAs are somehow involved in XCI and MSCI/MSUC, possibly through the establishment of DNA methylation aided by Miwi2 and LINE1-derived piRNAs on the LINE1 elements residing on chromosome X.

In order to confirm the localization of Miwi2 in the sex body, I next used three different antibodies directed against the C terminus of Miwi2, our custom-made antibody, the commercial antibody, and an additional antibody obtained from the Hannon lab (Aravin *et al. Mol Cell* 2008). Only the third antibody showed sex body staining pattern, whereas our custom-made antibody and the commercial antibody gave no signal (data not shown). This raised a question of the specificity of the signal in the sex body.

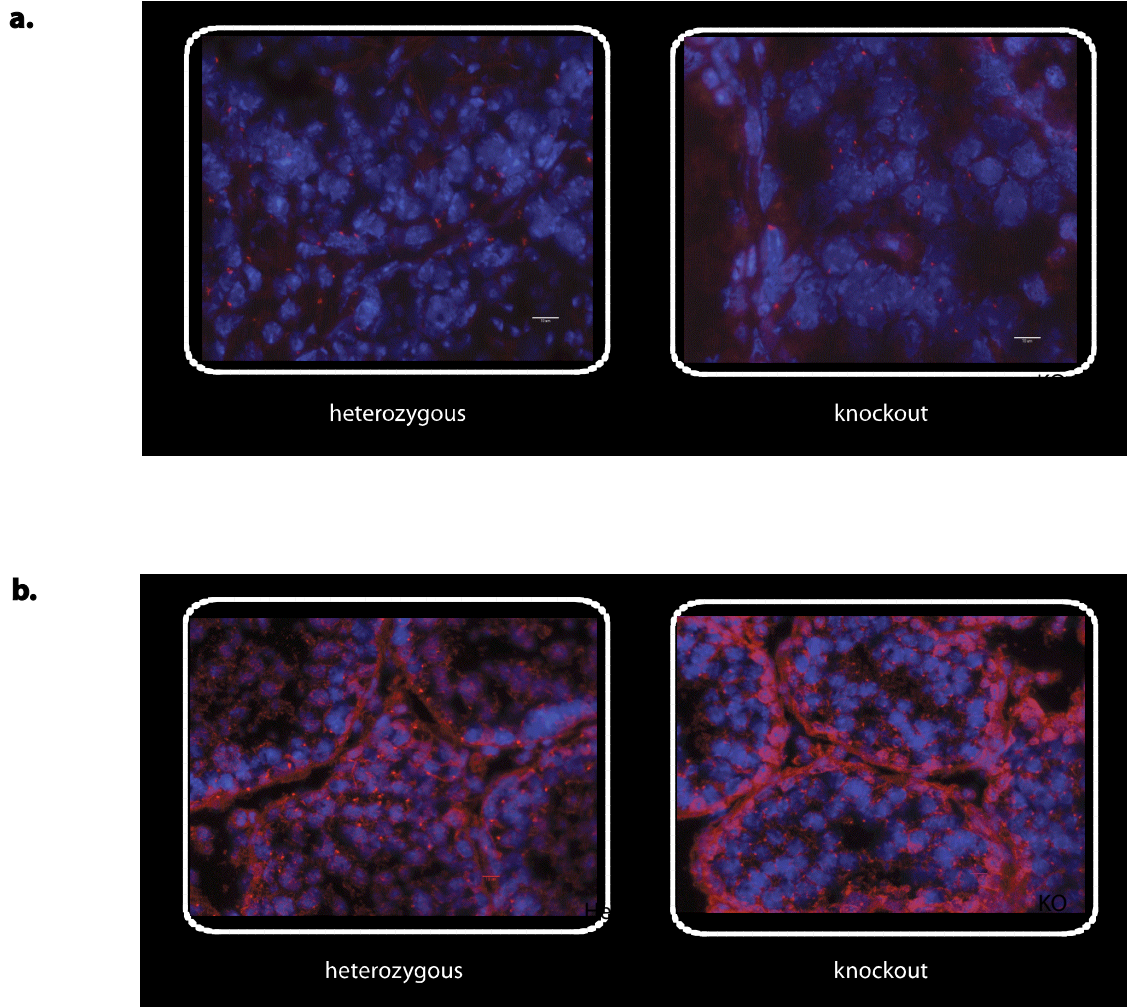
5. II.7. N-terminal antibody specificity.

N-terminally GFP-tagged Miwi2 could not be detected in the postnatal spermatogenic cells through fluorescence. However, as described, our N-terminal antibody staining suggests that the protein could still be present, with a peculiar staining pattern showing prominent localization of Miwi2 to the sex body of meiotic spermatocytes. This staining could be reproduced with only one of the three tested C-terminal antibodies.

These discrepancies raised the question of the specificity of the N-terminal antibody. As shown in Fig. 19, the antibody gives a staining pattern clearly distinct from the pre-immune serum one, and the signal can be blocked upon incubation of the antibody with the epitope-specific peptide. This is consistent with the Western blot (Fig. 17b).

Testicular tissue from the knockout mice does not contain the full-length protein, making it the best control for the antibody specificity. However, in the knockout created by Carmell *et al.*, the 5' end of miwi2 is still transcribed. It is therefore not possible to exclude the existence of a protein having the epitope for the N-terminal antibody in the tissue from these mice. Staining the testicular tissue from

this knockout model showed no difference between the heterozygote and the knockout (Fig. 22a).



α-Miwi2 N-terminal
DAPI

F

figure 22. Immunofluorescent staining of heterozygous and KO testis tissue from both models described in the literature using the N-term antibody.

The signal in the sex body is detected in the knockout tissue, suggesting it's unspecific.

- a.** miwi2 Het and KO testicular tissue from the Carmell *et al.* model.
- b.** miwi2 Het and KO testicular tissue from the Kuramochi-Miyagawa *et al.* model.

On the other hand, the epitope for our N-terminal antibody is deleted in the knockout mice produced by Kuramochi-Miyagawa *et al.*, and staining the testicular tissue from these mice should therefore not generate a signal.

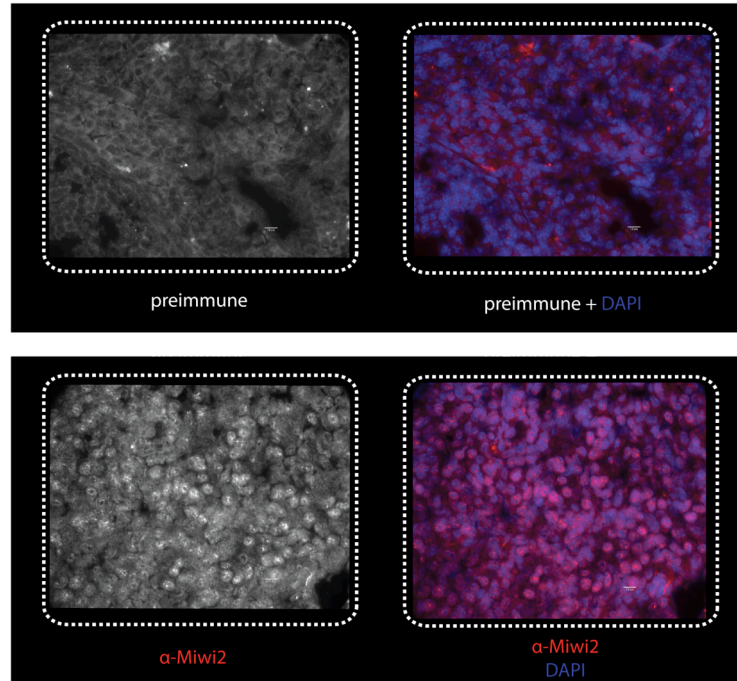
However, there was no difference between the staining patterns in the heterozygous and knockout tissue (Fig. 22b). This strongly suggested that the N-terminal antibody is cross-reacting with another protein in tissue stainings, and that the observed signal in the sex body is not specific. This result, therefore, spoke against localization of Miwi2 in the sex body and its involvement in MSCI/MSUC.

The presence of the signal in the miwi2 knockout mice could only be explained by the possibility that either Miwi2 protein or mRNA is inherited from the heterozygous parents. *Drosophila* Piwis in complexes with piRNAs get maternally deposited into early embryos, and the ping-pong cycle cannot be initiated in the absence of maternally inherited piRNAs, causing hybrid dysgenesis (Malone and Brennecke, *unpublished*). The putatively inherited form of Miwi2 should be capable of associating with unsynapsed chromatin independently of the endogenously produced one, as in the knockouts the presumptive protein still localizes to the sex body. However, if the observed failure of sex body formation in the knockout mice created by Carmell *et al.* is correct, the inherited form of Miwi2 could not fully exert its function without the presence and/or activity of the endogenous protein.

5. II. 8. Detection of the putative alternative form of Miwi2 in tissues and cells using the C-terminal antibody.

In order to try and determine the localization of the putative C-terminal form of Miwi2, I performed immunohistochemistry. The results were difficult to assess, as the two antibodies behave very differently in the Western blot analysis, and it is not clear which, if any, of the two is specific. The custom-made antibody showed similar nuclear localization pattern in all tissues examined (Fig. 23; examples for ovary and kidney). The commercial antibody did not give any distinct signal above background; in particular, no difference could be observed between the tissues where the protein could not be detected on the Western blot (testis, kidney), and the tissues in which the ~75 kD protein was detected (spleen, thymus) (data not shown).

a. Ovary.



b. Kidney.

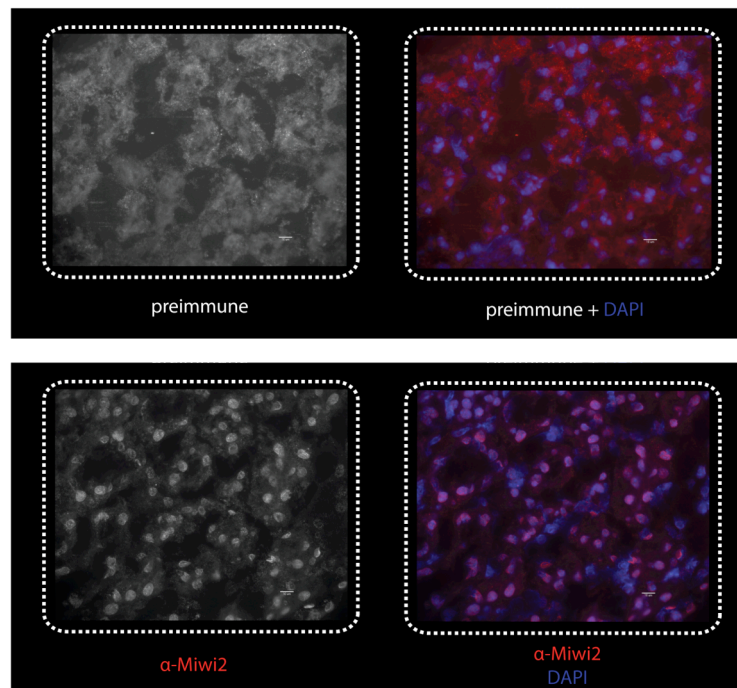


Figure 23. Immunodetection of Miwi2 in adult tissues using the custom-made C-terminal antibody.

- a.** Staining of adult ovary.
- b.** Staining of adult male kidney.

Both antibodies, however, gave the same signal when used for immunocytochemistry on MEFs, which express only the C-terminal version of the protein. Miwi2 could be detected in the nuclei of some cells (marked by arrows), as well as in cytoplasmic granules (marked by the arrowhead) (Fig. 24). In order to check if these granules are P-bodies, I stained the cells using two different antibodies for a P-body marker, GW182. However, none of the antibodies worked well for immunocytochemistry.

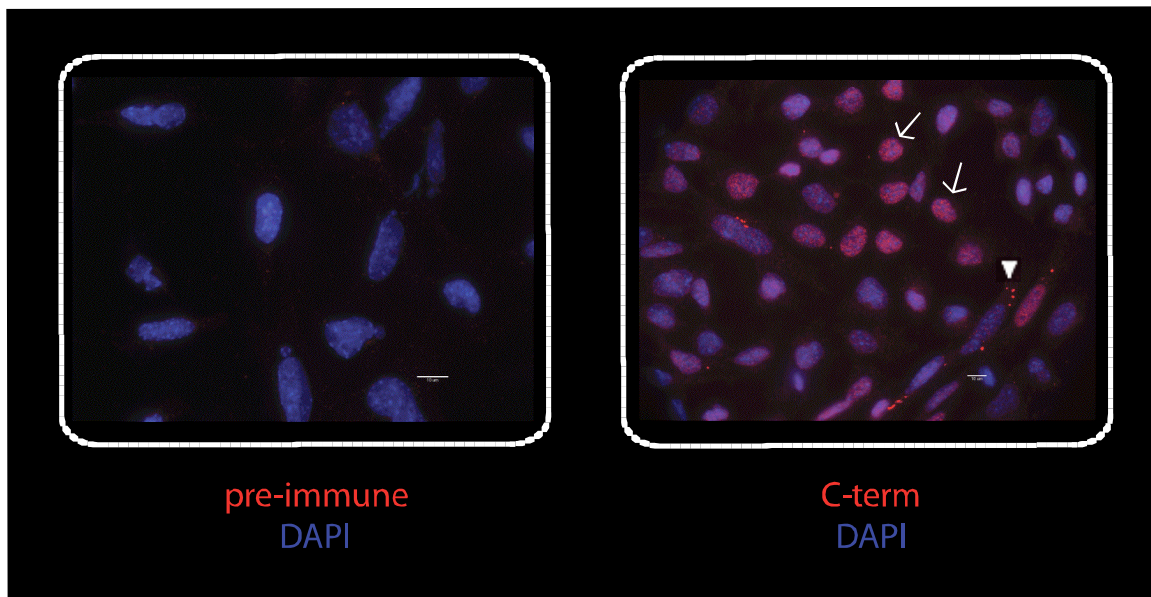


Figure 24. Immunodetection of Miwi2 in MEFs using the custom-made C-terminal antibody.

The signal is detected in the nuclei of some cells (marked by arrows), and cytoplasmic bodies resembling the P-bodies in others.

Specific expression in only a few cells in the heterogenous population of MEFs suggests that the expression of C-terminal variant of Miwi2 in MEFs might be cell cycle-specific. These questions were not addressed at this point, but could be tested by using markers specific for different stages of the cell cycle, as well as by performing cell cycle synchronization experiments.

In summary, as expected from the Western blot analysis, the IF experiments did not give a conclusive result. The staining pattern using the two C-terminal antibodies

was the same only in MEFs, where the protein was detected in the nucleoplasm and the cytoplasmic granules. Due to the inconsistent results of the Western blot analysis, and the uncertainty of the presence of the protein, it is not certain that this result is truthful.

5. III. A possible alternative function of Miwi2 in regulation of imprinting.

In the adult mouse, the only sequences subjected to the long-term stable silencing based on DNA methylation are transposable elements, imprinted genes and genes associated with pluripotency (Reik, *Nature* 2007). As Miwi2 has a role in establishing DNA methylation of transposable elements in the germ line, which occurs at the same time as the methylation on the regulatory regions of imprinted genes, I wanted to investigate if it could also have a role in controlling the expression of imprinted genes.

I analyzed the mRNA expression of the two genes, Igf2 and H19, by qRT-PCR in both somatic and germ line tissues of wild type, heterozygous and knockout adult male *miwi2* mice. Igf2 and H19 are closely linked on the mouse chromosome 7 and show reciprocal imprinting patterns. Non-coding H19 RNA is maternally expressed due to developmental silencing of the paternal promoter, and the protein-coding Igf2 is paternally expressed due to transcriptional insulation of the maternal promoters (Kaffer *et al. Mol. Cell. Biol.* 2001). Additionally, a common *cis*-acting element located upstream of the H19 promoter controls expression of both loci (Thorvaldsen *et al. Genes Dev.* 1998; Kaffer *et al. Genes Dev.* 2000). Loss of imprinting at this locus is associated with tumors and developmental diseases (Paulsen and Ferguson-Smith, *J. Pathol.* 2001).

DNA methylation of the main DMR of this locus, located between the two genes, was at present not analyzed.

I first analyzed the expression of Igf2 and H19 in the adult testis. Both heterozygous and knockout mice used for these experiments were progeny from

the crosses between heterozygous *miwi2* parents. It is important to point out that the cellular composition of the adult knockout testicular tissue, due to the arrest of spermatogenesis in meiotic prophase I, corresponds to the pre-meiotic tissue. Therefore, the levels of the transcripts from the knockout tissue had to be normalized and compared to their levels in the tissue from an 11-13 days old mouse, whereas the levels in the heterozygous tissue were compared to the levels in the adult, as the heterozygous tissue develops normally.

Analysis of Igf2 expression in testis showed that the levels of Igf2 in both heterozygous *miwi2* models were very similar to the adult wild type (Fig. 25a). The expression of Igf2 in both knockout models was the same as in the 11 days-old wild type tissue (Fig. 25a).

H19 non-coding RNA was reported to be expressed from the maternal allele in several fetal tissues and mostly absent from adult tissues (Leighton *et al. Genes Dev.* 1995). Consistent with this, the levels of this transcript decrease drastically between 11 and 35 old mice. The levels in the heterozygous tissue differed little from the levels in the adult wild type tissue; they were somewhat higher in the heterozygous tissue from the Kuramochi-Miyagawa *et al. miwi2* model (Fig. 25b). However, the levels of H19 transcript in the knockout tissue were drastically lower from the levels in the corresponding 11-days old wild type tissue (Fig. 25b). Namely, two different primer pairs showed ~10-fold down-regulation of H19 in both *miwi2* knockout models (a representative example is shown in Fig. 25b).

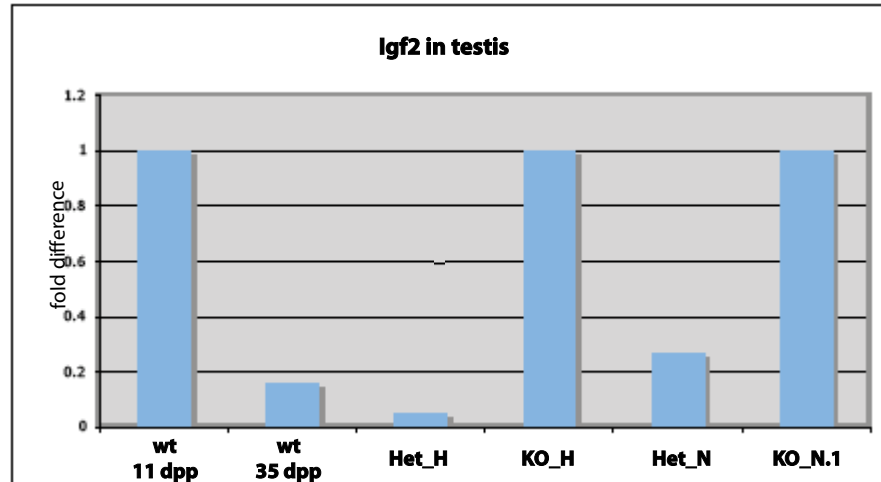
Next, I analyzed the expression of Igf2 and H19 in adult male spleen from the Kuramochi-Miyagawa *et al. miwi2* model. Adult somatic tissues have not been described to have any phenotype in the *miwi2* knockout mice.

qRT-PCR analysis of total mRNA from adult male spleen showed almost equal levels of the Igf2 transcript in the wild type and heterozygous tissue (Fig. 26a). In one of the knockout samples the levels were ~2-fold higher than in the heterozygote, and ~6-fold in the other (Fig. 26a).

In contrast, the expression of H19 is strongly upregulated in the adult male spleen from the *Miwi2* knockout mice (Fig. 26b). Two additional primer pairs confirmed

this observation, with the observed levels of H19 being at least 40-fold higher in the knockout compared to the wild type and heterozygous tissue.

a.



b.

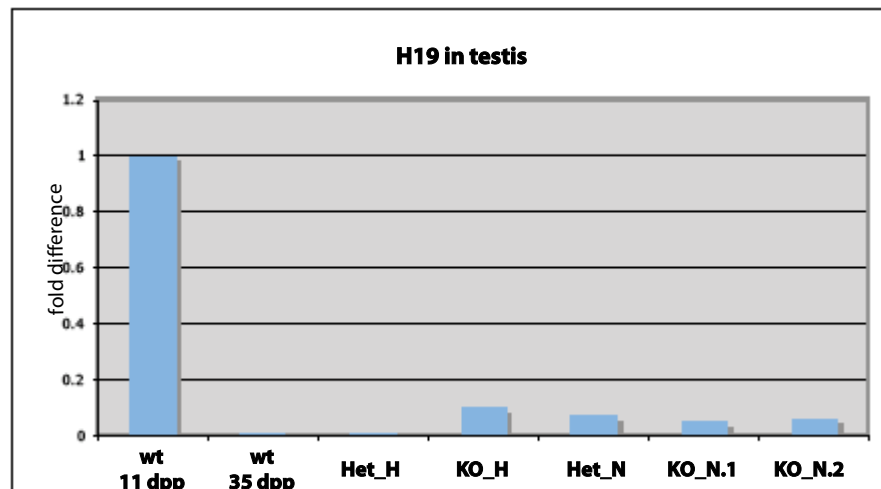
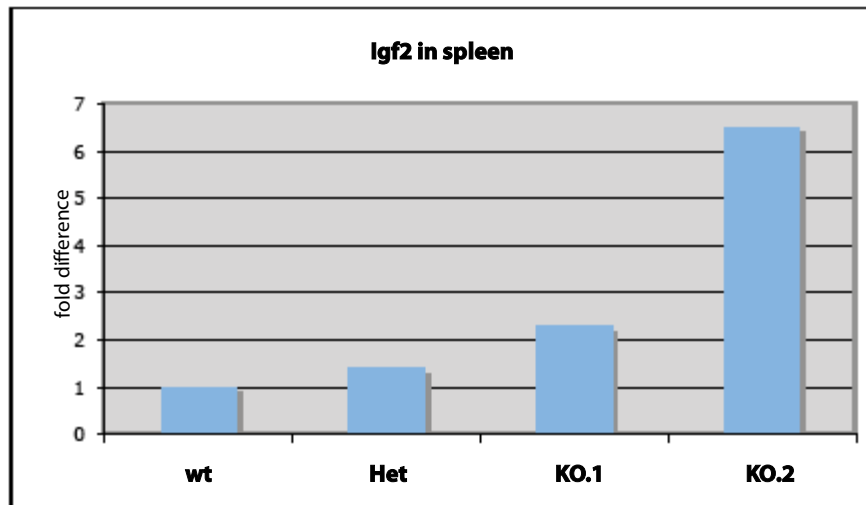


Figure 25. The expression of the maternally imprinted Igf2 gene is not affected in the miwi2 knockout testis (a). In contrast, expression of the paternally imprinted H19 is ~10-fold lower than in the corresponding wild-type tissue (b).

testis Het_N; testis KO_N: adult heterozygous and knockout mice created by Kuramochi-Miyagawa et al.; testis Het_H; testis KO_H: adult heterozygous and knockout mice created by Carmell et al. N.1 and N.2 represent two different knockout samples. All heterozygous and knockout samples are from adult mice.

a.



b.

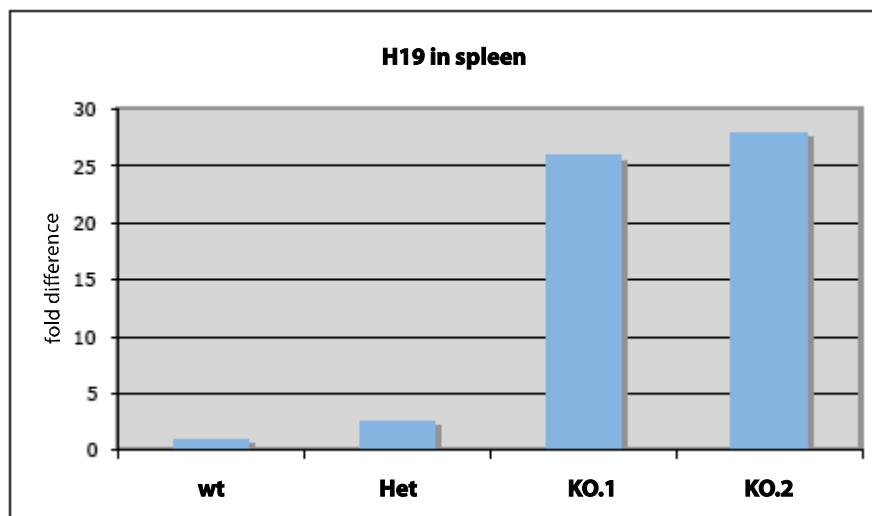


Figure 26. Analysis of the mRNA expression from the H19/Igf2 imprinted locus in adult male spleen.

- a. The expression of Igf2 is slightly up-regulated in the spleen of Miwi knockout mice.
- b. The strong upregulation of H19 is observed in spleen of the Miwi2 knockout mice.

This data suggest that full-length Miwi2 is not involved in establishing imprints of Igf2 in the male germ line, which is not unexpected as it is maternally imprinted. Analysis of its expression in the ovary from the knockout mice would show if

there's any impact of the absence of Miwi2 on maternal imprinting of Igf2. It is at present not clear which, if any, protein form of Miwi2 is expressed in the ovary.

The effect of the absence of the full-length Miwi2 on the control of Igf2 expression in adult spleen was also mild. If present, this effect might be stochastic, as the two different samples from the knockout mice show different levels of up-regulation. This has been previously observed in the imprinting defects caused by the lack of Dnmt3L (Arnaud *et al. Hum Mol Genet* 2006). More samples would have to be analyzed in order to investigate this observation in more detail.

On the other hand, there seems to be an impact of the absence of Miwi2 on the levels of the non-coding H19 transcript in testis. The observed down-regulation of the transcript levels is not the immediately expected result, if Miwi2 would be directly involved in the establishment of DNA methylation, and without the analysis of the DNA methylation status at the DMRs in the knockout mice, it cannot be concluded that Miwi2 directly regulates the expression of the H19 non-coding RNA. More paternally imprinted loci would have to be analyzed in order to see if there is a general effect of the absence of Miwi2. An additional control would be pre-meiotic heterozygous tissue, in order to exclude any effects of the observed increased transcription of the 3' half of miwi2 on the expression of the full-length form.

Surprisingly, a strong up-regulation of the levels of the H19 mRNA was observed in the spleen from the knockout males. The full-length Miwi2 is not expressed in somatic tissue, and the mice have not been described to have any other phenotype apart from male sterility (Carmell *et al. Dev Cell* 2007). Failure of imprinting of H19 causes several developmental disorders, such as Beckwith-Wiedemann syndrome (BWS), characterized by general and organ overgrowth and tongue enlargement (Catchpoole *et al. J Med Genet* 1997). The knockout mice would have to be inspected for these defects in order to address the phenotypical impact of the observed transcript up-regulation. Somatic imprints are established early in embryonic development, and maintained through the action of DNA methyltransferases. Without the detailed analysis of expression of Miwi2 in all periods of embryonic and post-natal splenic development, the mechanism through which it is involved

in the regulation of H19 expression cannot be anticipated. The involvement of Miwi2 in the maintenance of the imprints is less likely, as it is not expressed in the post-birth somatic tissues. Based on the available data, the only possible explanation for the observed up-regulation of the H19 transcript is the involvement of Miwi2 in the process of imprinting very early in development. Therefore, Miwi2 would either have to be expressed in the short developmental window in very early embryo, or it could be inherited. To my knowledge, neither of this was analyzed to date, and there is no evidence for either of the two speculated possibilities. It is also not clear if the observed up-regulation is the consequence of de-repression of the paternal allele, stronger expression of the maternal allele, or both. Also, the important question of the potential involvement of Miwi2 in the methylation of the regulatory regions at the H19/1gf2 locus should be answered by analysis of its DNA methylation status by bisulphate sequencing or methylation-sensitive Southern blot.

In summary, these results suggest that the full-length Miwi2 might have an impact on the regulation of expression of at least one paternally imprinted gene, H19, both in the male germ line and a somatic tissue, spleen. Surprisingly, the effect of the absence of Miwi2 on the expression of H19 is stronger in the spleen, the tissue in which the full-length protein has never been detected, then in the testis, where Miwi2 has a defined role in regulating the expression of transposable elements. Another interesting question not addressed to date is whether the absence of Miwi2 also has an effect on the expression of transposable elements, which were suggested to be involved in the regulation of imprinted genes, in the somatic tissues. The effect of Miwi2 on the regulation of maternally imprinted genes seems to be slighter in the spleen, and not present in the testis.

Even though this limited analysis suggests that Miwi2 might be involved in the regulation of H19 imprinting, an extended analysis would be required in order to confirm or disprove the role of Miwi2 in the regulation of imprinting.

6. Conclusions.

During my doctoral thesis, I studied the Piwi subfamily of Argonaute proteins in the mouse, the Miwis. The original project, addressing the potential association of small RNAs associated with the Miwis, had to be changed after the discovery of piRNAs associated with Miwi and Mili shortly after (Girard *et al. Nature* 2006; Aravin *et al. Nature* 2006). I therefore focused on Miwi2, primarily on its expression and function in the post-birth tissues.

My results reveal several previously not documented issues about the regulation of expression and function of Miwi2.

Transcriptional analysis suggested the possible existence of complex phenomena regulating the expression of *miwi2*. Despite the apparent lack of phenotype in any other tissues apart from testis, many different non-coding and potentially coding transcripts can be detected in virtually all tissues. Different promoters might be utilized for transcription of different regions within the gene. Particularly interesting are several predicted transcriptional starting sites roughly in the middle of the locus, which might give rise to the ubiquitously expressed 3' transcripts. These transcripts could theoretically be translated into a protein product, as they contain in-frame START codons. However, the analysis using antibodies directed against the C-terminal part of Miwi2 did not provide conclusive results about the presence of the alternative Miwi2 form in the tissues. Creating a knockout mouse in which the activity of the putative promoters would be affected, or the one lacking regions in the 3' part of the gene, could address this issue genetically. Inserting a GFP-tag into the 3' region of *miwi2* could also help answer this question, and would enable determining the sub-cellular localization of the putative alternative Miwi2 form. Addressing the seemingly complex regulation of *miwi2* expression might also provide new insights into its putative alternative functions, and, if true, challenge the dogma about the germ line-specific roles of Piwi proteins.

The main characterized function of Miwi2 is the control of expression of transposable elements in the male embryonic germ line. My result suggest that Miwi2 might have additional functions outside of the germ line, namely in the control of imprinted genes, as the expression of the imprinted non-coding RNA H19 changes in some tissues in the miwi2 knockout mice. These two processes share mechanistic similarities, both depending on DNA methylation. As it was shown that absence of Miwi2 results in decrease of DNA methylation on the regulatory regions of transposons, analyzing the regulatory regions of imprinted gene(s), whose expression is affected by the absence of Miwi2, could confirm or dispute the role for Miwi2 in DNA methylation of these regions.

In that context, the detailed analysis of Miwi2 expression throughout development, particularly during embryogenesis, when the somatic imprints are established, could further clarify this issue, and help confirm or dispute the impact of the absence of Miwi2 on the expression of imprinted genes.

It would also be very interesting to analyze the impact of the absence of Miwi2 on the expression of pluripotency-associated genes, as they are also silenced by DNA methylation in differentiated tissues. Piwis were initially described as the factors affecting division and maintenance of stem cells. In testis, the absence of Miwi2 results in progressive death of germ line stem cells, spermatogonia. My preliminary data, not presented here, show that a gene involved in the maintenance of pluripotency, Oct4, might be downregulated in the Miwi2 knockout mice. To my knowledge, the role of Piwi subfamily in the expression of pluripotency-associated genes was not addressed in detail to date, and it might be the link between the molecular function of Piwis and their role in stem cell maintenance, which has not been clarified yet.

In conclusion, these results posed some intriguing questions about regulation of expression and possible alternative functions of Miwi2, and addressing them could potentially provide new insights into the alternative roles of the Piwi proteins.

7. References.

- Ambros, V., R. C. Lee, et al. (2003). "MicroRNAs and Other Tiny Endogenous RNAs in *C. elegans*." Current Biology **13**(10): 807-818.
- Angell, D. B. S. (1997). "Consistent gene silencing in transgenic plants expressing a replicating potato virus X RNA." The EMBO Journal **16**(12).
- Aravin, A. A., N. M. Naumova, et al. (2001). "Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline." Current Biology **11**(13): 1017-1027.
- Aravin, A. A., M. Lagos-Quintana, et al. (2003). "The Small RNA Profile during *Drosophila melanogaster* Development." Developmental Cell **5**(2): 337-350.
- Aravin, A., D. Gaidatzis, et al. (2006). "A novel class of small RNAs bind to MILI protein in mouse testes." Nature **442**(7099): 203-207.
- Aravin, A. et al. (2007). "The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race." Science **318**(5851): 761-764
- Aravin, A. (2008). "A piRNA Pathway Primed by Individual Transposons is Linked to DeNovo DNA Methylation in Mice." Molecular Cell.
- Arnaud, P., K. Hata, et al. (2006). "Stochastic imprinting in the progeny of *Dnmt3L*^{-/-} females." Hum. Mol. Genet. **15**(4): 589-598.
- Baarends, W. M., E. Wassenaar, et al. (2005). "Silencing of Unpaired Chromatin and Histone H2A Ubiquitination in Mammalian Meiosis." Mol. Cell. Biol. **25**(3): 1041-1053.
- Bailey, J. A., L. Carrel, et al. (2000). "Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis." PNAS **97**(12): 6634-6639.
- Bartel, D. (2009). "MicroRNAs: target recognition and regulatory functions." Cell **136**(2): 215-233.
- Basyuk, E., F. Suavet, et al. (2003). "Human let-7 stem-loop precursors harbor features of RNase III cleavage products." Nucl. Acids Res. **31**(22): 6593-6597.

- Bellani, M. A., P. J. Romanienko, et al. (2005). "SPO11 is required for sex-body formation, and Spo11 heterozygosity rescues the prophase arrest of Atm^{-/-} spermatocytes." J Cell Sci **118**(15): 3233-3245.
- Bernstein, E., A. A. Caudy, et al. (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." Nature **409**(6818): 363-366.
- Bohmert, K, et al. (1998). "AGO1 defines a novel locus of Arabidopsis controlling leaf development." EMBO J. **17**(1): 170-180.
- Bohnsack, M. T., K. Czapinski, et al. (2004). "Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs " RNA **10**: 185-191.
- Bonetta, L. (2009). "RNA-based therapeutics: ready for delivery?" Cell **136**(4): 581-584.
- Bouché, N., D. Laressergues, et al. (2006). "An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs." EMBO Journal **25**(14): 3347-56.
- Bourc'his, D., G.-L. Xu, et al. (2001). "Dnmt3L and the Establishment of Maternal Genomic Imprints." Science **294**(5551): 2536-2539.
- Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." Nature **431**(7004): 96-99.
- Bourc'his, D. and T. H. Bestor (2006). "Origins of extreme sexual dimorphism in genomic imprinting." Cytogenetic and Genome Research **113**(1-4): 36-40.
- Brennecke, J., A. Stark, et al. (2005). "Principles of MicroRNA-Target Recognition." PLoS Biology **3**(3): e85.
- Brennecke, J. et al. (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila." Cell **128**(6): 1089-1103.
- Caplen, N. J., S. Parrish, et al. (2001). "Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems." PNAS **98**: 9742-9747.
- Carmell, M. A. et al. (2002). "The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis " Genes & Development **16**: 2733-2742.
- Carmell, M. A., et al. (2007). "MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline." Developmental Cell **12**(4): 503-514.
- Carthew, R. W. and E. J. Sontheimer (2009). "Origins and Mechanisms of miRNAs and siRNAs."

Cell **136**(4): 642-655.

Catalanotto, C., G. Azzalin, et al. (2002). "Involvement of small RNAs and role of the qde genes in the gene silencing pathway in *Neurospora*." Genes & Development **16**: 790-795.

Catchpoole, D., W. Lam, et al. (1997). "Epigenetic modification and uniparental inheritance of H19 in Beckwith-Wiedemann syndrome." Journal of Medical Genetics **34**(5): 353-359.

Cerutti, L., N. Mian, et al. (2000). "Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain." Trends in Biochemical Sciences **25**(10): 481-482.

Chen, T., Y. Ueda, et al. (2002). "A Novel Dnmt3a Isoform Produced from an Alternative Promoter Localizes to Euchromatin and Its Expression Correlates with Active de Novo Methylation." J. Biol. Chem. **277**(41): 38746-38754.

Chendrimada, T. P., R. I. Gregory, et al. (2005). "TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing." Nature **436**(7051): 740-744.

Chung, W.-J., K. Okamura, et al. (2008). "Endogenous RNA Interference Provides a Somatic Defense against *Drosophila* Transposons." Current Biology **18**(11): 795-802.

Clemens, J. C., C. A. Worby, et al. (2000). "Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways " PNAS **97**(12): 6499-6503.

Cogoni C. and Macino G. (1997). "Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in *Neurospora crassa*." PNAS **16**(94): 10233-10238.

Cox, D. N., A. Chao, et al. (1998). "A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal " Genes & Development **12**: 3715-3727.

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

Czech, B., C. D. Malone, et al. (2008). "An endogenous small interfering RNA pathway in *Drosophila*." Nature **453**(7196): 798-802.

Das, P. P., et al. (2008). "Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline." Molecular Cell **31**(1): 79-90.

Deleris, A., J. Gallego-Bartolome, et al. (2006). "Hierarchical Action and Inhibition of Plant Dicer-Like Proteins in Antiviral Defense." Science **313**(5783): 68-71.

Deng, W. and H. Lin (2002). "miwi, a Murine Homolog of piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis." Developmental Cell **2**(6): 819-830.

Denli, A. M., B. B. J. Tops, et al. (2004). "Processing of primary microRNAs by the Microprocessor complex." Nature **432**(7014): 231-235.

Dougherty, W. (1994). "RNA-mediated virus resistance in transgenic plants: exploitation of a cellular pathway possibly involved in RNA degradation." Mol Plant Microbe Interact. **7**(5): 544-552.

Elbashir, S. M., J. Harborth, et al. (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-498.

Elbashir, S. M., W. Lendeckel, et al. (2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes & Development **15**: 188-200.

Elbashir, S. M., J. Martinez, et al. (2001). "Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate." EMBO Journal **20**: 6877-6888.

Faehnle, C. R. and L. Joshua-Tor (2007). "Argonautes confront new small RNAs." Current Opinion in Chemical Biology **11**(5): 569-577.

Fagard, M., S. Boutet, et al. (2000). "AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals." PNAS **97**(21): 11650-11654.

Farazi, T. A., S. A. Juranek, et al. (2008). "The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members." Development **135**(7): 1201-1214.

Fimia, G. M., A. Morlon, et al. (2001). "Transcriptional cascades during spermatogenesis: pivotal role of CREM and ACT." Molecular and Cellular Endocrinology **179**(1-2): 17-23.

Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." Nature **391**(6669): 806-811.

Foerstemann, K., M. D. Horwich, et al. (2007). "Drosophila microRNAs Are Sorted into Functionally Distinct Argonaute Complexes after Production by Dicer-1." Cell **130**(2): 287-297.

Foerstemann, K., Y. Tomari, et al. (2005). "Normal microRNA Maturation and Germ-Line Stem Cell Maintenance Requires Loquacious, a Double-Stranded RNA-Binding Domain Protein." PLoS Biology **3**(7): e236.

Ghildiyal, M., H. Seitz, et al. (2008). "Endogenous siRNAs Derived from Transposons and mRNAs in *Drosophila* Somatic Cells." Science **320**(5879): 1077-1081.

Ghildiyal, M. and P. D. Zamore (2009). "Small silencing RNAs: an expanding universe." Nat

Rev Genet **10**(2): 94-108.

Girard, A. I., R. Sachidanandam, et al. (2006). "A germline-specific class of small RNAs binds mammalian Piwi proteins." Nature **442**(7099): 199-202.

Gregory, R. I., K.-p. Yan, et al. (2004). "The Microprocessor complex mediates the genesis of microRNAs." Nature **432**(7014): 235-240.

Grishok, A., A. E. Pasquinelli, et al. (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., E. Beyret, et al. (2006). "A novel class of small RNAs in mouse spermatogenic cells." Genes & Development **20**: 1709-1714.

Haase, A. D. and H. Z. Lukasz Jaskiewicz, Sébastien Lainé, Ragna Sack, Anne Gatignol & Witold Filipowicz (2005). "TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing " EMBO Reports **6**(10): 961–967

Hajkova, P., S. Erhardt, et al. (2002). "Epigenetic reprogramming in mouse primordial germ cells." Mechanisms of Development **117**(1-2): 15-23.

Hamilton, A. J. and D. C. Baulcombe (1999). "A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants." Science **286**(5441): 950-952.

Hammond, S. M., E. Bernstein, et al. (2000). "An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells." Nature **404**(6775): 293-296.

Hammond, S. M., S. Boettcher, et al. (2001). "Argonaute2, a Link Between Genetic and Biochemical Analyses of RNAi." Science **293**(5532): 1146-1150.

Han, J., Y. Lee, et al. (2004). "The Drosha-DGCR8 complex in primary microRNA processing." Genes & Development **18**: 3016-3027.

Harborth, J., S. M. Elbashir, et al. (2001). "Identification of essential genes in cultured mammalian cells using small interfering RNAs." J Cell Sci **114**(24): 4557-4565.

Hata, K., M. Okano, et al. (2002). "Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice." Development **129**(8): 1983-1993.

Hoeghe, C., B. Pfander, et al. (2002). "RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO." Nature **419**(6903): 135-141.

Homolka, D., R. Ivanek, et al. (2007). "Chromosomal rearrangement interferes with meiotic X chromosome inactivation." Genome Research **17**: 1431-1437.

- Houwing, S., L. M. Kamminga, et al. (2007). "A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon Silencing in Zebrafish." Cell **129**(1): 69-82.
- Howell, C. Y., T. H. Bestor, et al. (2001). "Genomic Imprinting Disrupted by a Maternal Effect Mutation in the Dnmt1 Gene." Cell **104**(6): 829-838.
- Hoyer-Fender, S. (2003). "Molecular aspects of XY body formation." Cytogenetic and Genome Research **103**(3-4): 245-255.
- Hutvagner, G., J. McLachlan, et al. (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 P. D. Zamore (2002). "A microRNA in a Multiple-Turnover RNAi Enzyme Complex." Science **297**(5589): 2056-2060.
- Ideraabdullah, F. Y. et al. (2008). "Genomic imprinting mechanisms in mammals." Mutation Research **647**(1-2): 77-85.
- Ingolia, N. T., S. Ghaemmamghami, et al. (2009). "Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling." Science **324**(5924): 218-223.
- Iwasaki, S., T. Kawamata, et al. (2009). "Drosophila Argonaute1 and Argonaute2 Employ Distinct Mechanisms for Translational Repression." **34**(1): 58-67.
- Kaffer, C. R., A. Grinberg, et al. (2001). "Regulatory Mechanisms at the Mouse Igf2/H19 Locus." Mol. Cell. Biol. **21**(23): 8189-8196.
- Kaffer, C. R., M. Srivastava, et al. (2000). "A transcriptional insulator at the imprinted H19/Igf2 locus." Genes & Development **14**(15): 1908-1919.
- Kaneda, M., M. Okano, et al. (2004). "Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting." Nature **429**(6994): 900-903.
- Kato, Y., M. Kaneda, et al. (2007). "Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse." Hum. Mol. Genet. **16**(19): 2272-2280.
- Kawamura, Y., K. Saito, et al. (2008). "Drosophila endogenous small RNAs bind to Argonaute-2 in somatic cells." Nature **453**(7196): 793-797.
- Kennerdell, J. R. and R. W. Carthew (1998). "Use of dsRNA-Mediated Genetic Interference to Demonstrate that frizzled and frizzled 2 Act in the Wingless Pathway." Cell **95**(7): 1017-1026.
- Ketting, R. F., S. E. J. Fischer, et al. (2001). "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans." Genes & Development **15**: 2654-2659.

- Klattenhoff, C., et al. (2007). "Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response." Developmental Cell **12**(1): 45-55.
- Kleene, K. C. (2001). "A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells." Mechanisms of Development **106**(1-2): 3-23.
- Knight, S. W. and B. L. Bass (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., M.-H. J. Ng, et al. (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.
- Kotaja, N., H. Lin, et al. (2006). "Interplay of PIWI/Argonaute protein MIWI and kinesin KIF17b in chromatoid bodies of male germ cells." J Cell Sci **119**(13): 2819-2825.
- Kozak, M. (1996). "Interpreting cDNA sequences: Some insights from studies on translation." Mammalian Genome **7**(8): 563-574.
- Kuramochi-Miyagawa, S. (2008). "DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes." Genes & Development **22**: 908-917.
- Kuramochi-Miyagawa, S., T. Kimura, et al. (2004). "Mili, a mammalian member of piwi family gene, is essential for spermatogenesis." Development **131**(4): 839-849.
- Kuramochi-Miyagawa, S., T. Kimura, et al. (2001). "Two mouse piwi-related genes: miwi and mili." Mechanisms of Development **108**(1-2): 121-133.
- Lagos-Quintana, M., R. Rauhut, et al. (2001). "Identification of Novel Genes Coding for Small Expressed RNAs." Science **294**(5543): 853-858.
- Lai, E. C. (2002). "Micro RNAs are complementary to 3[prime] UTR sequence motifs that mediate negative post-transcriptional regulation." Nat Genet **30**(4): 363-364.
- Landthaler, M., A. Yalcin, et al. (2004). "The Human DiGeorge Syndrome Critical Region Gene 8 and Its D. melanogaster Homolog Are Required for miRNA Biogenesis." Current Biology **14**(23): 2162-2167.
- Lau, N. C., L. P. Lim, et al. (2001). "An Abundant Class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis elegans*." Science **294**(5543): 858-862.

Lau, N. C., A. G. Seto, et al. (2006). "Characterization of the piRNA Complex from Rat Testes." Science **313**(5785): 363-367.

Lee, D. W., R. J. Pratt, et al. (2003). "An argonaute-like protein is required for meiotic silencing." Genetics **164**(2): 821-828.

Lee, R. C. and V. Ambros (2001). "An Extensive Class of Small RNAs in *Caenorhabditis elegans*." Science **294**(5543): 862-864.

Lee, R. C., R. L. Feinbaum, et al. (1993). "The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*." Cell **75**(5): 843-854.

Lee, Y., C. Ahn, et al. (2003). "The nuclear RNase III Drosha initiates microRNA processing." Nature **425**(6956): 415-419.

Lee, Y. and S.-Y. P. Inha Hur, Young-Kook Kim Mi Ra Suh and V Narry Kim (2006). "The role of PACT in the RNA silencing pathway." The EMBO Journal **25**(3): 522-32.

Lee, Y., M. Kim, et al. (2004). "MicroRNA genes are transcribed by RNA polymerase II." EMBO Journal **23**: 4051-4060.

Lees-Murdock, D. J., M. De Felici, et al. (2003). "Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage." Genomics **82**(2): 230-237.

Leighton, P., J. Saam, et al. (1995). "An enhancer deletion affects both H19 and Igf2 expression." Genes & Development **9**(17): 2079-2089.

Leuschner, P. J. (2006). "Cleavage of the siRNA passenger strand during RISC assembly in human cells." EMBO Rep. **7**(3): 314-320.

Lewis, B. P., I. h. Shih, et al. (2003). "Prediction of Mammalian MicroRNA Targets." Cell **115**(7): 787-798.

Li, E., C. Beard, et al. (1993). "Role for DNA methylation in genomic imprinting." Nature **366**: 362-365.

Li, J., Z. Yang, et al. (2005). "Methylation Protects miRNAs and siRNAs from a 3'-End Uridylation Activity in *Arabidopsis*." Current Biology **15**(16): 1501-1507.

Lim, L. P., N. C. Lau, et al. (2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs." Nature **433**(7027): 769-773.

Lin, H. and A. C. Spradling (1997). "A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary." Development **124**(12): 2463-2476.

Lin, S.-Y., S. M. Johnson, et al. (2003). "The *C. elegans* hunchback Homolog, *hbl-1*, Controls Temporal Patterning and Is a Probable MicroRNA Target." Developmental Cell **4**(5): 639-650.

Lingel, A., B. Simon, et al. (2003). "Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain." Nature **426**(6965): 465-469.

Liu, J., M. A. Carmell, et al. (2004). "Argonaute2 Is the Catalytic Engine of Mammalian RNAi." Science **305**(5689): 1437-1441.

Liu, Q., T. A. Rand, et al. (2003). "R2D2, a Bridge Between the Initiation and Effector Steps of the Drosophila RNAi Pathway." Science **301**(5641): 1921-1925.

Lund, E., S. Guttinger, et al. (2004). "Nuclear Export of MicroRNA Precursors." Science **303**(5654): 95-98.

Ma, J. B., et al. (2004). "Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain." Nature **429**(6989): 318-322.

MacRae, I. J., K. Zhou, et al. (2006). "Structural Basis for Double-Stranded RNA Processing by Dicer." Science **311**(5758): 195-198.

Mallory, A. C. and H. Vaucheret (2006). "Functions of microRNAs and related small RNAs in plants." Nat Genet.

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

Martinez, J., A. Patkaniowska, et al. (2002). "Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi." Cell **110**(5): 563-574.

Matranga, C., Y. Tomari, et al. (2005). "Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes." Cell **123**(4): 607-620.

Meister, G., M. Landthaler, et al. (2004). "Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs." Molecular Cell **15**(2): 185-197.

Mertineit, C., J. A. Yoder, et al. (1998). "Sex-specific exons control DNA methyltransferase in mammalian germ cells." Development **125**(5): 889-897.

Mette, M. F., W. Aufsatz, et al. (2000). "Transcriptional silencing and promoter methylation triggered by double-stranded RNA." EMBO Journal **19**: 5194-5201.

Miyoshi, K., H. Tsukumo, et al. (2005). "Slicer function of Drosophila Argonautes and its involvement in RISC formation." Genes & Development **19**: 2837-2848.

Mochizuki, K., et al. (2002). "Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*." Cell **110**(6): 689-699.

- Morlando, M., M. Ballarino, et al. (2008). "Primary microRNA transcripts are processed co-transcriptionally." Nat Struct Mol Biol **15**(9): 902-909.
- Mourelatos, Z., J. Dostie, et al. (2002). "miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs." Genes & Development **16**: 720-728.
- Namekawa, S. H., P. J. Park, et al. (2006). "Postmeiotic Sex Chromatin in the Male Germline of Mice." Current Biology **16**(7): 660-667.
- Napoli, C., C. Lemieux, et al. (1990). "Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans." Plant Cell **2**(4): 279-289.
- Noma, K.-i., T. Sugiyama, et al. (2004). "RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing." Nat Genet **36**(11): 1174-1180.
- Nykaenen, A., B. Haley, et al. (2001). "ATP Requirements and Small Interfering RNA Structure in the RNA Interference Pathway." Cell **107**(3): 309-321.
- Okamura, K., S. Balla, et al. (2008). "Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*." Nat Struct Mol Biol **15**(9): 998-998.
- Okamura, K., W.-J. Chung, et al. (2008). "The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs." Nature **453**(7196): 803-806.
- Okamura, K., J. W. Hagen, et al. (2007). "The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in *Drosophila*." Cell **130**(1): 89-100.
- Okano, M., D. W. Bell, et al. (1999). "DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development." Cell **99**(3): 247-257.
- Pak, J. and A. Fire (2007). "Distinct Populations of Primary and Secondary Effectors During RNAi in *C. elegans*." Science **315**(5809): 241-244.
- Pal-Bhadra, M., U. Bhadra, et al. (2002). "RNAi Related Mechanisms Affect Both Transcriptional and Posttranscriptional Transgene Silencing in *Drosophila*." Molecular Cell **9**(2): 315-327.
- Pane, A., K. Wehr, et al. (2007). "zucchini and squash Encode Two Putative Nucleases Required for rasiRNA Production in the *Drosophila* Germline." Developmental Cell **12**(6): 851-862.
- Parker, J. S., M. S. Roe, et al. (2004). "Crystal structure of a PIWI protein suggests mechanisms

for siRNA recognition and slicer activity." The EMBO Journal **23**: 4727-4737.

Pasquinelli, A. E., B. J. Reinhart, et al. (2000). "Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA." Nature **408**(6808): 86-89.

Paulsen, M. and A. C. Ferguson-Smith (2001). "DNA methylation in genomic imprinting, development, and disease." The Journal of Pathology **195**(1): 97-110.

Peaston, A. E., A. V. Evsikov, et al. (2004). "Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos." Developmental Cell **7**(4): 597-606.

Pham, C. T., D. M. MacIvor, et al. (1996). "Long-range disruption of gene expression by a selectable marker cassette." PNAS **93**: 13090-13095.

Rand, T. A., S. Petersen, et al. (2005). "Argonaute2 Cleaves the Anti-Guide Strand of siRNA during RISC Activation." Cell **123**(4): 621-629.

Ratcliff, F., B. D. Harrison, et al. (1997). "A Similarity Between Viral Defense and Gene Silencing in Plants." Science **276**(5318): 1558-1560.

Reik, W. (2007). "Stability and flexibility of epigenetic gene regulation in mammalian development." Nature **447**(7143): 425-432.

Reinhart, B. J., F. J. Slack, et al. (2000). "The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*." Nature **403**(6772): 901-906.

Rhoades, M. W., B. J. Reinhart, et al. (2002). "Prediction of Plant MicroRNA Targets." Cell **110**(4): 513-520.

Romano N, M. G. (1992). "Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences." Molecular Microbiology **6**(22): 3343-3353.

Ross, M. T., D. V. Grafham, et al. (2005). "The DNA sequence of the human X chromosome." Nature **434**(7031): 325-337.

Royo, H. and Cavaille, J. (2008). "Non-coding RNAs in imprinted gene clusters." Biology of the Cell **100**(3): 149-166.

Rubin, C. M., C. A. VandeVoort, et al. (1994). "Alu repeated DNAs are differentially methylated in primate germ cells." Nucl. Acids Res. **22**(23): 5121-5127.

Ruby, J. G., C. H. Jan, et al. (2007). "Intronic microRNA precursors that bypass Drosha processing." Nature **448**(7149): 83-86.

Ruiz, M. T., O. Voinnet, et al. (1998). "Initiation and Maintenance of Virus-Induced Gene

Silencing." Plant Cell **10**(6): 937-946.

Saito, K., A. Ishizuka, et al. (2005). "Processing of Pre-microRNAs by the Dicer-1- Loquacious Complex in Drosophila Cells." PLoS Biology **3**(7): e235.

Saito, K., K. M. Nishida, et al. (2006). "Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome." Genes & Development **20**: 2214-2222.

Saleh, M.-C., M. Tassetto, et al. (2009). "Antiviral immunity in Drosophila requires systemic RNA interference spread." Nature **458**(7236): 346-350.

Sasaki, T., A. Shiohama, et al. (2003). "Identification of eight members of the Argonaute family in the human genome[small star, filled]." Genomics **82**(3): 323-330.

Schimenti, J. (2005). "Synapsis or silence." Nat Genet **37**(1): 11-13.

Schmidt, A., G. Palumbo, et al. (1999). "Genetic and Molecular Characterization of sting, a Gene Involved in Crystal Formation and Meiotic Drive in the Male Germ Line of Drosophila melanogaster." Genetics **151**(2): 749-760.

Schwarz, D. S., G. Hutvagner, et al. (2003). "Asymmetry in the Assembly of the RNAi Enzyme Complex." Cell **115**(2): 199-208.

Seto, A. G., R. E. Kingston, et al. (2007). "The Coming of Age for Piwi Proteins." Molecular Cell **26**(5): 603-609.

Sharma, A. K., M. C. Nelson, et al. (2001). "Human CD34+ stem cells express the hiwi gene, a human homologue of the Drosophila gene piwi." Blood **97**(2): 426-434.

Shiu, P. K. T., N. B. Raju, et al. (2001). "Meiotic Silencing by Unpaired DNA." Cell **107**(7): 905-916.

Shovlin, T. C., D. Bourc'his, et al. (2007). "Sex-specific promoters regulate Dnmt3L expression in mouse germ cells." Hum. Reprod. **22**(2): 457-467.

Sijen, T., et al. (2001). "On the role of RNA amplification in dsRNA-triggered gene silencing." Cell **107**(4): 465-476.

Sijen, T., et al. (2007). "Secondary siRNAs Result from Unprimed RNA Synthesis and Form a Distinct Class." Science **315**(5809): 244-247.

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

Solari, A. J. (1974). "The behavior of XY pair in mammals." Review of Cytology **38**: 273-317.

- Song, J.-J., J. Liu, et al. (2003). "The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes." Nat Struct Mol Biol **10**(12): 1026-1032.
- Song, J.-J., S. K. Smith, et al. (2004). "Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity." Science **305**(5689): 1434-1437.
- Stark, A., J. Brennecke, et al. (2003). "Identification of Drosophila MicroRNA Targets." PLoS Biology **1**(3): e60.
- Svoboda, P., P. Stein, et al. (2000). "Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference." Development **127**(19): 4147-4156.
- Tabara, H., M. Sarkissian, et al. (1999). "The rde-1 Gene, RNA Interference, and Transposon Silencing in *C. elegans*." Cell **99**(2): 123-132.
- Taverna, S. D., et al. (2002). "Methylation of histone h3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*." Cell **110**(6): 701-711.
- Tam, O. H., A. A. Aravin, et al. (2008). "Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes." Nature **453**(7194): 534-538.
- Tang, G., B. J. Reinhart, et al. (2003). "A biochemical framework for RNA silencing in plants." Genes & Development **17**: 49-63.
- Thorvaldsen, J. L., K. L. Duran, et al. (1998). "Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2." Genes & Development **12**: 3693-3702.
- Tolia, N. H. and Joshua-Tor, L. (2007). "Slicer and the Argonautes". Nat. Chem. Biol. **3**(1): 36-43.
- Tomari, Y., T. Du, et al. (2007). "Sorting of Drosophila Small Silencing RNAs." Cell **130**(2): 299-308.
- Tomari, Y., C. Matranga, et al. (2004). "A Protein Sensor for siRNA Asymmetry." Science **306**(5700): 1377-1380.
- Trasler, J. M., A. A. Alcivar, et al. (1992). "DNA methyltransferase is developmentally expressed in replicating and non-replicating male germ cells." Nucl. Acids Res. **20**(10): 2541-2545.
- Turner, J. M. A. (2007). "Meiotic sex chromosome inactivation." Development **134**(10): 1823-1831.

Turner, J. M. A., O. Aprelikova, et al. (2004). "BRCA1, Histone H2AX Phosphorylation, and Male Meiotic Sex Chromosome Inactivation." Current Biology **14**(23): 2135-2142.

Turner, J. M. A., S. K. Mahadevaiah, et al. (2005). "Silencing of unsynapsed meiotic chromosomes in the mouse." Nat Genet **37**(1): 41-47.

Unhavaithaya, Y., Y. Hao, et al. (2009). "MILI, a PIWI-interacting RNA-binding Protein, Is Required for Germ Line Stem Cell Self-renewal and Appears to Positively Regulate Translation." J. Biol. Chem. **284**(10): 6507-6519.

Vagin, V. V., A. Sigova, et al. (2006). "A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline." Science **313**(5785): 320-324.

Vaistij, F. E., L. Jones, et al. (2002). "Spreading of RNA Targeting and DNA Methylation in RNA Silencing Requires Transcription of the Target Gene and a Putative RNA-Dependent RNA Polymerase." Plant Cell **14**(4): 857-867.

van der Krol, A. R., L. A. Mur, et al. (1990). "Flavonoid Genes in Petunia: Addition of a Limited Number of Gene Copies May Lead to a Suppression of Gene Expression." Plant Cell **2**(4): 291-299.

Vaucheret, H. (2008). "Plant ARGONAUTES." Trends in Plant Science **13**(7): 350-358.

Ventura, A. and T. Jacks (2009). "MicroRNAs and Cancer: Short RNAs Go a Long Way." Cell **136**(4): 586-591.

Verdel, A., S. Jia, et al. (2004). "RNAi-Mediated Targeting of Heterochromatin by the RITS Complex." Science **303**(5658): 672-676.

Voinnet, O., P. Vain, et al. (1998). "Systemic Spread of Sequence-Specific Transgene RNA Degradation in Plants Is Initiated by Localized Introduction of Ectopic Promoterless DNA." Cell **95**(2): 177-187.

Wargelius, A., S. Ellingsen, et al. (1999). "Double-Stranded RNA Induces Specific Developmental Defects in Zebrafish Embryos." Biochemical and Biophysical Research Communications **263**(1): 156-161.

Watanabe, T., Y. Totoki, et al. (2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." Nature **453**(7194): 539-543.

Webster, K. E., M. K. O'Bryan, et al. (2005). "Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis." PNAS **102**(11): 4068-4073.

Wheelan, S., D. Church, et al. (2001). "Spidey: a tool for mRNA-to-genomic alignments." Genome Research **11**(11): 1952-1957.

Wianny, F. and M. Zernicka-Goetz (2000). "Specific interference with gene function by

double-stranded RNA in early mouse development." Nat Cell Biol **2**(2): 70-75.

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

Williams, R. W. and G. M. Rubin (2002). "ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos." PNAS **99**: 6889-6894.

Yekta, S., I. h. Shih, et al. (2004). "MicroRNA-Directed Cleavage of HOXB8 mRNA." Science **304**(5670): 594-596.

Yi, R., Y. Qin, et al. (2003). "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs " Genes & Development **17**: 3011-3016.

Yigit, E., P. J. Batista, et al. (2006). "Analysis of the *C. elegans* Argonaute Family Reveals that Distinct Argonautes Act Sequentially during RNAi." Cell **127**(4): 747-757.

Yoshikawa, P., Park, R. Scott Poethig (2005). "A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*." Genes & Development **19**: 2164-2175.

Yu, B., Z. Yang, et al. (2005). "Methylation as a Crucial Step in Plant microRNA Biogenesis." Science **307**(5711): 932-935.

Zamore, P. D. (2006). "RNA Interference: Big Applause for Silencing in Stockholm." Cell **127**(6): 1083-1086.

Zamore, P. D., T. Tuschl, et al. (2000). "RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals." Cell **101**(1): 25-33.

Zilberman, D., X. Cao, et al. (2003). "ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation." Science **299**(5607): 716-719.

Zilberman, D., X. Cao, et al. (2004). "Role of *Arabidopsis* ARGONAUTE4 in RNA-Directed DNA Methylation Triggered by Inverted Repeats." Current Biology **14**(13): 1214-1220.

7. Acknowledgments.

At the end of my doctoral studies, which have been a constant rollercoaster of ups and downs and the most intense and invaluable period of my life, I would like to thank a few people. You've made it all possible, and I'm forever grateful for your help and support throughout these four years.

I would like to thank Javier, for giving me the opportunity to work in the great place that IMBA is in the first place, and experience high-level science and unlimited opportunities. I feel privileged to have worked in a great environment that Martinez lab always was. I thank you for allowing us to go to so many meetings and conferences, they were a fantastic experience in terms of meeting people and learning. Thank you for all your support and for believing in me, sometimes against all odds. And thank you for the independence I had in my work, as it enabled me a tough trial-and-error learning process, which I consider truly invaluable.

I would like to thank my PhD Thesis Committee members, Leonie Ringrose and Renée Schroeder, who always had time for me and gave me a lot of support, particularly in times when I had problems with the project.

I would like to thank Alexei Aravin from Greg Hannon's lab, for sharing material and data, and a lot of helpful discussions.

The greatest THANK YOU to all the people in the Martinez lab, Katrin, Christoph, Andrea, Philipp, Stefan, Gregor, Johannes, Jutta, Jiradet, and Anne, for all the help, discussions and support, and creating a great working environment. Work was always a fun place to come to.

I would particularly like to thank Katrin, Christoph, and Andrea, truly great people and colleagues. I'm lucky to have you as my friends. I learned a lot from you, and I'm really happy we shared this time and experience.

Dear Andrea, you are and will always be deeply missed.

Finally, I thank the people that make all the difference between life, and mere existence - my friends and family.

Biggest, greatest hugs and thankyou's go to Lada and Kata, my best friends and an untouchable shield; Jelena and Branka; Katrin, Anja and Andrea; Natko and Marin.

To the Secret Keepers of PD, keep the bubble and feel free to dance ☺ What is for us, will not pass us by. One PD to rule them all!

In the end, I'm grateful beyond words to the most important people in my life, my safe haven - my family: my parents and grandparents and my brother and sister, Nikola and Marta. Just knowing you're there for me no matter what, is support incomparable to any other. I would not have made it without you.

8. Curriculum vitae

Dubravka Pezic

Group Javier Martinez
Institute of Molecular Biotechnology of Austrian Academy of Sciences (IMBA)
Dr. Bohr-Gasse 3
A-1030 Vienna
Austria
Phone: +43 1 79044 4858
Fax: +43 1 79044 110
Email: Dubravka.Pezic@imba.oeaw.ac.at

Date of birth	2 June 1981
Place of birth	Zagreb, Croatia
Nationality	Croatian

Education

- 2005-present PhD studies
Vienna Biocenter International PhD Program in
molecular life sciences
Institute of Molecular Biotechnology (IMBA), Vienna,
Austria
- 1999 – 2005 Molecular biology studies
Faculty of Science and Mathematics, University of
Zagreb, Croatia
Average grade: 4.88/5.00

Employment

- 2005-present Institute of Molecular Biotechnology (IMBA)
Vienna, Austria

PhD Graduate Student

Project: Expression and function of mouse Argonaute
protein Miwi2.
Supervisor: Dr. Javier Martinez

- 2004-2005 Science Laboratory for Plant Virology, Biology Dept., Faculty of and Mathematics, University of Zagreb Zagreb, Croatia
Diploma Student
Project: Analysis of origin and evolution of different strains of plant pathogen *Phytoplasma* using single strand conformation polymorphism (SSCP).
Supervisor: Dr. Dijana Škorić

Summer Internships

- 2003, July and August INSERM/U571; Necker Medical School
Paris, France
Dr. Miroslav Radman lab
Project: Construction of λ phage-based vector for studying directed evolution in bacteria.
- 2002, July and August Physiology Department, Medical Faculty,
McGill University
Montreal, Canada
Dr. John Hanrahan lab
Project: Oligomerization of cystic fibrosis transmembrane conductance regulator (CFTR); activity and function of the oligomeric state.

Publications

- Heindl K, **Pezic D**, Gebeshuber C.
Small RNAs and their big implications: in vivo, in vitro, in silico.
RNA Biol. 2007 Nov;4(3):165-8
(Meeting Report)
- Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, **Pezic D**, Toth KF, Bestor T, Hannon GJ.
A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice.
Mol Cell. 2008 Sep 26;31(6):785-99.

Fellowships

- 'Top scholarship for Top students', awarded by Croatian weekly journal 'Nacional'; 2004-2005.
- Special student scholarship from County of Zagreb; 2002-2004.
- 'Partnerships for tomorrow' scholarship from Canadian International Development Agency for summer internship at McGill University, Montreal; 2002.
- Scholarship from the Government of Republic of Croatia, Ministry of Science, Education and Sport; 2001-2004.

Conferences and presentations

- June 2008 RNA Society meeting; Berlin, Germany (attendant)
- May 2008 Microsymposium on Small RNAs; Vienna, Austria (oral presentation)
- March 2008 Keystone Symposium: RNAi, MicroRNA, and Non-Coding RNA; Whistler, British Columbia, Canada (poster presentation)
- October 2007 IMP-IMBA Annual Recess; Vienna, Austria (poster presentation)
- June 2007 International PhD Student Conference; Cambridge, UK (oral presentation)
- May 2007 Microsymposium on Small RNAs; Vienna, Austria (attendant)
- January 2007 Keystone Symposium: MicroRNAs and siRNAs: Biological Functions and Mechanisms; Keystone, Colorado, USA (attendant)
- June 2006 Regulatory RNAs: LXXI Cold Spring Harbor Symposium on Quantitative Biology; CSHL, Cold Spring Harbor, NY, USA (attendant)
- May 2006 Microsymposium on Small RNAs; Vienna, Austria (attendant)

Teaching experience

- May 2008 University of Vienna Advanced Cell Biology Course for undergraduate students (lecture on Piwi proteins and piRNAs).
- April 2008 Vienna Biocenter PhD Student Lecture Series on RNA silencing and epigenetics (lecture on Piwi proteins and piRNAs).

References

- **Dr. Javier Martinez** Institute for Molecular Biotechnology of the Austrian Academy of Sciences (IMBA)
Dr. Bohr-Gasse 3
A-1030 Vienna, Austria
+43 1 79044 4840
Javier.Martinez@imba.oeaw.ac.at
- **Prof. Dr. Renee Schroeder** Max F. Perutz Laboratories
Dept. of Biochemistry
Dr. Bohrgasse 9/5
A-1030 Vienna, Austria
+43 1 4277 54690
renee.schroeder@univie.ac.at
- **Dr. Leonie Ringrose** Institute for Molecular Biotechnology of the Austrian Academy of Sciences (IMBA)
Dr. Bohr-Gasse 3
A-1030 Vienna, Austria
+ 43 1 79044 4650
Leonie.Ringrose@imba.oeaw.ac.at
- **Prof. Dr. Dijana Škorić** Botanical Section, Biology Department
Faculty of Science and Mathematics
University of Zagreb
Marulicev trg 9a
10000 Zagreb, Croatia
+385 1 4898 079

9. Appendix

Publication.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Fejes-Toth, K., Bestor, T., Hannon, G. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. Molecular Cell **31** 785-799.

A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice

Alexei A. Aravin,^{1,2} Ravi Sachidanandam,^{1,2,6} Deborah Bourc'his,³ Christopher Schaefer,^{3,4} Dubravka Pezic,⁵ Katalin Fejes Toth,^{1,2} Timothy Bestor,⁴ and Gregory J. Hannon^{1,2,*}

¹Watson School of Biological Sciences

²Howard Hughes Medical Institute

Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

³Institute Jacques Monod, Inserm U741/Paris 7 University, Tour 43-44/2eme Etage/Couloir 43-44, 2 Place Jussieu, Paris 75251, Cedex 05, France

⁴Department of Genetics and Development, College of Physicians and Surgeons, 701 West 168th Street, Columbia University, New York, NY 10032, USA

⁵Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr. Bohr-Gasse 3, Vienna 1030, Austria

⁶Present address: Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, 5 East 98th Street, New York, NY 10029, USA

*Correspondence: hannon@cshl.edu

DOI 10.1016/j.molcel.2008.09.003

SUMMARY

piRNAs and Piwi proteins have been implicated in transposon control and are linked to transposon methylation in mammals. Here we examined the construction of the piRNA system in the restricted developmental window in which methylation patterns are set during mammalian embryogenesis. We find robust expression of two Piwi family proteins, MIWI2 and MILI. Their associated piRNA profiles reveal differences from *Drosophila* wherein large piRNA clusters act as master regulators of silencing. Instead, in mammals, dispersed transposon copies initiate the pathway, producing primary piRNAs, which predominantly join MILI in the cytoplasm. MIWI2, whose nuclear localization and association with piRNAs depend upon MILI, is enriched for secondary piRNAs antisense to the elements that it controls. The Piwi pathway lies upstream of known mediators of DNA methylation, since piRNAs are still produced in *dnmt3L* mutants, which fail to methylate transposons. This implicates piRNAs as specificity determinants of DNA methylation in germ cells.

INTRODUCTION

It is of paramount importance that germ cell genomes be protected from the uncontrolled propagation of mobile genetic elements. This prevents both long-term reductions in fitness through accumulation of mutations and short-term reductions in fertility due to germ cell loss. For protection to be effective, mobile elements must be distinguished from endogenous genes and selectively silenced. This presents a significant challenge, as few characteristics unambiguously mark the many families of mobile elements that colonize animal genomes (Girard and Hannon, 2008).

In *Drosophila*, germ cells express a class of small RNAs, piRNAs, that are specialized for mobile element repression (Brennecke et al., 2007; Saito et al., 2006; Vagin et al., 2006). These act in a nucleic acid-based innate immune system that comprises both genetically encoded (primary piRNAs) and adaptive (secondary piRNAs) resistance mechanisms. Primary piRNAs are generated from dedicated loci, called piRNA clusters, that contain the highest density of transposon-related sequences in the fly (Brennecke et al., 2007).

Primary piRNAs join Piwi proteins and guide these to the selection of their targets. Two outcomes follow recognition of a transposon mRNA by a piRNA. First, the mRNA is cleaved, resulting in its destruction. The cleavage event also promotes the production of a secondary piRNA derived from the mRNA itself (Brennecke et al., 2007; Gunawardane et al., 2007). Though this is in the sense orientation with respect to the transposon, it can target antisense transposon RNAs that are generated from piRNA clusters. Cleavage of such transcripts reproduces original antisense piRNAs that can again target transposons. As a whole, this creates a cycle, called the “ping-pong cycle” that optimizes the piRNA population to target active elements (Brennecke et al., 2007; reviewed in Aravin et al., 2007a).

Features of the *Drosophila* piRNA system have been conserved in vertebrates. Signatures of ping-pong amplification cycles have been detected in zebrafish (Houwing et al., 2007) and mammals (Aravin et al., 2007b). Moreover, repeat-enriched mouse piRNA clusters give rise to small RNA species in both male and female germ cells (Aravin et al., 2007b; Tam et al., 2008). Finally, LTR and non-LTR transposons are overexpressed in male germ cells deficient in either of two Piwi family members, *mili* or *miwi2* (Aravin et al., 2007b; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008).

In contrast to *Drosophila*, epigenetically stable repression of transposable elements in mammals requires CpG DNA methylation (Bourc'his and Bestor, 2004; Walsh et al., 1998). Transposon methylation patterns are extensively remodeled during mammalian development. Following fertilization, gametic methylation patterns are mostly lost (Lane et al., 2003). De novo

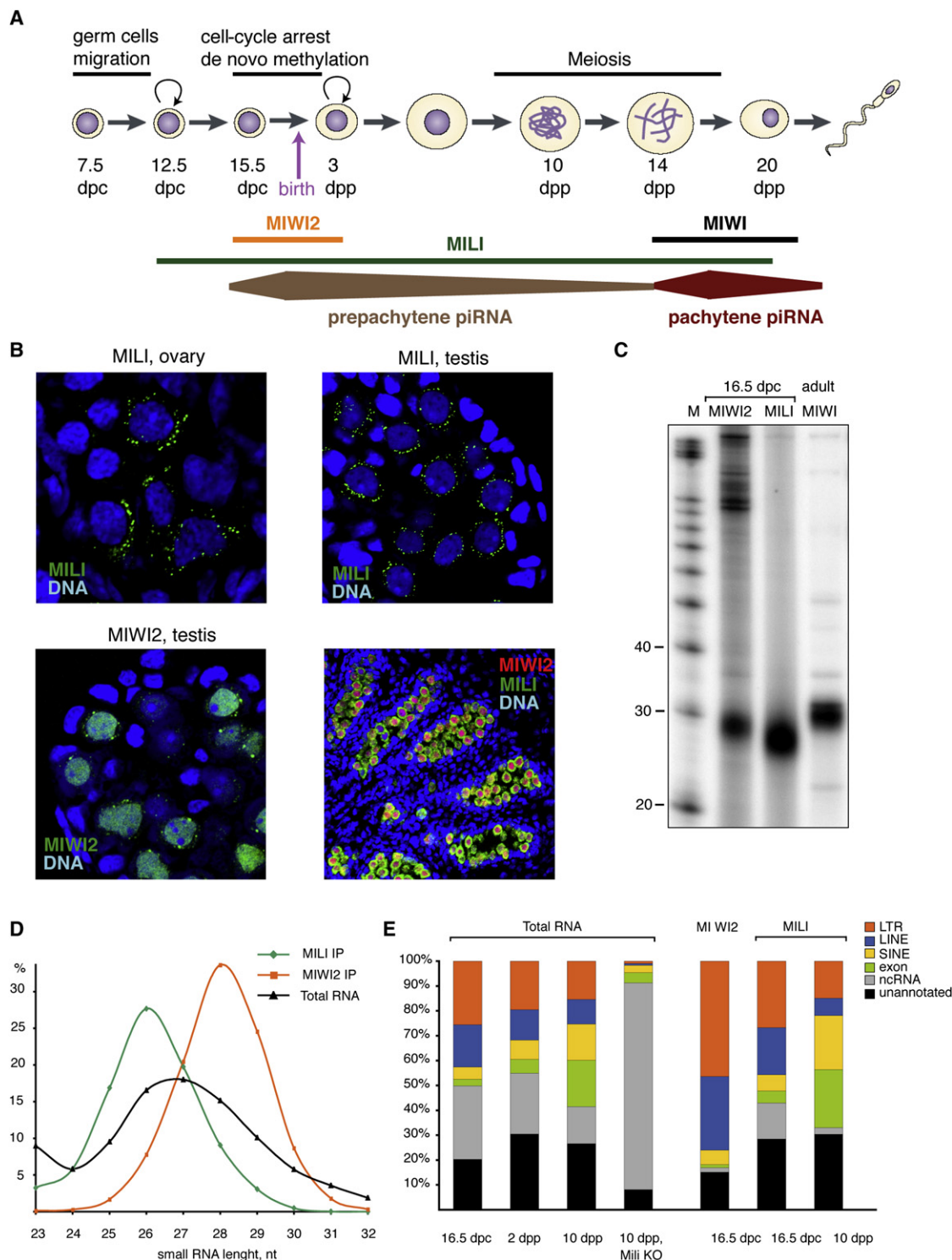


Figure 1. Expression of Piwi Proteins through Germ Cell Development

(A) A scheme of spermatogenesis is shown with the timing of expression of *mili*, *miwi*, and *miwi2* indicated. After migration, primordial germ cells (PGCs) arrive at the gonad around 11.5 dpc and expand prior to undergoing cell-cycle arrest at 15.5 dpc. The timing of cell-cycle arrest coincides with establishment of de novo DNA methylation patterns on transposable elements and imprinted genes. Germ cells resume division after birth at around 3 dpp and initiate meiotic division at 10 dpp. The first cells at pachytene and haploid round spermatid stages appear at days ~14 and ~20, respectively.

methylation patterns are then established around implantation and subsequently maintained in somatic cells throughout the life of the organism. The germ line undergoes a wave of transposon demethylation soon after its emergence during embryogenesis, as it colonizes the gonadal compartment (Hajkova et al., 2002). The subsequent establishment of gametic methylation patterns differs between the two sexes. In males, after migration of primordial germ cells (PGCs) into embryonic gonads and their initial expansion, germ cells arrest their cell cycle around 14.5 days postcoitum (dpc) as prospermatogonia, only resuming division 2–3 days after birth (dpp). This is the critical window during which male gametic methylation patterns are established (Kato et al., 2007; Lees-Murdock et al., 2003).

Several members of the DNA methyltransferase family, DNMT3A, DNMT3B, and DNMT3L, act in de novo methylation of transposable elements. Catalytically active DNMT3A and DNMT3B are important in both germ and somatic cells, where they perform complementary and nonoverlapping functions. Constitutive *dnmt3a* or *dnmt3b* mutations are embryonically lethal (Okano et al., 1999), and germ cell-conditional inactivation of *dnmt3a* results in sterility (Kaneda et al., 2004). In contrast, DNMT3L operates as a central regulator of de novo methylation specifically in the germline. *dnmt3L*-deficient animals globally fail to establish de novo methylation of transposons in their germ cells with no other phenotypic manifestation (Bourc'his and Bestor, 2004; Kato et al., 2007). In males, this results in uncontrolled transposon expression and eventually in spermatogenesis failure and sterility. Recent studies have begun to unravel the biochemistry of de novo DNA methylation machinery, showing that DNA methylation might be preceded by specific histone modifications (Jia et al., 2007; Ooi et al., 2007). However, it is still not clear how transposon sequences are specifically recognized to receive such modifications.

In the male germline, deficiency in either of two Piwi family members, *mili* or *miwi2*, results in loss of the DNA methylation marks on transposons, and mutant animals display a phenotype remarkably similar to that of *dnmt3L*-deficient mice (Aravin et al., 2007b; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2004, 2008). These data led to the hypothesis that Piwi/piRNA complexes might serve as sequence-specific guides that direct the de novo DNA methylation machinery to transposable elements (Aravin et al., 2007b; Kuramochi-Miyagawa et al., 2008). Thus far, we understand little of how the piRNA system operates in the restricted developmental window that is critical to the stable epigenetic silencing of repeat elements. This prompted investigation of both Piwi protein and piRNA expression during the time in which methylation marks are established.

RESULTS

Expression of Piwi Family Members during Germline Development

The mouse genome contains three Piwi family members: *mili*, *miwi*, and *miwi2*. *miwi* is expressed from the pachytene stage of meiosis to the haploid round spermatid stage (Deng and Lin, 2002) (Figure 1A). MILI is also present during meiosis, at which point both MILI and MIWI interact with an extremely abundant class of small RNAs, the pachytene piRNAs (Aravin et al., 2006; Girard et al., 2006). These are derived from specific genomic loci and form a complex population of small RNAs that match only to those sites from which they are derived. The function of this subclass of piRNAs is elusive.

mili is also expressed earlier in development. Just before entry into meiosis and the onset of pachytene piRNA expression, MILI binds piRNAs, which are different in character and genomic origin from meiotically expressed piRNAs (Aravin et al., 2007b). These are derived from a set of clustered loci that are repeat enriched and thus give rise to small RNA populations corresponding to transposons.

Loss of *miwi2* gives both morphological (e.g., meiotic arrest and progressive germ cell loss) and molecular phenotypes (increased transposon expression) that resemble those resulting from *mili* deficiency (Aravin et al., 2007b; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2004, 2008). However, neither the developmental timing of *miwi2* expression nor its subcellular localization has been reported.

We investigated *mili* and *miwi2* expression both by using specific antibodies raised against each family member and by using transgenic animals that express GFP- and myc-tagged MILI and MIWI2 proteins under the control of their endogenous promoters. For each protein, examination of several GFP- and myc-transgenic lines and comparison to results obtained with antibodies against native proteins revealed essentially identical patterns.

We analyzed the expression of *mili* and *miwi2* during germ line development in male and female mice. *mili* expression could be detected in both sexes as early as 12.5 dpc, a time when migrating PGCs have reached the somatic genital ridge (Figure 1A, data not shown). MILI localized to numerous perinuclear cytoplasmic granules in male PGCs (Figure 1B). These structures were remarkably similar to nuage, which contain the Piwi proteins, AUB and AGO3, in *Drosophila* female germ cells. MILI was also detected in female PGCs and localized to cytoplasmic granules (Figure 1B). *mili* expression continued in both male and female germ cells after birth (see Figure S1A available online). In ovary it localized to cytoplasmic granules of both arrested and growing oocytes. In adult testis, it was expressed throughout

(B) Expression of GFP-MILI and GFP-MIWI2 transgenes in prenatal (17.5 dpc) germ cells is shown. MILI exclusively expresses in germ cells and concentrates in perinuclear cytoplasmic granules in both developing testis and ovary. MIWI2 is absent in female germ cells and localizes in the nucleus as well as in cytoplasmic granules in male germ cells. MILI and MIWI2 are coexpressed in male germ cells at 17.5 dpc and are absent from somatic cells.

(C) MILI, MIWI2, and MIWI complexes were immunoprecipitated from embryonic gonad at 16.5 dpc (MILI and MIWI2) or adult (MIWI) testes. Associated piRNAs were 5' labeled using sequential phosphatase/kinase reactions.

(D) Size profiles are shown for MILI and MIWI2-bound piRNAs cloned from prenatal testis.

(E) Small RNA libraries were prepared from size-selected (24–33 nt) total RNA and MILI and MIWI2-immunopurified complexes from the indicated developmental stages and tissue sources. Their small RNA content was classified as indicated.

spermatogenesis until the round spermatid stage, at which point the majority of MILI localized in a single prominent granule, the chromatoid body, as evident from colocalization with chromatoid body marker MVH (DDX4) (Figure S1B).

MIWI2 could be detected in male germ cells beginning around 14.5–15.5 dpc but was absent from female germ cells (Figure 1B). MIWI2 was present in the nucleus as well as the cytoplasm. In the cytoplasm, MIWI2 occupied granules similar to but fewer in number than those containing MILI. From 15.5 dpc and until birth, both proteins were present in male germ cells (Figure 1B). We did not detect MILI or MIWI2 in somatic cells of the embryonic gonad. MIWI2 expression declined soon after birth, reaching undetectable levels in 4-day-old mice. Remarkably, the very short window of MIWI2 expression during male germ cell development (15.5 dpc to 3 dpp) corresponds to the time of cell-cycle arrest and de novo DNA methylation.

MILI- and MIWI2-Bound piRNA Populations during Germline Development

MILI and MIWI2 complexes were recovered from male embryonic gonads isolated from 16.5 dpc embryos. Each protein associated with piRNAs of a specific size. MILI bound ~26 nt. RNAs (Aravin et al., 2006, 2007b) and MIWI2 associated with ~28 nt RNAs (Figure 1C).

We prepared libraries of piRNAs from MILI and MIWI2 complexes at 16.5 dpc and 24–33 nt total RNAs from the same stage. Additionally, we cloned libraries from 24–33 nt RNAs and MILI complexes at days 2 and 10 after birth. The small RNA libraries were sequenced, typically yielding 2–3 million small RNA reads per library. Between 40% and 70% of sequences matched perfectly to the mouse genome, and these were considered for further analysis (Table S1). MILI- and MIWI2-associated sequences showed a normal distribution, with peaks at 26 and 28 nt, respectively (Figure 1D). The profile of total cellular small RNA suggested that MILI-bound piRNAs are slightly more abundant than MIWI2 piRNAs.

At 16.5 dpc, 47.5% of all cellular small RNAs were derived from transposon sequences (Figure 1E and Table S1). The other small RNAs represent fragments of abundant, larger noncoding RNAs (29.6%), sequences derived from unannotated genomic regions (20.3%), or exons of protein-encoding genes (2.5%). The overall fraction of transposon-derived piRNAs remained relatively stable during development from 16.5 dpc to 10 dpp. However, different types of transposons showed distinct patterns. The fraction of LTR- and LINE-derived piRNAs decreased, while SINE-derived sequences increased during development. The fraction of exon-derived small RNAs also increased substantially (from 2.5% at 16.5 dpc to 18.6% at 10 dpp). In *mili* knockout animals, LTR- and LINE-derived small RNAs were almost completely eliminated, indicating that they represent bona fide piRNAs. SINE and exon-derived small RNA were decreased in abundance but not eliminated in knockouts, suggesting that a small portion of this class derives from the degradation product of SINE-containing and genic transcripts.

Both MILI and MIWI2 associated with repeat-derived piRNAs during prenatal development (Figure 1E). However, MIWI2 demonstrated greater specificity for transposons. Seventy-six

percent of MIWI2-bound piRNAs mapped to LTR and LINE retrotransposons as compared to 45.7% for MILI-bound species. MILI complexes also contained piRNAs derived from exons of protein-coding genes (4.8%). As observed for total RNA profiles, the fraction of MILI-bound LTR and LINE piRNAs decreased and SINE and exon-derived piRNAs increased from 16.5 dpc to 10 dpp. A large fraction of total cellular small RNAs (48.7%) and MILI-bound piRNAs (60.1%) could be mapped to a unique genomic position (Figure S2). Both total small RNAs and MILI-associated piRNAs contained a subset (14%–17%) that matched highly repetitive sequences (100 or more genomic mappings). These dramatically decreased in *mili* mutants. In MIWI2 complexes, the fraction of uniquely mapped piRNAs was lower (35.6%) and the highly repetitive fraction was substantially higher (30.5%).

Overall, our analysis revealed that both MILI and MIWI2 bind transposon-derived piRNAs during embryonic development and that the fraction of piRNAs that match active transposons classes decreases after birth. Interestingly, the partners of the two Piwi proteins are different, with MIWI2 complexes being particularly enriched in transposon-derived piRNAs.

Transposon-Derived piRNAs

A substantial fraction of piRNAs matched the three major classes of transposable elements present in mammalian genomes: LTR, LINE, and SINE retrotransposons. We analyzed piRNAs derived from the representative element of each class that produced the largest number of small RNAs: the IAP LTR-retrotransposon, LINE1, and SINE B1. During development, the abundance of LINE1 piRNAs decreased and SINE increased, while IAP piRNAs showed a double peak with higher levels before birth and at 10 dpp (Figure 2A). LINE- and IAP-derived piRNAs were associated with both MILI and MIWI2 (Figure 2B). Interestingly, for both elements, piRNAs were almost equally distributed among MIWI2 and MILI complexes (Figure 2B). Sequences derived from exons were almost exclusively bound to MILI (Figure 2B).

To probe the functions of Piwi-bound small RNAs, we analyzed strand orientation of transposon- and gene-derived piRNAs. Exon-derived piRNAs were highly enriched for sense sequences. All three transposons produced substantial numbers of antisense piRNAs, with different elements showing different characteristics (Figure 2C). For LINE1, piRNAs were generally enriched in antisense sequences throughout development. In prenatal testis, MILI and MIWI2 had opposite strand preferences, with MILI binding more sense and MIWI2 binding more antisense piRNAs. Interestingly, strand orientation of LINE1 piRNA in MILI complexes was reversed after birth. IAP and SINE piRNAs were nearly equally divided between sense and antisense sequences in both complexes and in total piRNA populations in prenatal cells. However, at 10 dpp, sense IAP and SINE piRNAs became predominant. For all three transposons, the fraction of antisense sequences dramatically decreased in MILI-deficient animals.

These results indicate that LINE1 piRNAs are sorted into MILI and MIWI2 complexes according to strand orientation, similar to what is observed for Piwi family members in *Drosophila* (Brennecke et al., 2007). However, this bias was slight for SINE B1 and was absent for IAP. These data also reveal that strand

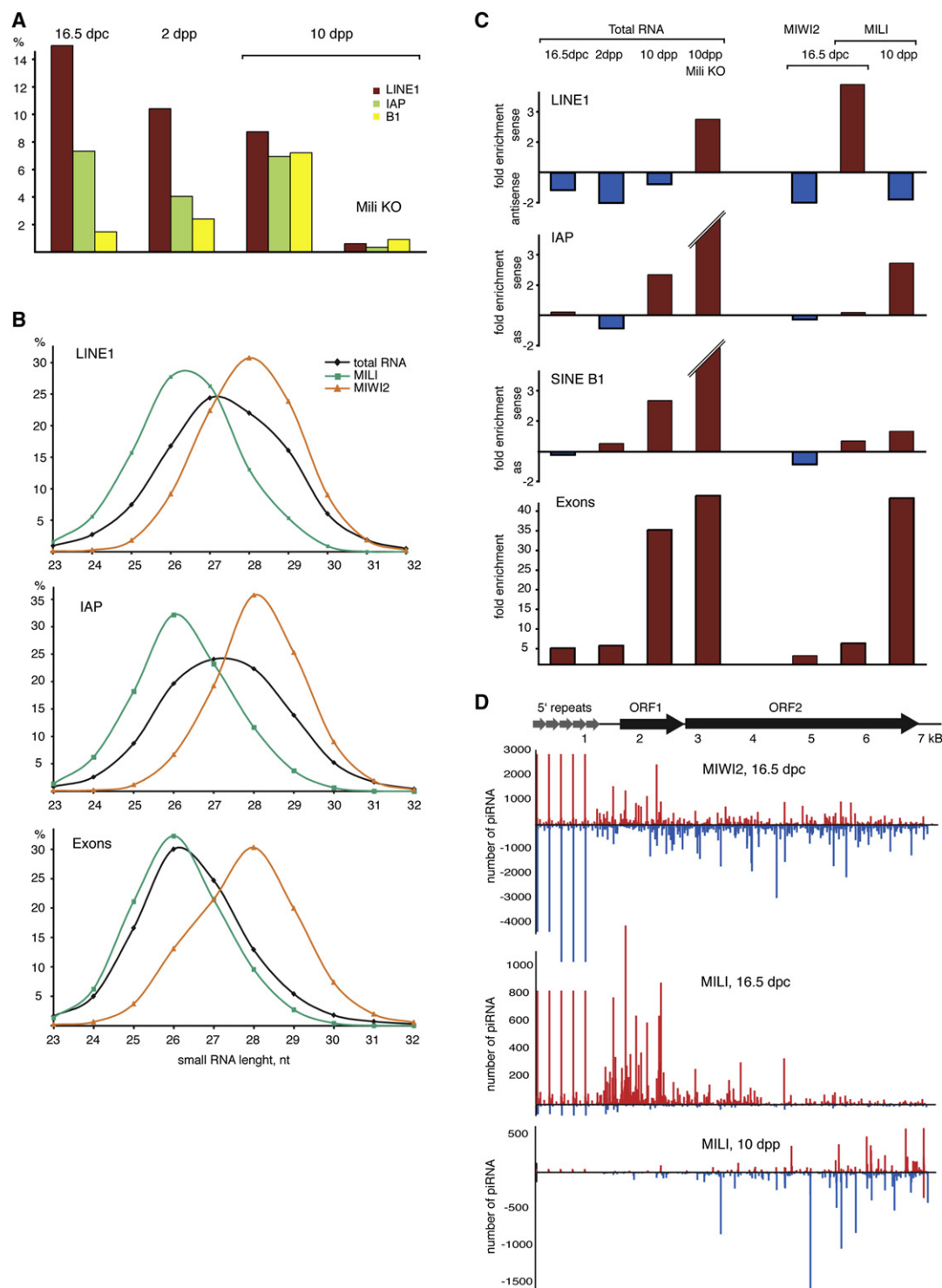


Figure 2. Repeat and Gene-Derived piRNAs

(A) Shown is the fraction of LINE1, LTR IAP, and SINE B1 piRNA in total small RNA libraries at the indicated developmental time points.

(B) The size distribution of LINE1, IAP, and exon-derived piRNAs in total RNA, MILI, and MIWI2 complexes is plotted.

(C) The strand orientation of piRNAs derived from transposable elements and exons of protein-coding genes is shown as fold enrichment for sense (red) or antisense (blue) piRNAs.

(D) The distribution of piRNAs on LINE1 retrotransposon consensus sequences is shown for the indicated piRNA libraries. piRNAs were mapped to LINE1-A consensus with up to three mismatches. The 5' end of LINE1 consists of several ~200 bp repetitive units.

bias even within a given Piwi complex can be dynamic during development.

We next analyzed the distribution of piRNAs along transposon consensus sequences (Figure 2D). Notably, MILI-bound LINE1 piRNAs were enriched in antisense and biased toward the 5' end of LINE1 in prenatal germ cells. At 10 dpp, piRNAs were less strand biased and mapped more frequently toward the 3' ends of the elements, likely reflecting the higher abundance of those sequences in the genome (Kazazian, 2004). The enrichment for piRNAs mapping to LINE1 5' ends at 16.5 dpc implies that, at this developmental stage, piRNAs are processed from full-length, potentially active copies.

The Ping-Pong Cycle in Prenatal piRNAs

Two different mechanisms, primary processing and ping-pong amplification, have been proposed to generate piRNAs (Brennecke et al., 2007; reviewed in Aravin et al., 2007a). Primary processing samples single-stranded piRNA precursor transcripts, generating a diverse set of piRNA sequences that share a preference for 5' uridine (1U). Pachytene piRNAs are exclusively produced by the primary processing mechanism. A subset of prepachytene piRNAs in mouse (Aravin et al., 2007b) and a substantial fraction of *Drosophila* piRNAs (Brennecke et al., 2007) are generated by a mechanism that depends upon the endonuclease activity of Piwi proteins and that is referred to as the ping-pong amplification cycle.

The ping-pong cycle requires the presence of transcripts that are complementary to primary piRNAs. Recognition by primary piRNAs guides the endonuclease activity of Piwi proteins, which cleave the transcript 10 nt from the 5' end of the original piRNA (Brennecke et al., 2007; Gunawardane et al., 2007). This event generates the 5' end of a new secondary piRNA. These show a strong bias for adenine at position 10 (10A), complementing the 1U bias of primary piRNAs (Figures 3A and 3B). Secondary piRNAs can also generate new piRNAs by recognizing and cleaving complementary transcripts to regenerate the piRNA that initiated the cycle. Thus, only secondary piRNAs are enriched for 10A. Although piRNAs with the sequence identical to original primary piRNA can be created during the cycle (following cleavage by a 10A secondary piRNA), these 1U-biased species remain reflective of a primary species that initiated the cycle. Thus, we consider the 1U population to be reflective of primary piRNA biogenesis.

We investigated the existence of the ping-pong cycle in prenatal germ cells and the roles of MILI and MIWI2 in this process. Tracking a specific feature of ping-pong piRNA pairs, the 10 nt offset between 5' ends of piRNAs, showed that both MILI and MIWI2 participate in the amplification cycle. The most prevalent signatures indicated MIWI2-MIWI2 and MIWI2-MILI cycles (Figure 3C).

In *Drosophila*, Piwi proteins participating prominently in the ping-pong cycle show piRNA strand specificity (Brennecke et al., 2007). We tested whether similar characteristics defined the ping-pong cycle in prenatal testis. Ping-pong pairs in which the sense strand associated with MILI and antisense with MIWI2 were more abundant than pairs with the opposite character (Figure 3D). Surprisingly, this held true not only for LINE1 piRNAs that are generally asymmetrically distributed in MILI

and MIWI2 complexes but also for IAP piRNAs that do not show a protein-dependent strand bias overall.

As was seen for strand orientation, MILI and MIWI2 complexes also discriminated primary and secondary piRNAs. We calculated the preference (primary/secondary [P/S] ratio) by taking the number of piRNAs that show 1U but no 10A bias (primary-like) and dividing it by the number that show a 10A but no 1U bias (secondary piRNAs). It should be noted that this approach ignores all sequences with both 1U and 10A, as these cannot be assigned to primary or secondary categories. Prenatal piRNAs are strongly enriched in secondary sequences as compared to pachytene piRNAs, which appear to be generated exclusively by primary processing (P/S ratios of 5.13 and 21.2, respectively). LINE1 and IAP piRNAs showed strong signals for secondary sequences as compared to exon-derived piRNAs, which, given a lack of antisense information, must be generated by primary processing (Figure 3E). MIWI2 complexes were ~2-fold enriched in secondary piRNAs as compared to MILI. Overall, our data indicate distinct roles for MILI and MIWI2 in the piRNA-processing pathway. MILI is biased toward primary piRNAs and 1U-containing piRNAs generated in the ping-pong cycle. MIWI2 is particularly enriched in secondary sequences.

Finally, we probed the correlation between the strand orientation of transposon piRNAs and their processing category (primary or secondary). In both MILI and MIWI2 complexes, antisense piRNAs were enriched for secondary sequences. MILI-associated sense piRNAs had a P/S ratio of 4.46 versus 2.02 for antisense, and MIWI2-associated sense and antisense piRNAs had P/S ratios of 2.54 and 1.40, respectively (Figure 3F).

Overall, these data are consistent with a model in which sense transcripts, most likely mRNAs of active transposons, represent the major substrate for primary processing and result in piRNAs associated with MILI. MILI-associated primary sense piRNAs recognize and cleave transcripts that contain transposon sequences in the antisense orientation and generate secondary piRNAs that join MIWI2 complexes. This is precisely opposite to the bias observed in *Drosophila*, in which primary piRNAs are mostly derived from piRNA clusters and are enriched for antisense strands while secondary piRNAs are sense.

Ping-pong relationships between MILI and MIWI2 suggest a need for physical proximity. Colocalization of MILI and MIWI2 was investigated by immunofluorescence in 17.5 dpc germ cells. Though MIWI2 was mainly present in the nucleus, MIWI2-containing cytoplasmic granules colocalized with or were in close proximity to MILI-containing granules (Figure 4A). MILI granules were more abundant as compared to MIWI2 granules, and many MILI granules did not interact with MIWI2 foci. In *miwi2*-deficient cells, MILI localization in cytoplasmic granules did not change (Figure 4B, lower panel). In contrast, in *mili* mutants MIWI2 relocalized from the nucleus to the cytoplasm, where it was uniformly distributed rather than concentrating in granules (Figure 4B, upper panel). Notably, we detected no MIWI2-associated piRNAs in *mili* mutants (Figure 4C), indicating that MIWI2 remains unloaded when MILI is absent. This epistatic relationship between MILI and MIWI2 supports the proposed directionality of their interaction, wherein MILI initiates the cycle with primary piRNAs and the production of secondary piRNAs associated with MIWI2 depends upon the prior existence of these species.

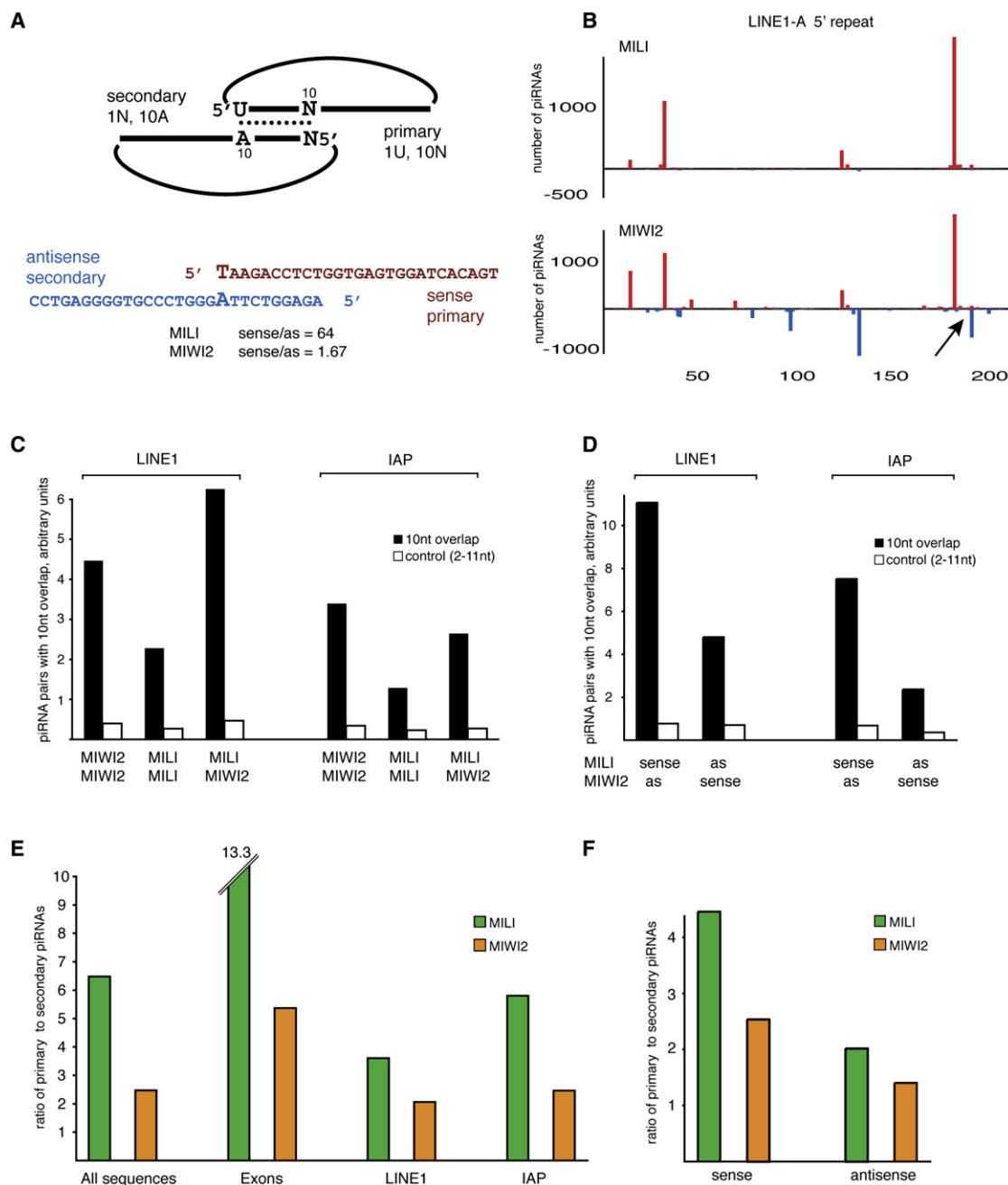


Figure 3. Ping-Pong Amplification in Prenatal piRNAs

(A) A schematic ping-pong pair is shown. piRNAs generated in the ping-pong cycle are complementary to each other and have a 10 nt offset between their 5' ends. Primary piRNAs have a bias for uridine at position 1 and do not have nucleotide bias at position 10. Secondary piRNAs, generated by primary piRNA guided cleavage, have a bias for adenine at position 10 and do not have a bias at position 1. Below is an example of an actual ping-pong pair derived from the 5' repeats of LINE1-A as shown in (B).

(B) The distribution of MILI- and MIWI2-associated piRNAs is shown across the 5' repeats of LINE1-A. MILI has a preference for sense piRNAs, while MIWI2 is bound to both sense and antisense sequences. The most prominent ping-pong pair is indicated by arrow, and its sequences are shown in (A).

(C) The ping-pong interaction between piRNAs associated with each Piwi family member was measured as the number of piRNA pairs that have a 10 nt overlap between their 5' ends normalized to the total number of sequences (arbitrary units). Comparison to a control set (overlap at position 2–11) shows that both Piwi proteins are involved in the ping-pong cycle.

(D) The extent of the ping-pong interaction measured as in (C) was calculated for sense and antisense transposon piRNAs separately.

(E) The ratio of primary (1U, no-10A) to secondary (no-1U, 10A) piRNA was calculated for MILI- and MIWI2-bound piRNAs.

(F) The correlation between strand orientation of transposon-derived piRNA and their processing category is displayed.

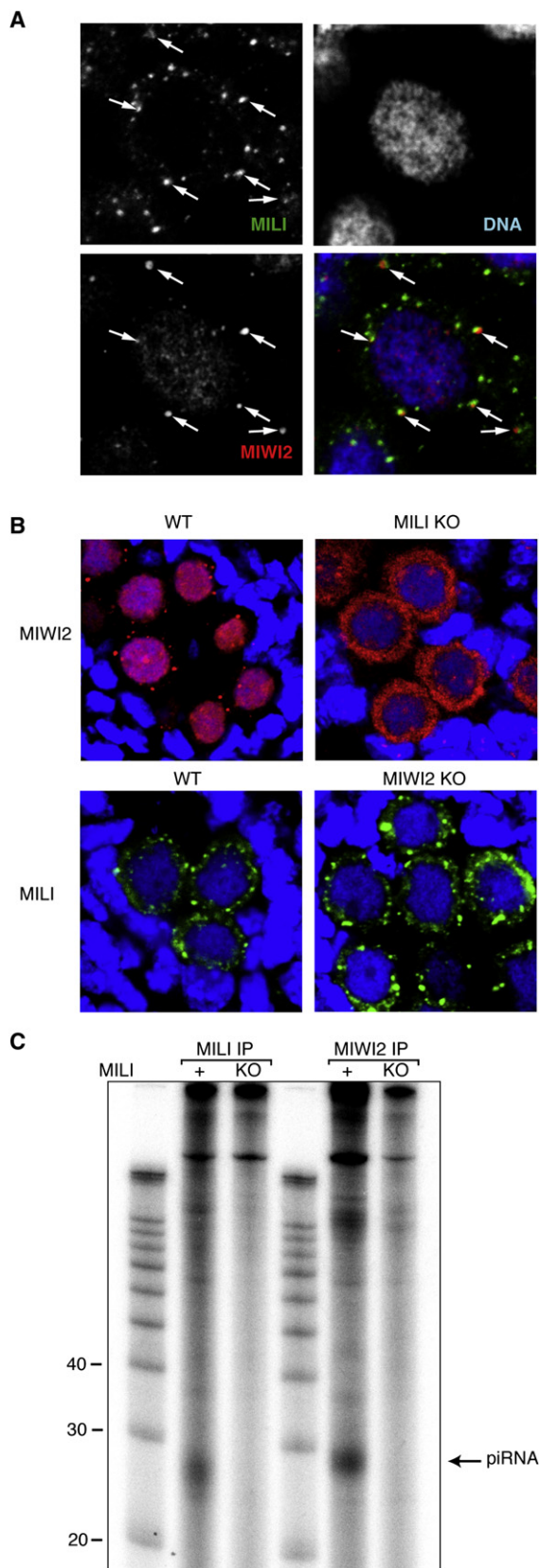


Figure 4. Interaction between MILI and MIWI2 Complexes in Germ Cells

(A) Colocalization of MILI and MIWI2 granules is shown. Detection of MILI and MIWI2 in 17.5 dpc germ cells was performed in Myc-MIWI2 transgenic animals using anti-myc and anti-MILI antibodies. Note that MILI granules are smaller but more numerous as compared to MIWI2 granules.

(B) MIWI2 localization depends on MILI. MIWI2 was detected in 17.5 dpc prenatal testes of heterozygous and homozygous *mili* embryos (upper panels). MIWI2 is present in the germ cells of *mili* mutants but almost completely delocalizes from the nucleus to the cytoplasm, where it is uniformly distributed. MILI localization in cytoplasmic granules does not change in *miwi2*-deficient animals (bottom panels).

(C) MIWI2-associated piRNA cannot be detected in *mili* mutants. MILI and MIWI2 complexes were immunoprecipitated from testis extracts prepared from 16.5–18.5 dpc heterozygous (+) and homozygous (KO) *mili* embryos. Associated small RNA species were labeled using sequential phosphatase/kinase reactions.

Genomic Origins of Prenatal piRNAs

Derivation from clustered loci in the genome has been a defining feature of piRNAs, and these loci play an important role in piRNA generation in both *Drosophila* and vertebrates (Aravin et al., 2006, 2007b; Brennecke et al., 2007; Girard et al., 2006; Lau et al., 2006). To investigate the genomic origin of prenatal piRNAs, we searched for unambiguously mapping sequences in close proximity in the genome. Using a threshold value of 10 piRNAs per kilobase, we identified 3399 clusters, which were ranked by their relative contributions to piRNA populations. Though the most prominent cluster gave rise to almost 10% of all uniquely mapped piRNAs, the individual contribution of each subsequent cluster dropped dramatically (Figure 5A). Combining all clusters yielded about 8 Mb of genomic space, which could accommodate only ~50% of all uniquely mapping piRNAs. This suggests fundamental differences from the pachytene piRNAs (Aravin et al., 2006; Girard et al., 2006) and even those prepachytene piRNAs that occupy *mili* at 10 dpp (Aravin et al., 2007b). In fact, most prenatal clusters actually represent individual transposons and transposon fragments. Expression from all clusters was eliminated in *mili* mutants (Figure 5B).

Analyzing the expression of clusters that gave rise to the greatest numbers of piRNAs indicated developmentally regulated expression patterns. Only cluster #2 on chromosome 10 persisted in its expression through day 10 dpp. Interestingly, this cluster was also highly expressed in ovary (Figure 5B).

Unlike piRNA clusters identified in mammals thus far, six out of the eight most prominent prenatal clusters generate piRNAs from both genomic strands (double-strand clusters). This arrangement is very similar to piRNA clusters in *Drosophila* ovary (Brennecke et al., 2007). The two most prominent clusters (#1 chr. 7 and #2 chr. 10) show profound strand asymmetry, with the vast majority of piRNAs being derived from one genomic strand (single-strand clusters). As with *Drosophila flamenco*, these clusters also showed enrichment for similarly oriented transposon fragments and for the generation of piRNAs that are antisense to transposon mRNAs (Figure 5C). Surprisingly, double-stranded clusters are not particularly enriched in transposon sequences. Both single-strand and double-strand clusters produced piRNAs in MILI and MIWI2 complexes (Figures 5C and 5E). However, even in double-strand clusters, MILI-bound piRNAs were strongly biased for one of the two genomic strands (Figure 5E).

Size profiles of total cellular piRNAs indicated that single-strand clusters produced significantly more MILI-bound piRNAs (Figure 5D), while double-strand clusters contributed to MILI and MIWI2 complexes equally (Figure 5F).

piRNAs derived from clusters participate in the ping-pong cycle (data not shown). For cluster-derived piRNAs, MIWI2 remained biased toward secondary piRNAs as compared to MILI (Figure 5G). Single-strand clusters produced more primary piRNAs in both MILI and MIWI2 complexes than did double-strand clusters. Indeed, the P/S ratio of MILI-bound piRNAs derived from the single-strand cluster on chr. 7 (16.21) was close to that of pachytene piRNAs (21.2).

If ping-pong amplification operates by interaction of piRNAs from single-strand clusters with transposon mRNAs in *trans*, MIWI2, which is enriched in secondary piRNAs, should contain more transposon-derived piRNAs than MILI. Indeed, for the two single-strand transposon-rich clusters (#1 chr. 7 and #2 chr. 10), MIWI2 was particularly enriched in transposon-derived piRNAs (Figure 5H, left panel). The difference between MILI and MIWI2 complexes was even more dramatic when strand orientation was taken into account. For both clusters, MIWI2 was enriched in piRNAs that match transposons in the antisense orientation (Figure 5H, right panel). These data indicate that the ping-pong cycle strongly shapes piRNA populations and enhances the production of piRNAs that match transposons in the antisense orientation.

DNA Methylation and Piwi/piRNA Pathway

DNA methylation is critical to stable, epigenetically inherited silencing of transposons, and this is lost upon mutation of either MILI or MIWI2 (Aravin et al., 2007b; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008). DNMT3L, a catalytically defective member of the DNA methyltransferase family, is essential for both proper transposon methylation and transposon repression (Bourc'his and Bestor, 2004). DNMT3L acts together with the catalytically active de novo methyltransferases DNMT3A and DNMT3B to establish methylation patterns (Chen et al., 2005; Gowher et al., 2005; Suetake et al., 2004). We sought to order MILI and MIWI2 with respect to other pathway components. Since DNMT3L acts as a coordinator of methylation activities in the male germ line, we examined the integrity of the piRNA pathway in *dnmt3L*-deficient animals.

We immunoprecipitated MILI from testes of 10 dpp *dnmt3L*-deficient animals and their wild-type littermates and found that piRNAs were still present in mutant animals (Figure 6A). Cloning and analysis of small RNA libraries showed that the fraction of LTR and LINE retrotransposon piRNAs increased, while the fraction of SINE piRNAs decreased in *dnmt3L* mutants (Figure 6B, left panel). This was consistent with northern blotting for an abundant IAP-derived piRNA, which also increased in abundance in the mutant (Figure 6C).

Several lines of evidence indicated that increases in LTR and LINE piRNAs might be linked to derepression of these elements and the increased capacity of transposon mRNAs for entry into the piRNA pathway. First, LINE and LTR elements lost methylation and silencing, showed increased expression, and contributed a greater number of sense piRNAs in *dnmt3L* mutants. In contrast, SINEs (B1) were neither affected at the level of DNA

methylation and expression nor contributed increasingly to piRNA populations in mutant animals (A.A.A. and G.J.H., unpublished data). Moreover, LINE and LTR piRNAs that increased in *dnmt3L* mutants corresponded to elements that are close to consensus and thus potentially expressed under circumstances in which methylation was lost (Figure 6B, right panel).

As an example, piRNAs derived from IAP elements showed the greatest change in *dnmt3L* mutants, and this was mainly due to a dramatic increase in sense small RNAs (Figures 6C and 6D). This pattern was also obvious when the distribution of piRNAs along the IAP consensus was displayed (Figure 6E). Normally, the ratio of primary to secondary sequences in MILI complexes decreases after birth and approaches that of MIWI2 in prenatal germ cells (Figure 6F, compare to Figure 3E). In *dnmt3L*-deficient animals, the ratio of primary to secondary IAP piRNA increased ~6-fold as compared to wild-type. This strongly supports the model in which mRNAs from IAP elements, which show increased expression in *dnmt3L* mutants, flow into the piRNA pathway as a source of primary piRNAs.

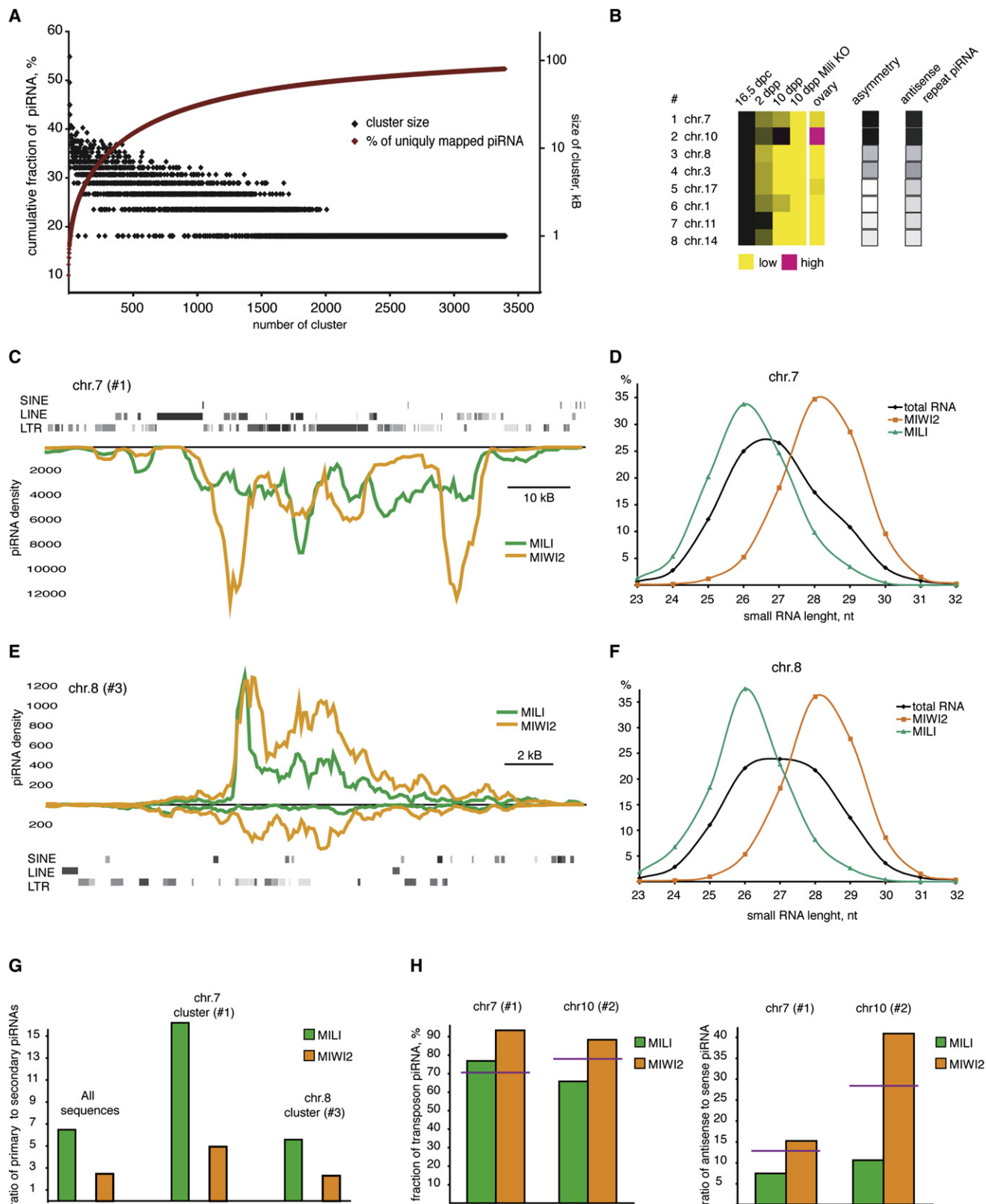
Considered, these data support a model in which the piRNA pathway acts upstream of DNMT3L, and consequently DNMT3A and DNMT3B, to help establish patterns of DNA methylation on repeat elements. This interpretation rests on the observation that while mutations in methyltransferase family members impact cytosine methylation, they leave the piRNA pathway largely intact, affecting only the composition of small RNA populations in a manner that can be rationalized by the impact of loss of methylation on transposon expression. In contrast, loss of the piRNA pathway prevents the recognition and silencing of potentially active transposons by the DNMT3L pathway.

DISCUSSION

Dynamic Expression of Piwis and piRNAs

The mouse genome encodes three Piwi proteins. MIWI is expressed beginning at the pachytene stage of meiosis (Deng and Lin, 2002) and together with MILI binds to nonrepetitive pachytene piRNAs whose function remains unknown (Aravin et al., 2006; Girard et al., 2006). MIIL1 is expressed very early in germ cell development and persists through the completion of meiosis. Deficiency in either MILI or MIWI2 causes increased transposon expression and similar defects in meiosis and germ cell survival (Aravin et al., 2007b; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008), suggesting their cooperation in the same pathway. Results reported herein support an intimate connection between MILI and MIWI2 in transposon control.

MIWI2 protein had not previously been detected, leaving its developmental expression profile and subcellular localization unknown. We found that both MILI and MIWI2 are expressed in prospermatogonia during embryogenesis. MILI can be detected in both male and female germ cells starting at 12.5 dpc, soon after PGCs arrive at the embryonic gonad, and continues to be expressed after birth throughout both oogenesis and spermatogenesis. In contrast, *miwi2* expression can be detected exclusively in male germ cells in a window from 15 dpc to soon after birth. The timing of *miwi2* expression precisely corresponds to the stage of cell-cycle arrest when prospermatogonia establish de novo methylation patterns (Lees-Murdock et al., 2003;



Trasler, 2006; Walsh et al., 1998). Furthermore, it coincides with the time when expression of LINE1 elements (Trelogan and Martin, 1995) and an IAP transgene (Dupressoir and Heidmann, 1996) can be detected in wild-type testes. Overall, our data show that two Piwi members are coexpressed in primordial male germ cells at the moment that is critical for establishing epigenetically stable silencing of mobile elements.

Developmental profiling revealed unexpectedly dynamic piRNA populations. Previously we reported that the expression of abundant, repeat-poor piRNAs, pachytene piRNAs, during meiosis is preceded by the appearance of another transposon-enriched piRNA population, prepachytene piRNAs (Aravin et al., 2007b). Here we showed that the composition of the prepachytene piRNA pool changes throughout development, even though at least one binding partner, MILI, is expressed throughout the entire window examined. Changes are observed in the types of transposons targeted, the strand bias of populations, the distribution of piRNAs along transposon consensus sequences, and the degree to which primary and secondary piRNAs populate the pathway.

It is likely that the dynamics of piRNA populations reflect a complex interplay between the expression of Piwi proteins, transposable elements, and piRNA clusters. Particularly, we note a switch from the “MILI/MIWI2” ping-pong cycle that operates before birth to one containing MILI alone after birth. Since amplification of piRNAs in the ping-pong loop depends on the presence of transcripts from active elements, piRNA populations should be greatly influenced by the expression level of transposons. Accordingly, we found the highest fraction of LINE and IAP-derived piRNAs at 16.5 dpc, the time at which both elements are actively expressed (Dupressoir and Heidmann, 1996; Trelogan and Martin, 1995). Finally, the expression of piRNA clusters changes during development. The majority of clusters expressed in prenatal testes no longer produce piRNAs postnatally, at which point a different set of clusters predominates. Overall, our data highlight the highly dynamic nature of piRNA populations that are shaped to provide an adequate response to active transposable elements.

The piRNA Ping-Pong Cycle

Analysis of piRNA sequences in both flies and mouse indicates the existence of two classes of piRNAs, primary and secondary (reviewed in Aravin et al., 2007a). Most primary piRNAs in *Drosophila* are antisense to transposons and are derived from few large piRNA clusters that contain a diversity of transposon fragments (Brennecke et al., 2007). In mice, during prenatal development, the situation appears to be quite different. We find that primary piRNAs in mouse are predominantly sense oriented with respect to transposon transcripts and appear to derive mainly from transposon mRNAs instead of from piRNA clusters. The processing of primary piRNAs from transposon mRNAs is supported by several observations. Primary LINE1 piRNAs are biased toward the 5' end of this element, suggesting that they are generated from transcripts of active full-length LINE1 elements; the majority of LINE1 fragments in the genome are inactive, 5' truncated copies (Kazazian, 2004). Furthermore, increases in sense, primary IAP piRNAs are observed in *dnmt3L* mutants that derepress these elements. These observations indicate that transposon mRNAs can be efficiently used as substrates for primary piRNA processing. Though long piRNA clusters do exist in mouse prenatal germ cells, dispersed transposon copies outweigh their contribution to the system.

Currently, it is unclear how transposon transcripts are distinguished from cellular mRNAs as an input into the piRNA system. An alternative hypothesis is that they are indeed not distinguished but that the transcriptome is instead essentially sampled at random, with only those sequences that encounter complementary antisense transcripts truly engaging the pathway and increasing their relative abundance through a ping-pong cycle. This hypothesis is supported to some degree by the prevalence of piRNAs from exons of normal protein-coding genes. These genic piRNAs are almost exclusively primary and might be considered as by-products of a random sampling of transcriptome for piRNA generation. In contrast, piRNAs from transposon-poor double-stranded piRNA clusters might represent by-products of the ping-pong amplification acting on loci in the genome that happen to be transcribed bidirectionally. In this case,

Figure 5. Genomic Origins of Prenatal piRNAs

(A) Prenatal MIWI2 piRNA clusters were identified by scanning the genome using a 1 kb sliding window to find loci that produce at least 10 uniquely mapped piRNAs per kb. More than 3000 piRNA clusters were thus identified and arranged by the number of uniquely mapped piRNAs that they produce from left to right. Shown are the genomic size of each cluster (black diamonds) and the cumulative fraction of piRNAs contributed by clusters (red curve).

(B) Expression patterns and features of the eight most prominent piRNA clusters. Expression is calculated as a fraction of cluster-derived piRNAs in total RNA populations normalized for expression level at 16.5 dpc. Also shown is the genomic strand orientation of piRNAs produced by each cluster (strand asymmetry) and the fraction of cluster-derived piRNAs that matches the antisense strand of transposons. piRNA density and size profiles of piRNAs for clusters #1 and #3 are shown in (C) and (D) and (E) and (F), respectively.

(C) piRNA density is shown for the most prominent piRNA cluster (chr. 7, 6526000–6612000). The cluster spans ~70 kb and is enriched in transposon sequences (LINE and LTR). The majority of transposons are located on the plus genomic strand, and piRNAs are exclusively derived from the minus strand. Therefore, the majority of cluster-derived piRNAs are antisense to transposons.

(D) The size profile of piRNAs derived from cluster #1 is shown.

(E) piRNA density is plotted for cluster #3 (chr. 8, 48701000–48723000). This cluster is not significantly enriched in transposon sequences.

(F) The size profile of piRNAs derived from cluster #3 is shown.

(G) Processing features are displayed for piRNAs derived from clusters. The ratio of primary (1U, no-10A) to secondary (no-1U, 10A) piRNAs is shown for total MILI- and MIWI2-bound populations and for piRNAs uniquely mapped to two piRNA clusters (#1, single-strand, [C], and #3, double-strand, [E]). For the double-strand cluster #3, the ratio of primary to secondary piRNAs in both complexes is similar to that of total population. For the single-strand cluster #1, both MILI and MIWI2 are enriched in primary piRNAs.

(H) Shown are fractions of repeat-derived piRNAs (left panel) and their sense/antisense ratio (right panel) for the two most prominent piRNA clusters. The genomic sequences of both clusters are enriched in transposable elements. The expected level of repetitive piRNAs and their sense/antisense ratios are shown (red line) if cluster sequences are randomly sampled.

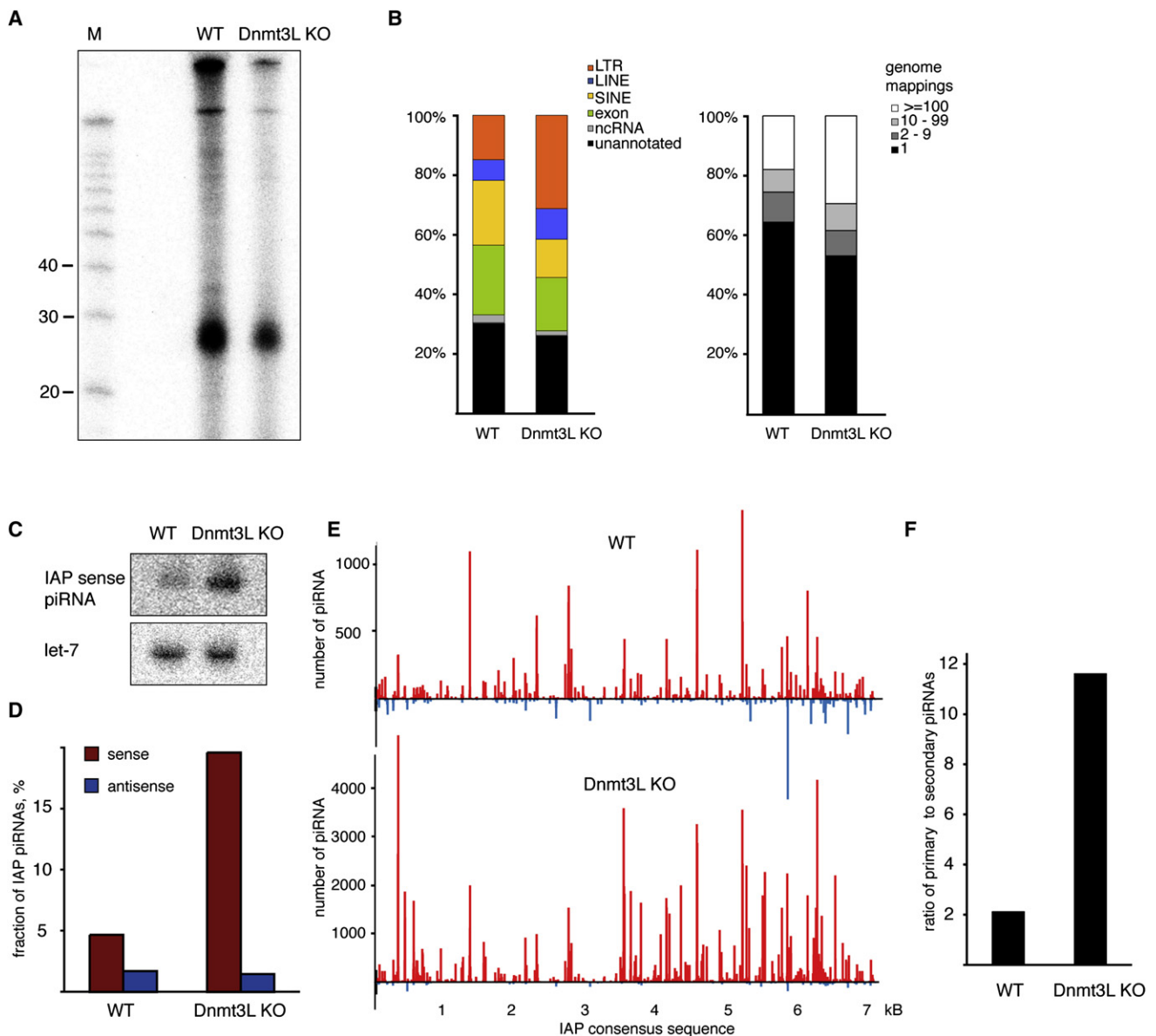


Figure 6. Links between the DNA Methylation and piRNA Pathways

(A) MILI-piRNA complexes were immunoprecipitated from testes of 10 dpp wild-type and *dnmt3L* knockout animals, and isolated RNAs were 5' labeled. (B) piRNA populations were analyzed in *dnmt3L* mutants. piRNAs isolated from MILI complexes shown in (A) were cloned, sequenced, and annotated. Shown are the annotation (left panel) and genomic mapping (right panel) for RNAs from wild-type and *dnmt3L* mutants. (C) Northern hybridization with LNA probe to detect an abundant IAP sense piRNA in 10 dpp testes of wild-type and *dnmt3L* KO mice. Hybridization with let-7 miRNA was used as a loading control. (D) The fraction of sense and antisense IAP piRNAs is shown for wild-type and *dnmt3L* mutant libraries. (E) The distribution of piRNAs on the IAP retrotransposon consensus is shown for MILI complexes from wild-type and *dnmt3L* mutant animals. (F) Primary and secondary IAP piRNA in *dnmt3L* mutants. Shown is the ratio of primary (1U, no-10A) to secondary (no-1U, 10A) piRNAs in the indicated libraries.

complementary transcripts are readily available in *cis*, allowing robust operation of the ping-pong cycle in the absence of transcripts in *trans*.

Overall, our data indicate that in mouse the initiating input for the ping-pong cycle is different than in *Drosophila*. In mouse, primary processing of mRNAs of active elements generates sense piRNAs that start the cycle, while in *Drosophila* primary piRNAs

are often antisense to transposons and are processed from few specialized piRNA clusters (Figure 7). Recognition of transposon mRNAs as a source for primary processing and an independence from piRNA clusters might allow greater flexibility in response to expansion of transposable elements in mammals. However, it also poses the problem of how transposons are distinguished from host genes.

Specialized Functions of MILI and MIWI2 Proteins

Both the content of their associated piRNA populations and their intracellular location suggest that MILI and MIWI2 play quite specialized roles in the piRNA pathway and in transposon silencing. Several lines of evidence suggest that MILI might be the principal recipient of piRNAs generated by the primary processing mechanism. MILI complexes are enriched for 1U-containing species, and exonic piRNAs for which no ping-pong partners are found populate MILI complexes almost exclusively. Single-strand clusters, which appear to be processed mainly by the primary mechanism, contribute disproportionately to MILI. Finally, MILI seems to act upstream of MIWI2, consistent with its being a recipient of primary piRNAs.

Subcellular compartmentalization of these proteins also supports functional specialization. MIWI2 is found predominantly in the nucleus, while MILI is cytoplasmic. The impacts of mutations in this cytoplasmic family member on DNA methylation can be explained by downstream effects on MIWI2. Indeed, we found that MIWI2 lacks piRNA partners and is lost from the nucleus in MILI mutants. A subset of MIWI2 is cytoplasmic and is found in or near MILI-containing granules. This colocalization is consistent with MILI and MIWI2 acting as partners in the ping-pong cycle.

Interactions between Piwi/piRNA and Chromatin/DNA Methylation Pathways

During the development of PGCs into prospermatogonia, methylation on mobile elements is erased and re-established (Trasler, 2006). We and others previously showed that mutation of either MIWI2 or MILI results in substantial demethylation and derepression of IAP and LINE1 elements (Aravin et al., 2007b; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008). Results presented here implicate these proteins in establishment rather than in maintenance of methylation patterns and support the idea that piRNAs serve as guides that direct the DNA methylation machinery to transposon sequences. This conclusion is supported not only by the nuclear localization of MIWI2 at the critical time when methylation patterns are established but also by indications from small RNA sequencing that both MILI and MIWI2 act upstream of DNMT3L, which in turn acts upstream of the de novo methyltransferases DNMT3A and 3B.

Overall, our data can be synthesized into a model (Figure 7) in which loading of MIWI2 complexes with piRNAs depends on MILI function. This occurs through a ping-pong cycle that functions in cytoplasmic granules reminiscent of nuage and results in MIWI2 becoming enriched in antisense piRNAs targeting active elements. MIWI2 likely shuttles between the cytoplasm, where it acts in the ping-pong cycle, and the nucleus, where the majority of protein appears. Here it may recognize chromatin-bound nascent primary transcripts from active elements. Using coimmunoprecipitation from embryonic testes, we did not detect interaction between MIWI2 and de novo methyltransferases, Dnmt3a and Dnmt3b (data not shown); therefore, it is unlikely that piRNA-containing complexes directly bring DNA methyltransferases to their targets. With reference to pathways that are becoming increasingly well understood, particularly in *S. pombe* (Verdel and Moazed, 2005), piRNA complexes, bound to nascent transcripts, may simultaneously cleave those RNAs

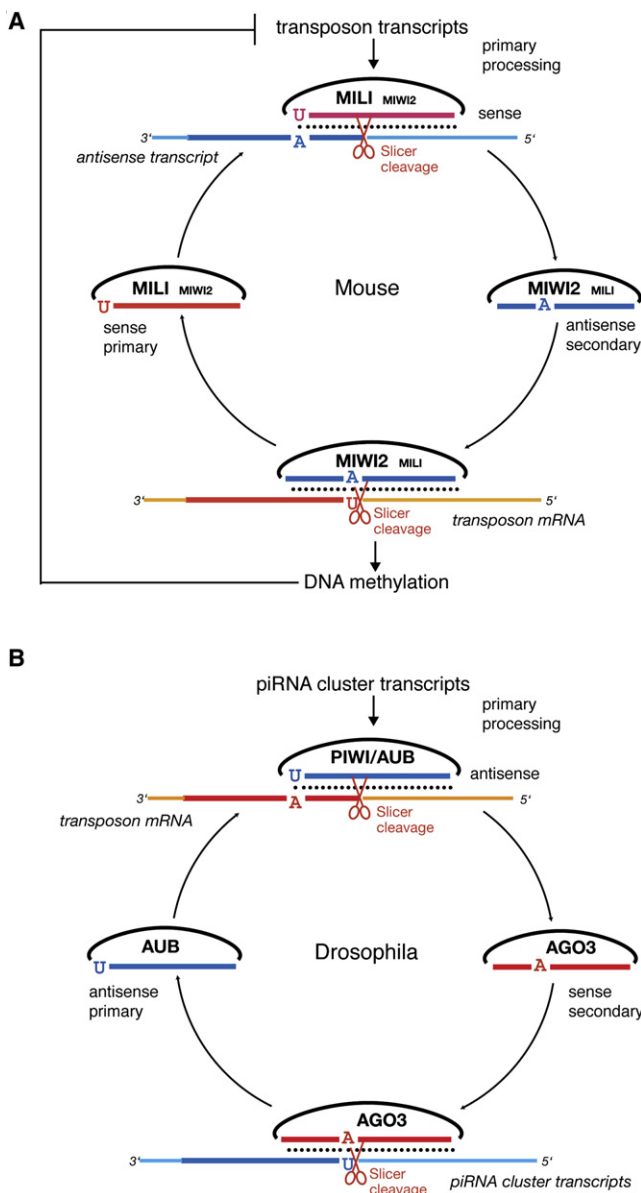


Figure 7. A Comparison of Ping-Pong Amplification in Mouse and Fly

The mouse (A) ping-pong cycle is reversed as compared to *Drosophila* (B). In *Drosophila*, primary processing of long transcripts derived from piRNA clusters produces both sense and antisense piRNAs that enter a ping-pong cycle that involves AGO3 and AUB. AGO3 primarily binds sense secondary piRNAs, and AUB binds primary antisense piRNAs. In mouse piRNA clusters are not the major source of primary piRNAs (Figure 5A). mRNAs of active transposable elements likely represent the substrate for primary processing resulting in sense piRNAs that preferentially associate with MILI. In prenatal testes, both MILI and MIWI2 participate in the amplification cycle. MIWI2 is specifically enriched in secondary antisense piRNAs as compared to MILI. Antisense piRNAs guide DNA methylation of transposable elements sequences in the nucleus, probably through recognition of nascent transposon transcripts. After birth, when MIWI2 is no longer expressed, MILI continues to operate the cycle alone. If DNA methylation of transposon sequences is impaired due to downstream mutations in methyltransferase proteins, overexpression of transposon transcripts boosts primary processing and increases the fraction of primary sense piRNAs.

and recruit chromatin-modifying enzymes to establish epigenetically stable repression marks that eventually lead to de novo DNA methylation. However, the precise pathway through which small RNAs act to induce the deposition of methylation marks remains a mystery at present.

EXPERIMENTAL PROCEDURES

Animals and Transgenic Constructs

CD-1 wild-type mice were purchased from Charles River Lab. The *miwi2* and *dnmt3L* knockout strains are described in Carmell et al. (2007) and (Bourc'his and Bestor (2004), respectively. The *mili* knockout strain was obtained from Haifan Lin (Yale University) and is described in Kuramochi-Miyagawa et al. (2004). Transgenic animals expressing tagged MILI and MIWI2 were produced by BAC recombineering (see the Supplemental Data).

Antibodies

Rabbit polyclonal antibodies against MILI were described in Aravin et al. (2007b). Synthetic peptides (N-MSGRARVRARGIC-C and N-GITGHSARE VGRSSRD-C) were used to raise rabbit polyclonal antibodies against MIWI2.

MILI and MIWI2 Detection

To detect GFP-fused MILI and MIWI2, dissected testes were fixed for 10–15 min in 4% paraformaldehyde in PBS, washed in PBS, and mounted in 30% glycerol in PBS. For immunofluorescence, testis were fixed in either Bouin's solution or 4% paraformaldehyde (for details, see the Supplemental Data).

Simultaneous detection of MILI and MIWI2 in germ cells was performed using slides prepared from 3xMyc-MIWI2 transgenic animals using anti-myc (clone 4A6, Upstate) and anti-MILI antibodies.

Small RNA Libraries

MILI and MIWI2 ribonucleoprotein complex immunopurification, RNA isolation, library construction, and annotation were performed as described previously (Aravin et al., 2007b). For ovarian libraries, 24–33 nt total RNA was isolated from ovaries dissected from CD1 females.

ACCESSION NUMBERS

Sequences reported in this manuscript are available in the Gene Expression Omnibus under accession number GSE12757.

SUPPLEMENTAL DATA

The Supplemental Data include Supplemental Experimental Procedures, two figures, and one table and can be found with this article online at <http://www.molecule.org/cgi/content/full/31/6/785/DC1/>.

ACKNOWLEDGMENTS

We thank Sang Yong Kim (Cold Spring Harbor Laboratory, CSHL) for generating transgenic animals. We thank Maria Mosquera, Lisa Bianco, Jodi Coblenz, and Gula Nourjanova (CSHL) for animal assistance and histology. We thank Stephen Hearn (CSHL) for microscopy assistance and Emily Hodges, Michelle Rooks, Dick McCombie, Dana Rebolini, and Laura Cardone for help with Illumina sequencing. We thank Catherine Schlingheyde for help with experiments and members of the Hannon laboratory, especially Antoine Molaro, for comments on the manuscript. G.J.H. is an investigator of the Howard Hughes Medical Institute. This work was supported by grants from the National Institutes of Health (NIH) to G.J.H. and an NIH Pathway to Independence Award K99HD057233 to A.A.A.

Received: June 2, 2008

Revised: September 5, 2008

Accepted: September 10, 2008

Published: September 25, 2008

REFERENCES

- 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.
- Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007a). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318, 761–764.
- Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007b). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 316, 744–747.
- Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96–99.
- 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.
- 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, 503–514.
- Chen, Z.X., Mann, J.R., Hsieh, C.L., Riggs, A.D., and Chedin, F. (2005). Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. *J. Cell. Biochem.* 95, 902–917.
- Deng, W., and Lin, H. (2002). miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* 2, 819–830.
- Dupressoir, A., and Heidmann, T. (1996). Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice. *Mol. Cell. Biol.* 16, 4495–4503.
- Girard, A., and Hannon, G.J. (2008). Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol.* 18, 136–148.
- 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.
- Gowher, H., Liebert, K., Hermann, A., Xu, G., and Jeltsch, A. (2005). Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J. Biol. Chem.* 280, 13341–13348.
- 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.
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* 117, 15–23.
- 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., et al. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129, 69–82.
- Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* 449, 248–251.
- Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429, 900–903.
- Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.* 16, 2272–2280.
- Kazanian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. *Science* 303, 1626–1632.
- Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., et al. (2004). Mili, a mammalian

member of piwi family gene, is essential for spermatogenesis. *Development* 131, 839–849.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* 22, 908–917.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* 35, 88–93.

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, 363–367.

Lees-Murdock, D.J., De Felici, M., and Walsh, C.P. (2003). Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage. *Genomics* 82, 230–237.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714–717.

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.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J. Biol. Chem.* 279, 27816–27823.

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, 534–538.

Trasler, J.M. (2006). Gamete imprinting: setting epigenetic patterns for the next generation. *Reprod. Fertil. Dev.* 18, 63–69.

Trelogan, S.A., and Martin, S.L. (1995). Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc. Natl. Acad. Sci. USA* 92, 1520–1524.

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.

Verdel, A., and Moazed, D. (2005). RNAi-directed assembly of heterochromatin in fission yeast. *FEBS Lett.* 579, 5872–5878.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* 20, 116–117.