Roles and biogenesis pathways of endogenous siRNAs and piRNAs in mouse germline cells

Toshiaki Watanabe

DOCTOR OF PHILOSOPHY

Department of Genetics

School of Life Scieence

The Graduate University for Advanced Studies

2008

Contents

Abstract

Small RNAs ranging in size between 20 and 35 nucleotides (nt) are found in many organisms including yeasts, plants, and animals. Small RNAs are associated with Argonaute proteins, which is effector of silencing, and involved in the regulation of gene expression through translational repression, mRNA degradation, and chromatin modification in a sequence dependent manner. In mammals, three classes of small RNAs have been found; microRNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi interacting RNAs (piRNAs). piRNAs and siRNAs are expressed only in specific tissues, mainly in germline cells. In contrast, miRNAs are ubiquitously expressed in many tissues.

In the first chapter, I described the results obtained from the analysis of fetal piRNAs. piRNAs are 24-30 nt small RNAs that associate with Piwi proteins, which is a subfamily of Argonaute proteins. piRNAs and Piwi proteins are mainly expressed in germline and implicate in germline development, silencing of retrotransposon. DNA methylation is also known as one of the mechanisms that repress retrotransposons. However, the mechanisms that determine the retrotransposon sequence to be subjected to DNA methylation have been unknown. Recent study in mouse shows that DNA methylation of the 5'-untranslated region of LINE-1 was decreased in Piwi KO, raising the possibility that DNA methylation is mediated by piRNAs. Though, piRNAs

expressed in neonatal and adult testes have been identified, piRNAs in fetal testes, when *de novo* methylation takes place, have remained unknown. To examine the link between piRNAs and DNA methylation and to know the possible candidate of piRNA-mediated DNA methylation, I analyzed piRNAs from fetal testes.

I cloned more than 100,000 small RNAs from fetal testes. A distinct set of piRNAs were expressed in fetal testes and most of them were derived from retrotransposons. L1Md and IAP1 retrotransposons, the DNA methylation levels of which have been known to be reduced in Piwi KO, were major classes in fetal piRNAs. This result indicates that piRNAs define genomic sequences that are subjected DNA methylation. To further investigate the link between methylation and piRNAs, I generated and analyzed *Zucchini* KO mice. *Zucchini* gene has been thought to be involved in the piRNA pathway in *Drosophila*. In *Zucchini* mutant mice, the expression level of piRNAs were decreased dramatically and methylation levels of L1Md and IAP1 were also decreased. Furthermore, decreased methylation was also observed at a specific region of *Rasgrf1* locus. To this region, piRNAs expressed in fetal testes were specifically mapped and RNA transcribed from *Rasgrf1* locus were targeted by these piRNAs. The results of my study support the piRNA-mediated DNA methylation in fetal mouse testes and help to understand the mechanism of piRNA-mediated DNA methylation.

In the second chapter, I described the results obtained from the analysis of oocyte endogenous siRNAs. siRNAs are generated from long double-stranded RNAs (dsRNAs) by *Dicer* and are mainly involved in defense against molecular parasites including viruses, transposons, and transgenes through RNAi in Plants and worms. RNA dependent RNA polymerase (RdRP) is involved in the generation of precursor dsRNAs. Gene regulation by endogenous siRNAs has been observed only in organisms possessing RdRP. Despite no report of RdRP activity in mammalian cells, endogenous siRNA molecules has been observed in mouse fully grown oocytes. However, only a small number of endogenous siRNAs has been identified and their biogenesis and function largely remain unclear. In order to obtain a comprehensive picture of endogenous siRNAs, I have analyzed small RNAs from mouse growing oocytes through deep sequencing.

I identified a large number of both ~25-27 nt Piwi-interacting RNAs (piRNAs) and ~21 nt siRNAs corresponding to mRNAs or retrotransposons in growing oocytes. piRNAs in oocytes play a role in the regulation of retrotransposons. siRNAs were exclusively mapped to retrotransposons or other genomic regions that produce transcripts capable of forming dsRNA structures. Inverted repeat structures, bidirectional transcription and antisense transcripts from various loci are sources of the dsRNAs. Some precursor transcripts of siRNAs were derived from expressed

pseudogenes, suggesting that one role of pseudogenes is to adjust the level of the founding source mRNA through RNAi. Loss of *Dicer* or *Ago2* resulted in decreased levels of siRNAs and increased levels of retrotransposon and protein-coding transcripts complementary to the siRNAs. Thus, RNAi pathway regulates both protein-coding transcripts and retrotransposons in mouse oocytes.

The results obtained here establish the existence of endogenous siRNAs in mammals and provide new insights into the pathway and function of small RNAs in mouse germline.

Abbreviations

dpc: days postcoitum

dsRNA: double-stranded RNA

IAP: Intracisternal A Particle

IP: Immuno Precipitated

KO: Knock Out

kbp: kilo base pairs

L1: LINE-1

LINE: Long Interspersed Nuclear Elements

LTR: Long Terminal Repeat

miRNA: microRNA

P: Postnatal day

PGCs: Primordial Germ Cells

piRNA: Piwi-interacting RNA

rasiRNAs: repeat-associated siRNAs

RdRP: RNA dependent RNA polymerase

RNAi: RNA interference

SINE: Short Interspersed Nuclear Elements

siRNA: small interfering RNA

Gene symbols

Ago: Argonaute

Gapdh: Glyceraldehyde-3-phosphate dehydrogenase

Kif4: Kinesin family member 4

Mael: Maelstrom

Mili: Miwi like

Miwi: Mouse Piwi-like

Mvh: Mouse vasa homologue

Pdzd11: PDZ domain containing 11

Ppp4r1: Protein phosphatase, regulatory subunit 1

Rangap1: RAN GTPase activating protein 1

Rasgrf1: Ras protein-specific guanine nucleotide-releasing factor1

Scp3: Synaptonemal complex protein 3

Tdrd1: Tudor domain containing protein 1

Zp3: Zona pellucida glycoprotein 3

Zuc: Zucchini

Introduction

RNA interference (RNAi) is a sequence-specific gene regulatory mechanism conserved among diverse eukaryotes. The sequence specificity in RNAi is determined by a family of 18–30 nt-regulatory small RNAs. In RNAi pathway, small RNAs are incorporated into Argonaute proteins. Argonaute proteins, also known as PAZ Piwi domain (PPD) proteins, are critical players in the RNAi pathway, effecting transcriptional and post-transcriptional gene regulation. In Eukaryotes, three classes of endogenous small RNAs have been characterized: microRNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi interacting RNAs (piRNAs). miRNAs, which are the best-characterized endogenous small RNAs in eukaryotes, have been identified in diverse plants and animals, and are mainly involved in development and differentiation. miRNAs are processed by an enzyme called *Dicer* from miRNA precursors (pre-miRNAs) with a stem-loop structure and regulate gene expression through translational repression or mRNA cleavage (Lee et al. 1993; Bartel 2004). siRNAs are generated from long double-stranded RNAs (dsRNAs) by *Dicer* or transcribed from single strand RNA templates by RNA dependent RNA polymerase (RdRP) and are mainly involved in defense against molecular parasites including viruses, transposons, and transgenes through RNAi in Plants and worms (Hamilton and Baulcombe 1999; Tabara et al. 1999; Pak and Fire 2007; Sijen et al. 2007). Generation of dsRNA

precursor is dependent on the activity of RdRP (Vazquez 2006). piRNAs are 24-31 nt small RNAs that associate with Piwi proteins, which is a subfamily of Argonaute proteins. piRNAs are mainly expressed in germline and implicate in germline development, silencing of selfish DNA elements (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Watanabe et al. 2006; Ambros and Chen 2007). Mouse have four Argonaute family genes (*Ago1~4*) and three Piwi family genes (*Miwi*, *Mili*, and *Miwi2*). AGO1~4, which are ubiquitously expressed in many tissues, recruit miRNAs (Liu et al. 2004). In contrast, Piwi family genes, which are expressed in germcells, recruit piRNAs.

Piwi family genes are expressed predominantly in male germline cells and have crucial roles in spermatogenesis. MIWI is reported to be expressed from spermatocytes to spermatids and is associated with 28-31 nt piRNAs (Kuramochi-Miyagawa et al. 2001; Girard et al. 2006). Disruption of *Miwi* causes spermatogenic arrest at the beginning of the round spermatid stage (Deng and Lin 2002). Expression of *Mili* has reported to be from gonocyte of fetal testes to spermatocyte. *Miwi2* is predominantly expressed in gonocyte of fetal testes (Kuramochi-Miyagawa et al. 2001; Kuramochi-Miyagawa et al. 2008). MILI and MIWI2 are associated with 24-27 and 25-28 nt piRNAs respectively

(Kuramochi-Miyagawa et al. 2008). Spermatogenesis in *Mili*-null and *Miwi2*-null mice is blocked completely at the early prophase of the first meiosis (Kuramochi-Miyagawa et al. 2004; Carmell et al. 2007).

By studies of several groups, piRNAs expressed in mouse spermatocytes and spermatids (pachytene piRNAs) have been identified (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Watanabe et al. 2006). Pachytene piRNAs are derived from ~100 of genomic loci ranging from 1 kbp to 100 kbp. From each locus, thousands kinds of piRNAs are generated. Almost all piRNAs in a given cluster are from the same strand. This strand bias implies that piRNAs are processed from a long primary transcript. However, because most of pachytene piRNAs are derived from unique genomic locus, the targets of pachytene piRNAs remain elusive. piRNAs expressed in neonatal testes (prepachytene piRNAs) have also been reported (Aravin et al. 2007a). In contrast to pachytene piRNAs, most of prepachytene piRNAs are derived from retrotransposons including SINE (short interspersed nuclear elements), LINE (long interspersed muclear elements) and LTR (long terminal repeat) retrotransposon, indicating the involvement of prepachytene piRNAs in retrotransposon suppression.

It was reported that the *Mili* or *Miwi2* targeted mice exhibited enhanced IAP (intracisternal A particle) and LINE-1 expression, and DNA methylation of the 5'-untranslated region of LINE-1 was shown to be reduced in these mutants (Aravin et

al. 2007b; Carmell et al. 2007). These data raise the possibility that piRNAs may be involved in the suppression of retrotransposons through DNA methylation. During spermatogenesis, the DNA methylation status of the regulatory region in retrotransposons changes dynamically. These regions are demethylated in PGCs (Primordial Germ Cells) around E12.5-13.5, and the reacquisition of DNA methylation (de novo DNA methylation) takes place in the gonocytes of fetal testes around E16.5-18.5 (Lane et al. 2003; Kato et al. 2007). To know the link between piRNAs and DNA methylation, it is important to identify the piRNAs expressed in the period of *de novo* methylation. However, there are no reports of piRNAs expressed in this period.

In the first chapter of my thesis, the results obtained from analysis of fetal testis small RNAs are described. To analyze piRNAs expressed in the period of de novo methylation, I cloned the small RNAs from fetal testes. piRNAs expressed in this stage were mostly derived from retrotransposons. To know the effect of piRNAs, *Zucchini*, which is reported to be involved in the accumulation of piRNAs in *Drosophila* (Pane et al. 2007), KO (knock out) mice was produced. In *Zucchini* mutant mice, the level of piRNAs were dramatically decreased. The DNA methylation levels of the region where piRNAs are mapped were also decreased in KO mice. The results of my study support the hypothesis that piRNA pathway is one of the mechanisms that define the genomic DNA sequences that are subject to DNA methylation in fetal testes

and help to understand the mechanism of piRNA-mediated DNA methylation.

Whether endogenous siRNAs are present in mouse has been unclear for two reasons: (I) there is no evidence for the presence of RdRP activity in mammals, which generates dsRNAs, namely the precursors of siRNAs, and (II) induction of the interferon pathway by dsRNAs usually results in cell death, suggesting that mammalian cells may not tolerate dsRNAs (Elbashir et al. 2001). However, in mouse oocytes and preimplantation embryos, the interferon response is suppressed and injection of long dsRNAs results in specific reduction in amount of the target mRNAs (Svoboda et al. 2000; Wianny and Zernicka-Goetz 2000; Stein et al. 2003; Yan et al. 2005). These studies suggest that dsRNA-induced RNAi pathway is active in mouse oocytes and early embryos. From these results, I hypothesized that endogenous siRNAs may be present in mouse oocytes and early embryos, and identified retrotransposn-derived siRNA-like molecules from mouse fully grown oocytes (Watanabe et al. 2006). However, the number of siRNA-like molecules identified so far is small, and their precise identity, biogenesis pathway and roles are unknown.

In the second chapter of my thesis, the results obtained from analysis of oocyte small RNAs are described. In order to obtain a comprehensive picture of siRNA molecules, we have sequenced more than 100,000 of small RNAs from mouse growing

oocytes. The results presented here establish the existence of endogenous siRNAs in mammals, and reveal the function and biogenesis pathway of endogenous siRNAs and a role of pseudogenes in gene regulation.

Material and Methods

Small RNA library construction and sequencing

Five microgram of total RNA from ~12,000 growing oocytes (35-60 μm) from B6D2F1 females, 50μg of total RNA from 12.5-19.5 dpc (days postcoitum) testis gonocyte and MILI-IP small RNAs from C57BL/6 P8 (postnatal day 8) ovaries (see above) were used. Small RNAs ranging 15-40 nt in length were cloned using a Small RNA Cloning Kit (TAKARA). For sequencing of the oocyte small RNA library and the fetal testis library, a 454 Life Sciences sequencer was used. MILI-IP small RNAs were sequenced using capillary sequencers. After trimming of the adaptor sequences, inserts were mapped to the mouse genome (mm8 assembly, Feb 2006) using blastn. Sequencing of the growing oocyte small RNA library yielded 176,267 reads of 17-40 nt small RNAs, 103,995 (59 %) of which comprising of 63,244 of non-redundant sequences were mapped to the genome with perfect match. Sequencing of the fetal testis small RNA library yielded 261,325 reads of 17-40 nt small RNAs, 127,997 (49 %) of which comprising of 64,632 of non-redundant sequences were mapped to the genome with perfect match. Capillary sequencing of the MILI-IP small RNA library produced 4937 reads of 17-40 nt small RNAs, 4129 of which were mapped to the genome with perfect match. We analyzed only these perfect match sequences.

Immunoprecipitation of MILI-piRNA complex and AGO2-siRNA complex

A whole cell extract was prepared from P8 ovaries of C57BL/6 females in a lysis buffer (20 mM HEPES pH 7.3, 150 mM NaCl, 2.5 mM $MgCl_2$, 0.1% NP40, 1x Roche-Complete). Cleared extract was incubated with anti-MILI antibody (Aravin et al. 2006) for 12 hr at 4^oC. Using protein G Sepharose, MILI-piRNA-antibody or AGO2-small RNA-antibody complexes were collected, and then washed four times in a wash buffer (20 mM HEPES pH 7.3, 320 mM NaCl, 2.5 mM $MgCl_2$, 0.1% NP40, 1x Roche-Complete) for 15 minutes.

Annotation of small RNAs

To know the exact positions of the small RNAs in the genome, the small RNA sequences were aligned with the genome using blastn (ftp://ftp.ncbi.nlm.nih.gov/genbank/) with a perfect match criterion. To identify small RNAs corresponding to various repeats (rRNA, tRNA, retrotransposon, DNA transposon etc.), the genomic positions of repeats were retrieved from the University of California, Santa Cruz (UCSC) web site (http://hgdownload.cse.ucsc.edu/downloads.html) and compared with the genomic positions of small RNAs. If the genomic position of a certain small RNA overlapped with any repeats by 15 nt, this small RNA was considered to be repeat-derived. Repeat

names were retrieved from all positions where the small RNA was mapped, and if multiple repeat names were retrieved, the class (such as LTR/MaLR or rRNA) and subclass (such as IAP), where applicable, were determined according to the majority of positions. If the top two repeats had the same number of positions, the class or subclass were not determined. To identify small RNAs corresponding to tRNAs, rRNAs, snRNAs, snoRNAs, scRNAs, miRNAs, piRNAs (known ones from adult testis) and mRNAs based on sequence similarity, the sequences of these RNAs were extracted from the flat files and downloaded the sequences from the following databases: tRNAs, Genomic tRNA Database (http://lowelab.ucsc.edu/GtRNAdb/Mmusc/); rRNAs, European ribosomal RNA database (http://www.psb.ugent.be/rRNA/index.html); snoRNAs, snoRNA database (http://www-snorna.biotoul.fr) and RNA database (http://jsm-research.imb.uq.edu.au/rnadb/); piRNAs, RNA database (http://jsm-research.imb.uq.edu.au/rnadb/); miRNAs, miRBase (http://microrna.sanger.ac.uk/sequences/index.shtml); mRNAs, Refseq Genes (ftp://ftp.ncbi.nih.gov/refseq/) and Ensemble Genes (http://www.ensemble.org/index.html). Then blastn search (ftp://ftp.ncbi.nih.gov/blast/) was performed using our small RNA sequences as queries and the sequences downloaded as above as a database. Since the sequences downloaded in the database could not cover all RNA species present in cells, the small RNA sequences were

aligned with the RNA sequences in the database using a 90% match criterion including indels. Namely, a single base mismatch for 17-19 nt small RNAs was allowed, 2 mismatches for 20-29 nt small RNAs, 3 mismatches for 30-39 nt small RNAs and 4 base mismatches for 40 nt small RNAs. Firstly, blastn program with an option not masking repeat sequences (-F, F) was used and extracted the best hit regions on the RNAs deposited in the database with extra 5 bases at both ends. Since blastn is a local alignment program, the small RNAs were re-aligned with the extracted RNA fragments using global alignment program. Finally, the repeat annotations based on genomic position and the annotations based on sequence similarity were combined. If a small RNA had more than one annotation, following order of priority was used: rRNA, tRNA, snoRNA, sc/srpRNA, miRNA, rasiRNA, piRNA and mRNA. The non-annotated sequences were classified as unknown.

Cluster identification

After annotation of small RNAs, small RNA clusters were identified. Only small RNAs that hit the genome 1-10 times and were annotated as repeat, piRNAs, mRNAs or unknown (the latter two can also include siRNAs and piRNAs) were used. Firstly, the genome was scanned using a 10 kb window and extracted the windows that had more than five small RNAs. Any overlapping windows that fulfilled these criteria

were combined. If the combined region had more than 3 unique hit small RNAs, the region was considered to be a cluster. The positions of the most 5' and 3' small RNAs were considered to be the boundaries of the cluster. In cases where the boundary of the next cluster was located within 100 kb, the two clusters were considered to be one cluster (because both clusters may have been derived from a single precursor).

If one or more MILI-IP small RNAs were mapped within or around 1 kb of the cluster described above, this cluster was classified as a piRNA cluster. For this analysis MILI-IP small RNAs that uniquely hit the genome were used.

Identification of siRNA clusters

For identification of candidates for siRNA clusters, only small RNAs that (1) constituted clusters, (2) hit the genome 1-10 times and (3) were annotated as repeat, mRNA, piRNA or unknown were considered. A total of 444 small RNA clusters were subjected to the selection. The series of criteria described below were used to select siRNA clusters of the respective classes.

Selection procedure for hp-siRNA cluster

1. Select clusters in which more than three small RNA sequences hit two times within a certain 5 kbp region. These two hits should lie in opposite orientations. (4/444)

2. In the clusters, the small RNAs of all pairs should be arranged in a symmetrical

manner to form an inverted repeat. (4/4) The boundaries of the cluster are re-determined by the positions of the outermost pairs.

3. More than three unique hit small RNA sequences should be mapped within the cluster. (4/4)

4. More than 90% of the unique small RNA sequences should lie in the same orientation. (4/4)

5. More than 80% of small RNA sequences that constitute the cluster should be 19-23 nt in length. (4/4)

If only step 5 is applied to the 444 small RNA clusters, 106 meet its criterion. Thus, it is highly significant that all 4 clusters selected by steps 1-4 meet the criterion of step 5 ($P = 0.034$; χ^2 test). The low p-value at step 5 and absence of clusters that dropped off at steps 2-4 suggest that inverted-repeat structures are found only among the 19-23 nt small RNA clusters.

Selection procedure for trans-nat-siRNA cluster

1. Select pairs of clusters, in which the partner clusters share more than two small RNAs unique to the pair. (43/444; 23 pairs) (Three of the 43 clusters paired with two clusters.)

2. In each cluster of a pair, more than 90 % of unique small RNAs should be derived from the same strand. (17/43; 9 pairs)

3. In each pair, the shared unique small RNAs identified in step 1 should lie in the same orientation as the other unique small RNAs in one cluster but should lie in the opposite orientation in the other cluster. (16/17; 8 pairs)

4. More than 80% of small RNA sequences that constitute the cluster should be 19-23 nt in length. (14/16; 7 pairs) ($P = 1.2 \times 10^{-8}$; χ^2 test)

The low p-value at step 4 and the fact that only one pair dropped off at step 3 suggest that the sense/antisense relationship is highly unique to the 19-23 nt small RNA clusters.

Selection procedure for cis-nat-siRNA cluster

1. Select clusters in which some unique small RNAs hit the sense strand while other unique small RNAs hit the antisense strand. [Number of sense (or antisense) unique small RNA sequences] / [Number of total unique small RNA sequences] should be between 0.25 and 0.75. (70/444)

2. The cluster should contain more than 10 small RNA sequences. (35/70)

3. More than 80% of small RNA sequences that constitute the cluster should be 19-23 nt in length. (17/35) ($P = 0.0013$; χ^2 test)

Some piRNA clusters contain both sense and antisense piRNAs, and therefore step 1 would select both piRNA and siRNA clusters. Step 3 was therefore designed to exclude piRNA clusters; however, some piRNA clusters were not filtered out due to the

presence of only a small number of small RNA sequences. Therefore step 2 was added to decrease the false positive rate. The low p-value at step 3 suggests that 19-23 nt small RNA clusters were enriched by steps 1 and 2.

Of the top 30 clusters in terms of the number of mapped small RNA species, 25 clusters were either piRNA clusters or one of three classes of siRNA cluster (hp-siRNA, trans-nat-siRNA or cis-nat-siRNA) (Table 2).

Distribution and frequency of piRNAs

The sequences of IAP1 (M17551) and L1_MdA (nucleotides 588– 7713 of M13002) were retrieved from the flat files of GenBank. All of the small RNA sequences cloned were BLAST-searched against the IAP1 and L1_MdA sequences, using a cutoff E-value of 0.0001. For *Rasgrf1* locus, small RNAs reported by Aravin et al. (Aravin et al. 2008) were BLAST-searched against *Rasgrf1* sequence. The number of hits was determined every 100 nt.

RT-PCR analyses

For expression analysis of small RNA pathway components, 100-1000 oocytes were collected from C57BL/6 females at various developmental stages and their total RNA (~100 ng) was reverse transcribed using random primers. For

expression analysis of *Zucchini*, about 1 μg of total RNA from various tissues or total testis RNA from various developmental stages were used. For expression analysis of retrotransposons in *Mili* KO oocytes, about 100 ng of total RNAs from 40-60 μm oocytes (500 oocytes) was used. For expression analysis of retrotransposons and protein-coding transcripts in *Zp-3* conditional *Dicer* KO oocytes (Tang et al. 2007) or *Zp-3* conditional *Ago2* KO oocytes, about 15 ng of total RNAs from 60-80 μm oocytes (25 oocytes) was used. Before isolation of total RNA from *Dicer* KO or *Ago2* KO oocytes, 10 pg of EGFP mRNA per oocyte was added.

Generation of *Zucchini* KO mouse

Targeting vector was constructed using lox Neo targeting vector (Saga laboratory). ~2.5 kbp short fragment derived from intron between 1st and 2nd exons was inserted into Xho1 site. ~7 kbp long fragment from upstream region of exon1 was inserted into EcoR1 site. I obtained 4 lines of R1 ES cell that were targeted. Of these, 3 lines were germline transmitted. Results presented herein were obtained from mice with a mixed 129/B6 background. Animals were backcrossed to B6 2–3 generations. Then, the mice were crossed to CAG-Cre mouse (B6 background) and Neomycin cassettes were removed. Mouse genotyping was performed by Southern blotting or PCR.

Antibody and Western Blotting

Rabbits were immunized with *Zucchini* peptide (CRKAGIQVRHDQDLGY 111-125), and specific antibody was affinity purified from the antisera using the same antigen coupled to NHS-activated HP. For western blotting of Zucchini, lysates of 16 day testis and NIH3T3 transfected with *Zucchini* expression vector were subjected to 15% acrylamide SDS-PAGE and transferred to nitrocellulose membrane.

Small RNA library for PCR

MILI-IP small RNAs from C57BL/6 P8 ovaries, AGO2-IP small RNAs from C57BL/6 P8 ovaries, total small RNAs from C57BL/6 P8 ovaries, total small RNAs from conditional *Dicer* KO P15 ovaries, total small RNAs from control P15 ovaries without Cre recombinase, *Mili* -/- P8-11 ovaries and *Mili* +/- P8-11 ovaries were used for the construction of small RNA libraries for PCR. A synthetic RNA of plant MIR-164 was added to the total RNAs from *Dicer* KO and control ovaries before gel fractionation. Small RNA libraries were constructed as described above and amplified. In each quantitative PCR (q-PCR) reaction, 0.5 ng of the amplified library was used. Individual small RNAs were amplified with specific primers complementary to the 3' part of the respective small RNAs and a universal primer corresponding to the 5' linker. Amplified products were sequenced and the sequences of the 5' part of the small RNA

were confirmed. Only confirmed small RNAs were analyzed. For quantification, each experiment was repeated three times.

5'RACE analyses

To determine the 5' end of cis-nat-siRNA precursors, 5'RACE was performed by using a GeneRacer kit (Invitrogen) according to the manufacturer's instructions. About 200 ng of total RNA from 40-60 μm oocytes (~1,000 oocytes) from C57BL/6 females was used.

Northern and Southern Blotting

Northern blotting and Southern blotting were done using formamide buffer (5xSSPE, 50 % formamide, 5xDenhalt's, 0.5 % SDS, 50 μg/ml sperm DNA). Probes were synthesized by random priming method. Hybridization was performed at 42°C. The membranes were washed three times with a 0.2× SSC/0.1%SDS solution at 68°C for 15 minutes. Small RNA Northern blots were performed as described previously (Watanabe et al. 2007), with 20 μg of total RNA loaded per well. Hybridization was performed at 42°C in the buffer (0.2 M NaHPO4 (pH 7.2), 1 mM EDTA, 1% BSA, and 7% SDS).

H.E. staining and *In situ* Hybridization

For H. E. staining, testes are fixed in Bouin's solution at 4 ^{o}C overnight. Testes were paraffin embedded and sectioned at a thickness of 6 μm. *In situ* hybridization analysis of L1Md was performed according to Soper et al (Soper et al. 2008).

Immunostainig

Testicular cells were dispersed using 0.25 % trypsin. Cells were fixed in 2 % paraformaldehyde for 10 min. on ice. Cells were stuck on slide glass using cytospin. Cells were permerlized using the Buffer (0.5% Triton in PBS) for 30 min. Blocking was performed using the Buffer (0.5% Triton, 1.5 % goat serum in PBS) for 30 min.

Testes were fixed in 2 % paraformaldehyde for 3 hr at 4^{o}C. After wash in PBS three times for ten minutes, testes were rotate in 15 % sucrose PBS for 3 hr at 4^{o}C, then move into 30 % sucrose PBS overnight at 4^{o}C. Testes were put into O.C.T. compound for 1 hr at 4^{o}C, and then put into fresh O.C.T. compound and frozen using liquid nitrogen. Testes were sectioned at a thickness of 10 μm and stuck on slide glass, dried for 30 min. using dryer and washed 2 times in PBS to remove O.C.T. compound. Blocking was performed using the Buffer (1.5 % goat serum in PBS). Primary antibody and secondary antibody were diluted in the Blocking buffer.

Germ cell isolation

Zucchini KO mice were crossed with Oct-4/GFP transgenic mice to obtain GFP-positive spermatogonia. Testis cells from P5-7 were collected by trypsin digestion, and the GFP-positive cells were sorted by FACS. Alternatively, testis cells from P5-7 were collected by trypsin digestion, stained with α-EpCAM and the positive cells were sorted by FACS.Genomic DNA was extracted from the sorted germ cells.

Bisulfite sequencing analysis

DNA methylation analysis of *Rasgrf1* locus was done as described in the thesis of Tomizawa. DNA methylation analyses of L1Md and IAP1 were done according to Miyagawa et al (Kuramochi-Miyagawa et al. 2008).

Oligo Sequences

Primer name	**Sequence**
Primers used for expression analysis of small RNA pathway components	
Dicer F	TGGCACCAGCAAGAGACTCA
Dicer R	CTGGGAGATGCGATTTTGGA
Ago1 F	GGCATCTCAAGAATACCTACTCAG
Ago1 R	CTACCACTGCTGTGATAGATGGT
Ago2 F	TCAAGCTGGAGAAGGACTATCAGC
Ago2 R	TGATCTTCGTGTCCACGGTTGTG
Ago3 F	TCCCTGCACCAGCATATTACG
Ago3 R	TGTGGATCTCGCCCATTGC
Ago4 F	AGATAAAATGGAAAGGGTGGGG
Ago4 R	AATGTGAAGGACGGCTGGTTC
Actin F	CCACCACAGCTGAGAGGGAA
Actin R	AGCCACCGATCCACACAGAG
Primers used for quantitative PCR of individual small RNAs	
Universal primer F (5’ linker sequence)	AAAGATCCTGCAGGTGCGTCA
Plant MIR164	TGCACGTGCCCTGCTTCT
let7f F	ACTATACAATCTACTACCTCA

miR-16 F	GCCAATATTTACGTGCTGCTA
miR-103 F	CATAGCCCTGTACAATGCTGCT
miR-183 F	GTGAATTCTACCAGTGCCATA
25-26 nt small RNA1 F	GACTCTAGATACCGGGGTTCA
25-26 nt small RNA2 F	TGTCCTGCTACTCCGTGCCTA
25-26 nt small RNA3 F	AAGCCAGTCTAATAGCCACAA
25-26 nt small RNA4 F	TACCAATCCCAGCAATGCC
hp-siRNA1 F (21 nt small RNA1 F)	GCTCTAAGGGCACCGTT
hp-siRNA2 F (21 nt small RNA2 F)	ACAGCATCTGCGAAGGC
cis-nat-siRNA1 F (21 nt small RNA3 F)	ACCGCCTAAAGGTTTGTCA

Primers used for 5'RACE analysis of cis-nat-siRNA clusters

Kif4 exon3 R	CTGCCCATAGGCCAGGACAGTTGCA
Kif4 exon2 R (for 2nd PCR)	CCTGTTCAGTAGAGGGGTCAAACACA
Pdzd11 exon3 R	CTGCCCATAGGCCAGGACAGTTGCA
Pdzd11 exon2 R (for 2nd PCR)	GTCATCATAGGGAATCCGGTTGTCCA

Primers used for quantitative PCR of mRNAs and oocyte retrotransposons

Rangap1 F	ATGGCTGAGACTCTGAAGACTCTG
Rangap1 R	ACGGACAGCATCTGCGATGGCAAC
Ppp4r F	CCTGCTCCAGTGGACAGCTCCT

Ppp4r R	GCAGACTTGCTGTTGGTGTCAG
Kif4 F	GTAAGAGTGGCACTGCGTTGTCG
Kif4 R	TAAAGACCTCTTCCTGTTCAGTAGAG
Pdzd F	AAGCTGGCTAGACCTCTGAGA
Pdzd R	GAACTCCTAGTACAGTACAGT
Beta-actin F	ACAGCTTCTTTGCAGCTCCT
Beta-actin R	ATTCCCACCATCACACCCTG
GAPD F	ATGACATCAAGAAGGTGGTG
GAPD R	CATACCAGGAAATGAGCTTG
EGFP F	GAACCGCATCGAGCTGAAGG
EGFP R	CGGATCTTGAAGTTCACCTTGATGC
IAP1 F	ACAAGAAAAGAAGCCCGTGA
IAP1 R	GCCAGAACATGTGTCAATGG
L1Md F	GAGACATAACAACAGATCCTGA
L1Md R	GAACTTTGGTACCTGGTATCTG
RLTR10 F	GTGAAGGTAGAGGTCTGATC
RLTR10 R	GAAGGTATGTCTGATTGCATG
MTA F	ATGTCTTGGGGAGGACTGTG
MTA R	AGCCCCAGCTAACCAGAACT

Targeting vector construction

Long F	TTGCTCACTGTGAGATGAATTCAGAG
Long R	TCACAGAATTCACTTTAGGGCGAAGCAGTACTCTG
ShortF	GACAAAGTCGACATGCAACTCATGTCACAGATGTG
ShortR	CCCTACGTCGACCGACTTTGTCCAGAGCAGGTAG

Northern Blot and RT-PCR analysis of Zucchini mRNA

NorthernF	CGGCCGAATTCATGGGGCGCTCGAGTTG
NorthernR	GCAAAGAATTCCTTGTGGTGCATGTAGCCTAG
RT-PCRF	ATGCACCACAAGTTTGCCATCGTTG
RT-PCRR	CAGTACTGTCCAACACTGTCA

Southern Blot and genotyping PCR of Zucchini gene

SouthernF	CAACACCAAGATTAGAAGGTCAGA
SouthernR	GATTGCTTCTAGCTGAGCAGTAG
GenotypingF1	TAGGTCCACAGTGGTCCACACTCAG
GenotypingF2	CTCCATGCCTTCACTTATGTCAG
GenotypingR1	GAAAGTGATGTTCTGTCTGTATC

Primers used for quantitative PCR of testes retrotransposons

L1_5UTRF	GGCGAAAGGCAAACGTAAGA
L1_5UTRR	GGAGTGCTGCGTTCTGATGA

L1_ORF2_1F	GGAGGGACATTTCATTCTCATCA
L1_ORF2_1R	GCTGCTCTTGTATTTGGAGCATAGA
L1_ORF2_2F	GAGACATAACAACAGATCCTGA
L1_ORF2_2R	GAACTTTGGTACCTGGTATCTG
IAP_gagF	AACCAATGCTAATTTCACCTTGGT
IAP_gagR	GCCAATCAGCAGGCGTTAGT
IAP_3LTRF	GCACATGCGCAGATTATTTGTT
IAP_3LTRR	CCACATTCGCCGTTACAAGAT

Primers used for in situ hybridization analysis of L1Md

L1_ISHF	AGTTCCCAACATAGAGTCCTGAGTT
L1_ISHR	AGTGGGCAGAGTATTCTCTGCAGG

Results

1. Small RNAs in fetal testes

Characterization of small RNAs in fetal testes

Based on the role of small RNAs from other organisms in gene silencing (Zilberman et al. 2003), it is possible that DNA methylation is mediated by small RNAs. Therefore, we analyzed many small RNA sequences (127,997 clones), 17–40 nt in length, from E12.5–E19.5 fetal male germ cells, to obtain a comprehensive picture of the piRNAs present at this stage, and compared findings with the results of previous studies on piRNAs in neonatal (prepachytene) testes (Aravin et al. 2007b) and adult testes (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Watanabe et al. 2006). The lengths of the small RNAs from the germ cells showed a bimodal pattern (Fig. 1A). One peak was observed at 21 nt, which corresponds to the length of miRNAs, and a second was observed at 25–27 nt, which is the length of piRNAs. Annotation of the small RNAs revealed that the library was enriched in retrotransposon sequences, with the exception of the breakdown products of abundant noncoding RNAs (rRNA/tRNA/snRNA/snoRNA/ scRNA/srpRNA) (Fig. 1B, Table 1).

The repeat-associated small RNAs (rasiRNAs) in the library, which reportedly bind MILI, showed a single peak at 25– 27 nt, suggesting that rasiRNAs in fetal testes are piRNAs (Fig. 1A). Detailed characterization of the repeat-derived piRNAs revealed

that there were some differences between the fetal germ cell piRNAs and the reported prepachytene piRNAs (Aravin et al. 2007b). As shown in Figure 1C and Table 2, the majority of the repeat-derived piRNAs in the E16.5 male germ cells were LTR retrotransposons (55%); namely, ERVK (37%), ERV1 (10%), MaLR (6%), and ERVL (2%). The others were LINEs (30%) and SINEs (11%). In contrast, among the prepachytene piRNAs, the SINE frequency (49%) was higher than the LTR (33.8%) or LINE (15.8%) frequency (Aravin et al. 2007b). Unique pachytene piRNAs were scarcely detected in the fetal male germcells (0.1%). Next, as the small RNAs that correspond to L1Md (more than 20 % of total repeat-derived piRNAs) and IAP1 (more than 10 %) were abundant, we examined the nucleotide composition of the piRNAs corresponding to L1Md_Gf and IAP1 retrotransposons. The first and tenth nucleotides of the piRNAs are shown in Figure 1, D and E. Most of the piRNAs started with uridine which is characteristics of piRNAs.

Piwi-mediated cleavage has been reported to promote the formation of secondary piRNAs (Brennecke et al. 2007; Gunawardane et al. 2007). This allows active transposons to participate in a feed-forward loop that amplifies silencing. Because Piwi proteins cleave targets opposite nucleotides between 10 and 11 of the piRNAs, secondary piRNAs produced by Piwi-mediated cleavage are enriched for adenine (A) at position 10. The tenth nucleotide in fetal repeat-derived piRNAs was

enriched in adenine, suggesting the involvement of feed-forward loop in the production of repeat-derived piRNAs.

Adult and neonatal piRNAs are clustered within the genome (Aravin et al. 2006, 2007; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Watanabe et al. 2006). Therefore, a cluster analysis of the male fetal gonadal small RNAs was performed and detected 205 clusters (Table 3), only seven and 25 of which were identified in adult and prepachytene testes, respectively (Fig. 1F). About 75% of the clustered small RNAs were 24–28 nt in length, and the percentage of small RNAs with U as the first nucleotide was high (71.5%). Therefore, I conclude that most of the clusters are piRNA clusters, and that the set of piRNAs expressed at the stage of de novo methylation is quite different from the piRNA sets expressed in the neonate and adult.

The distribution and frequency plots of the piRNAs that correspond to the type Gf Line-1 and IAPI1 genes are shown in Figure 1, G and H, respectively. For Line-1, the number of piRNAs that corresponded to the regulatory region was higher than that corresponding to the coding region, raising the possibility that Line-1 piRNAs mediate silencing through transcriptional regulation.

Reduced piRNA accumulation in *Zucchini* KO

To understand the roles and pathway of piRNAs in fetal testes, *Zucchini* KO mouse was generated. *Zucchini* gene is a member of phospholipase D superfamily, which includes nuclease and phospholipase (Zhao et al. 1997), and first identified in *Drosophila* (Pane et al. 2007). It has been reported that some transposable elements are upregulated and cognate piRNAs are diminished in *Drosophila* mutant. Mouse *Zucchini* mRNA was specifically expressed in testes and oocytes, where piRNAs are expressed (Fig. 2A, B). Mouse *Zucchini* gene consists of two exons, and first exon was removed for KO (Fig. 3A). I confirmed the depletion of *Zucchini* protein in KO testes using Western Blotting (Fig. 3B).

Mouse piRNA pathway component mutants (*Mili*, *Miwi2*, *Maelstrom*) have been reported to show common phenotypes (Kuramochi-Miyagawa et al. 2004; Aravin et al. 2007b; Carmell et al. 2007; Kuramochi-Miyagawa et al. 2008; Soper et al. 2008). They show impaired spermatogenesis at zygotene stage with accumulation of L1 retrotransposon. *Zucchini* KO mouse showed the same phenotypes. Testis from adult KO mouse exhibited smaller size than that from wild type (Fig. 3C). In KO testes from P16, when germ cells of first wave reach pachytene stage, abnormal cells were observed (Fig. 4A). Meiotic arrest at zygotene stage was confirmed by using SCP3 immunofluorescene staining (Fig. 4B). Accumulation of L1 RNA was also detected by q-PCR and in situ hybridization (Fig. 5A, B).

As *Zucchini* KO Mouse exhibited the same phenotypes as piRNA pathway mutants, I carried out Northern blotting analysis using probes that detect fetal piRNAs. The levels of piRNAs were decreased in KO testes (Fig. 6A). The components of piRNA pathway were known to be localized to nuage, which is ribonucleoprotein amorphous aggregates lacking limiting membranes observed in germ cells (Chuma et al. 2006; Aravin et al. 2008). To investigate the cause of impairment of piRNA pathway in *Zucchini* KO, I analyzed the localization of piRNA pathway components. I observed mislocalization of MILI, MVH and TDRD1, which are components of piRNA pathways and are localized to nuage, around the centrosome in *Zucchini* KO 16.5 dpc testes (Fig. 6B,C, data not shown).

DNA methylation and piRNAs

Because L1_Md and IAP1 piRNAs were major classes in fetal testes piRNAs, I measured the DNA methylation levels of L1_Md and IAP1 in *Zucchini* KO mice using Bisulfite sequencing. As shown in Fig. 7, the methylation levels of L1Md_A, L1Md_Gf and IAPIΔ1 were decreased dramatically in KO spermatogonia. The decrease of DNA methylation in retrotransposon has been also reported in *Mili*, *Miwi2* KO (Aravin et al. 2007b; Carmell et al. 2007). These data imply that piRNAs repress retrotransposn through DNA methylation.

Differentialy methylated region (DMR) of imprint genes are demethylated in PGCs around E12.5-13.5, and the reacquisition of DNA methylation (de novo DNA methylation) takes place in the fetal testes around E16.5-18.5 (Kato et al. 2007). Decreased methylation in *Rasgrf1* locus has been found in *Mili* KO (Miyagawa S. et al., personal communication). I also examined the methylation levels of DNA loci in DMRs of imprinted genes (*H19*, *IGDMR*, *Rasgrf1*) using *Zucchini* KO spermatogonia. Of three loci, *Rasgrf1* locus showed decreased methylation in KO spermatogonia. Tomizawa in our laboratory reports in his thesis that the DNA methylation was observed at two separate regions in *Rasgrf1* locus (primary methylated region and another methylated region). In *Zucchini* KO, decreased methylation was observed only in a specific locus of primary methylated region, suggesting that piRNA pathway is not a sole determinant of *Rasgrf1* methylation (Fig. 8). To investigate the link between piRNAs and DNA methylation in *Rasgrf1* locus, piRNAs from 16.5 dpc testes reported by Aravin et al. (Aravin et al. 2008) were mapped to the primary methylated region of *Rasgrf1* locus using criteria that allow some gaps or mismatches. piRNAs were densely mapped in specific region annotated as RMER4B by Repeatmasker. Remarkably, decreased methylation was observed around this mapped region.

To further investigate the connection between piRNAs and *Rasgrf1* locus, piRNAs mapped to the *Rasgrf1* RMER4B were analyzed closely. piRNAs were

predominantly mapped to the two loci in *Rasgrf1* RMER4B. The sequences of the two loci resembled closely each other and were orientated in the same directions, and each locus shared common piRNAs mapped (Fig. 9). In each locus, most of piRNAs (total of 161 clones comprising 17 kinds) were mapped in sense directions, and a few (total of 6 clones comprising 3 kinds) piRNAs were mapped in antisense direction. None of sense piRNAs were uniquely mapped to these loci; rather, most of them were aligned with mismatches (Fig. 9). Almost all of sense piRNAs were mapped to chr7 piRNA cluster, which is the largest cluster in fetal piRNA cluster (Table 3), and some were uniquely mapped to this chr7 cluster (data not shown). This result suggests sense piRNAs are produced from the chr7 piRNA cluster. On the other hand, all of antisense piRNAs were uniquely mapped to either of two loci (Fig. 9), suggesting that antisense piRNAs are produced from *Rasgrf1* locus. Overlap of sense piRNAs (99% start with U) by 10 nt with antisense piRNAs (0% start with U) was observed, which is characteristics of a set of primary piRNA and secondary piRNAs (Brennecke et al. 2007; Gunawardane et al. 2007). These results suggest that antisense piRNAs are secondary piRNAs produced from transcripts by the action of sense piRNA-mediated cleavage. Together, these results suggest that the transcripts transcribed from *Rasgrf1* locus are targeted by Piwi-piRNAs in 16.5 dpc testes.

2. Small RNAs in oocytes

Characterization of small RNAs in growing oocytes

In order to obtain a comprehensive picture of endogenous siRNAs, more than 100,000 of small RNAs from mouse growing oocytes have been sequenced. The length distribution of the total small RNAs showed a bimodal pattern (Fig. 10A): one peak was observed at 21 nt, corresponding to the length of miRNAs and siRNAs, and the other at 25-26 nt, corresponding to the length of piRNAs, which are a distinct class of small RNAs bound to Piwi family proteins (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Watanabe et al. 2006). Annotation of the small RNAs revealed that both the 21 nt and 25-26 nt small RNAs were mainly derived from repeat sequences, most of which were retrotransposons (Fig. 11A, Table 4) and were quite diverse in sequence (21969 clones of 21 nt small RNAs comprising 10194 different sequences and 18572 clones of 25-26 nt small RNAs comprising 12006 different sequences).

In order to examine expression of some components of the small RNA pathways in growing oocytes, RT-PCR and Western blotting were carried out. *Ago2*, *Ago3* and *Dicer*, which are the components of the siRNA and miRNA pathways, were expressed at high levels throughout oocyte growth (Fig. 10B). Of the three mouse Piwi family proteins, MILI was predominantly expressed at early stages of oocyte growth

(Fig. 10C). However, MIWI or MIWI2 were not detected in growing oocytes. To examine whether the 25-26 nt small RNAs in growing oocytes are piRNAs bound to MILI, a total small RNA library and a MILI-immunoprecipitated (IP) small RNA library were constructed from ovaries, and the abundance of several 25-26 nt small RNAs were examined in each library. All 25-26 nt small RNAs were enriched in the MILI-IP library relative to the control miRNAs and 21 nt small RNAs (Fig. 10D), indicating that the 25-26 nt small RNAs in oocytes are mostly MILI-bound piRNAs.

Biogenesis and role of endogenous siRNAs

Genomic mapping of the oocyte small RNAs (excluding those with more than 10 hits to the genome, which are likely to be repeat sequences) revealed the presence of a total of 444 small RNA clusters. piRNAs are known to be mapped to the genome in clusters (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Watanabe et al. 2006), and we identified 152 piRNA clusters by sequencing and mapping of MILI-IP small RNAs (Fig. 11B). Surprisingly, many of the largest clusters, which were determined by the number of small RNAs, were not included in these piRNA clusters (Fig. 12A). Furthermore, the length distribution of the small RNAs constituting these non-piRNA clusters was centered at 21 nt (Fig. 12B), which is the length of miRNAs and siRNAs. However, the lack of a short stem-loop structure (Fig.

13), which is a characteristics of miRNA precursors, in most of the genomic sequences encompassing the 21 nt small RNAs suggested that they are not miRNAs.

The largest novel cluster was located at the Au76 locus, a pseudogene of *Rangap1*. In this cluster, 979 clones of oocyte small RNAs comprising 485 different sequences were mapped in a 1447 nt region (Fig. 14a). Most (91%) of them were 19-22 nt in length. Close inspection of this region revealed an inverted repeat structure in the Au76 pseudogene (Figures 14a and 14b). The small RNAs were exclusively mapped to this inverted repeat structure and were orientated in the same direction. Combined with the fact that production of these small RNAs requires *Dicer* and *Ago2* (see below), these observations strongly suggest that the small RNAs mapped to the Au76 pseudogene locus are siRNAs that are produced from a precursor RNA with an intra-molecular dsRNA structure (Fig. 14b). This type of siRNA cluster was designated as a "hairpin siRNA (hp-siRNA) cluster". Using in-house program that detects inverted repeat structure in the small RNA clusters, three other hp-siRNA clusters were identified in the mouse genome (Table 5).

Only one mammalian *Dicer* protein has been identified and shown to be involved in the miRNA pathway (Hutvagner et al. 2001). The abundance of siRNAs derived from the Au76 pseudogene was dramatically decreased in conditional *Dicer* knockout (KO) ovaries (Fig. 14c), suggesting that *Dicer* is also involved in the

production of siRNAs from intra-molecular dsRNA precursors. Regulation of the founding source gene by a pseudogene has been reported in mammalian cells, but the existence of such regulation has been controversial (Hirotsune et al. 2003; Gray et al. 2006). The abundance of *Rangap1* mRNA, which is the founding source gene of Au76 and shows ~90% nt identity, was increased ~4 fold in *Dicer* KO oocytes (Fig. 14c), suggesting that Au76 negatively regulates the founding source gene in trans through an RNAi mechanism.

Identification of hp-siRNA clusters led to ask whether other types of siRNA clusters were present. Formation of dsRNAs can occur by transcription of natural antisense transcripts from the same loci or different loci. Such hypothetical siRNA clusters were designated as “cis-nat-siRNA clusters” and “trans-nat-siRNA clusters”, respectively. Such clusters were searched using in-house program designed to detect the siRNA mapping patterns expected for these. Seventeen loci met our criteria of cis-nat-siRNA clusters (Table 5). An example of the predicted cis-nat-siRNA clusters was found at the *Pdzd11*/*Kif4* locus, where the two genes are orientated in a head-to-head manner (Fig. 15). At this locus, 135 clones of small RNAs comprising 117 different sequences were mapped to the first exon of the *Kif4* gene (Fig. 15). Of these small RNAs, 93% were 19-22 nt in length. A 5’ RACE analysis of the transcripts from growing oocytes revealed that the first exons of *Pdzd11* and *Kif4* overlapped (Fig.

15). Strikingly, almost all of the small RNAs mapped to this locus were derived from this overlapping region, suggesting that these small RNAs were produced from an inter-molecular dsRNA formed between the oppositely oriented transcripts. In *Dicer* mutants, levels of the siRNAs derived from this locus were decreased ~7 fold (Fig. 16A) and both *Pdzd11* and *Kif4* mRNA levels were increased ~1.5 fold (Fig. 16B), suggesting that the bidirectional overlapping transcription regulate *Pdzd11* and *Kif4* expression through RNAi.

A bioinformatics search predicted seven sets of trans-nat-siRNA clusters (Table 5), all of which were pairs of an mRNA and its pseudogene. Of the seven pseudogenes, two were observed in the 3' untranslated region of unrelated mRNAs, and five were observed in intergenic regions. A representative trans-nat-siRNA cluster pair consisted of the *Ppp4r1* gene on chr17 and its processed pseudogene on chr8 (Fig. 17). *Ppp4r1* and its pseudogene showed ~90 % nt identity. At the *Ppp4r1* locus, most of the 72 small RNAs (96% were 19-22 nt in length) comprising 63 sequences mapped exclusively to the exons of *Ppp4r1*, and all unique small RNAs were orientated in the same direction as the gene (Fig. 17), suggesting that *Ppp4r1* mRNA was the source of the siRNAs. In the pseudogene locus, 77 small RNAs (88% were 19-22 nt in length) comprising 69 sequences were mapped, and almost all unique small RNAs were orientated in the antisense direction of the *Ppp4r1* sequence (Fig. 17). Oocyte ESTs

mapped to this region were orientated in the same direction as the unique small RNAs, suggesting that the transcripts were the sources of the siRNAs. I did not observe small RNAs in the 3' region of the last exon of *Ppp4r1* (right side in Fig. 17 top), even though it spanned ~1/4 of the total mRNA length. In the region of the pseudogene corresponding to this 3' region (left side in Fig. 17 bottom), no EST was observed. Thus, siRNAs were produced exclusively from the region where dsRNAs could be formed between the mRNA and its expressed pseudogene. These results are consistent with idea that dsRNAs were the only source of the siRNAs. In *Dicer* KO oocytes, *Ppp4r1* mRNA level was increased ~1.5 fold (Fig. 16C). Together, these results strongly suggest that the antisense transcripts from the *Ppp4r1* pseudogene suppress *Ppp4r1* expression through RNAi.

Two small RNA mechanisms in the suppression of retrotransposons

The majority of the siRNAs and piRNAs in growing oocytes corresponded to retrotransposons. I therefore examined the possibility that the small RNA pathways suppress retrotransposons in oocytes. In *Dicer* KO oocytes, I observed that the transcript level of RLTR10 was elevated ~5 fold and that of MTA, which gives rise to more than 10% of the total polII transcripts in mouse oocytes (Peaston et al. 2004), was elevated ~3 fold (Fig. 18A). In *Mili* mutant oocytes, the transcript level of IAP

retrotransposons was elevated ~3.5 fold (Fig. 18B). These data suggest that both piRNA and siRNA pathways suppress retrotransposons in mouse oocytes and that each pathway clearly has preferred targets.

A peculiar siRNA cluster was observed in a retrotransposon-rich region. This ~50 kb locus contained 983 small RNAs comprising 637 different sequences (Fig. 19). The locus produced both 25-26 nt piRNAs and ~21 nt small RNAs and was annotated as both a piRNA and an hp-siRNA cluster. A large fraction of the ~21 nt small RNAs matched the RLTR10 sequence and were exclusively mapped to an ~2.5 kb inverted repeat structure located at the end of the piRNA cluster, suggesting that they were siRNAs produced from intra-molecular dsRNAs. Overlaps between piRNA and siRNA clusters were also observed in other cases (seven out of 152 piRNA clusters and 36 siRNA clusters) (Table 2).

siRNAs function through RNAi mechanism

To confirm that the siRNAs are assembled in the RNA-induced silencing complex (RISC), an AGO2-IP small RNA library and a total small RNA library were constructed from ovaries and the abundance of some siRNAs were examined in each libarary. All siRNAs examined were enriched in AGO2-IP library (Fig. 20a), suggesting that these siRNAs are bound to AGO2. In *Ago2* conditional KO oocyte, the

levels of mRNAs that are complementary to the siRNAs and those of retrotransposons that are elevated in *Dicer* KO were increased (Fig. 20b, c), suggesting that siRNAs repress target RNA through RNAi mechanism.

2. Small RNAs in oocytes

Characterization of small RNAs in growing oocytes

In order to obtain a comprehensive picture of endogenous siRNAs, more than 100,000 of small RNAs from mouse growing oocytes have been sequenced. The length distribution of the total small RNAs showed a bimodal pattern (Fig. 10A): one peak was observed at 21 nt, corresponding to the length of miRNAs and siRNAs, and the other at 25-26 nt, corresponding to the length of piRNAs, which are a distinct class of small RNAs bound to Piwi family proteins (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Watanabe et al. 2006). Annotation of the small RNAs revealed that both the 21 nt and 25-26 nt small RNAs were mainly derived from repeat sequences, most of which were retrotransposons (Fig. 11A, Table 4) and were quite diverse in sequence (21969 clones of 21 nt small RNAs comprising 10194 different sequences and 18572 clones of 25-26 nt small RNAs comprising 12006 different sequences).

In order to examine expression of some components of the small RNA pathways in growing oocytes, RT-PCR and Western blotting were carried out. *Ago2*, *Ago3* and *Dicer*, which are the components of the siRNA and miRNA pathways, were expressed at high levels throughout oocyte growth (Fig. 10B). Of the three mouse Piwi family proteins, MILI was predominantly expressed at early stages of oocyte growth

(Fig. 10C). However, MIWI or MIWI2 were not detected in growing oocytes. To examine whether the 25-26 nt small RNAs in growing oocytes are piRNAs bound to MILI, a total small RNA library and a MILI-immunoprecipitated (IP) small RNA library were constructed from ovaries, and the abundance of several 25-26 nt small RNAs were examined in each library. All 25-26 nt small RNAs were enriched in the MILI-IP library relative to the control miRNAs and 21 nt small RNAs (Fig. 10D), indicating that the 25-26 nt small RNAs in oocytes are mostly MILI-bound piRNAs.

Biogenesis and role of endogenous siRNAs

Genomic mapping of the oocyte small RNAs (excluding those with more than 10 hits to the genome, which are likely to be repeat sequences) revealed the presence of a total of 444 small RNA clusters. piRNAs are known to be mapped to the genome in clusters (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Watanabe et al. 2006), and we identified 152 piRNA clusters by sequencing and mapping of MILI-IP small RNAs (Fig. 11B). Surprisingly, many of the largest clusters, which were determined by the number of small RNAs, were not included in these piRNA clusters (Fig. 12A). Furthermore, the length distribution of the small RNAs constituting these non-piRNA clusters was centered at 21 nt (Fig. 12B), which is the length of miRNAs and siRNAs. However, the lack of a short stem-loop structure (Fig.

13), which is a characteristics of miRNA precursors, in most of the genomic sequences encompassing the 21 nt small RNAs suggested that they are not miRNAs.

The largest novel cluster was located at the Au76 locus, a pseudogene of *Rangap1*. In this cluster, 979 clones of oocyte small RNAs comprising 485 different sequences were mapped in a 1447 nt region (Fig. 14a). Most (91%) of them were 19-22 nt in length. Close inspection of this region revealed an inverted repeat structure in the Au76 pseudogene (Figures 14a and 14b). The small RNAs were exclusively mapped to this inverted repeat structure and were orientated in the same direction. Combined with the fact that production of these small RNAs requires *Dicer* and *Ago2* (see below), these observations strongly suggest that the small RNAs mapped to the Au76 pseudogene locus are siRNAs that are produced from a precursor RNA with an intra-molecular dsRNA structure (Fig. 14b). This type of siRNA cluster was designated as a "hairpin siRNA (hp-siRNA) cluster". Using in-house program that detects inverted repeat structure in the small RNA clusters, three other hp-siRNA clusters were identified in the mouse genome (Table 5).

Only one mammalian *Dicer* protein has been identified and shown to be involved in the miRNA pathway (Hutvagner et al. 2001). The abundance of siRNAs derived from the Au76 pseudogene was dramatically decreased in conditional *Dicer* knockout (KO) ovaries (Fig. 14c), suggesting that *Dicer* is also involved in the

production of siRNAs from intra-molecular dsRNA precursors. Regulation of the founding source gene by a pseudogene has been reported in mammalian cells, but the existence of such regulation has been controversial (Hirotsune et al. 2003; Gray et al. 2006). The abundance of *Rangap1* mRNA, which is the founding source gene of Au76 and shows ~90% nt identity, was increased ~4 fold in *Dicer* KO oocytes (Fig. 14c), suggesting that Au76 negatively regulates the founding source gene in trans through an RNAi mechanism.

Identification of hp-siRNA clusters led to ask whether other types of siRNA clusters were present. Formation of dsRNAs can occur by transcription of natural antisense transcripts from the same loci or different loci. Such hypothetical siRNA clusters were designated as “cis-nat-siRNA clusters” and “trans-nat-siRNA clusters”, respectively. Such clusters were searched using in-house program designed to detect the siRNA mapping patterns expected for these. Seventeen loci met our criteria of cis-nat-siRNA clusters (Table 5). An example of the predicted cis-nat-siRNA clusters was found at the *Pdzd11*/*Kif4* locus, where the two genes are orientated in a head-to-head manner (Fig. 15). At this locus, 135 clones of small RNAs comprising 117 different sequences were mapped to the first exon of the *Kif4* gene (Fig. 15). Of these small RNAs, 93% were 19-22 nt in length. A 5’ RACE analysis of the transcripts from growing oocytes revealed that the first exons of *Pdzd11* and *Kif4* overlapped (Fig.

15). Strikingly, almost all of the small RNAs mapped to this locus were derived from this overlapping region, suggesting that these small RNAs were produced from an inter-molecular dsRNA formed between the oppositely oriented transcripts. In *Dicer* mutants, levels of the siRNAs derived from this locus were decreased ~7 fold (Fig. 16A) and both *Pdzd11* and *Kif4* mRNA levels were increased ~1.5 fold (Fig. 16B), suggesting that the bidirectional overlapping transcription regulate *Pdzd11* and *Kif4* expression through RNAi.

A bioinformatics search predicted seven sets of trans-nat-siRNA clusters (Table 5), all of which were pairs of an mRNA and its pseudogene. Of the seven pseudogenes, two were observed in the 3' untranslated region of unrelated mRNAs, and five were observed in intergenic regions. A representative trans-nat-siRNA cluster pair consisted of the *Ppp4r1* gene on chr17 and its processed pseudogene on chr8 (Fig. 17). *Ppp4r1* and its pseudogene showed ~90 % nt identity. At the *Ppp4r1* locus, most of the 72 small RNAs (96% were 19-22 nt in length) comprising 63 sequences mapped exclusively to the exons of *Ppp4r1*, and all unique small RNAs were orientated in the same direction as the gene (Fig. 17), suggesting that *Ppp4r1* mRNA was the source of the siRNAs. In the pseudogene locus, 77 small RNAs (88% were 19-22 nt in length) comprising 69 sequences were mapped, and almost all unique small RNAs were orientated in the antisense direction of the *Ppp4r1* sequence (Fig. 17). Oocyte ESTs

mapped to this region were orientated in the same direction as the unique small RNAs, suggesting that the transcripts were the sources of the siRNAs. I did not observe small RNAs in the 3' region of the last exon of *Ppp4r1* (right side in Fig. 17 top), even though it spanned ~1/4 of the total mRNA length. In the region of the pseudogene corresponding to this 3' region (left side in Fig. 17 bottom), no EST was observed. Thus, siRNAs were produced exclusively from the region where dsRNAs could be formed between the mRNA and its expressed pseudogene. These results are consistent with idea that dsRNAs were the only source of the siRNAs. In *Dicer* KO oocytes, *Ppp4r1* mRNA level was increased ~1.5 fold (Fig. 16C). Together, these results strongly suggest that the antisense transcripts from the *Ppp4r1* pseudogene suppress *Ppp4r1* expression through RNAi.

Two small RNA mechanisms in the suppression of retrotransposons

The majority of the siRNAs and piRNAs in growing oocytes corresponded to retrotransposons. I therefore examined the possibility that the small RNA pathways suppress retrotransposons in oocytes. In *Dicer* KO oocytes, I observed that the transcript level of RLTR10 was elevated ~5 fold and that of MTA, which gives rise to more than 10% of the total polII transcripts in mouse oocytes (Peaston et al. 2004), was elevated ~3 fold (Fig. 18A). In *Mili* mutant oocytes, the transcript level of IAP

retrotransposons was elevated ~3.5 fold (Fig. 18B). These data suggest that both piRNA and siRNA pathways suppress retrotransposons in mouse oocytes and that each pathway clearly has preferred targets.

A peculiar siRNA cluster was observed in a retrotransposon-rich region. This ~50 kb locus contained 983 small RNAs comprising 637 different sequences (Fig. 19). The locus produced both 25-26 nt piRNAs and ~21 nt small RNAs and was annotated as both a piRNA and an hp-siRNA cluster. A large fraction of the ~21 nt small RNAs matched the RLTR10 sequence and were exclusively mapped to an ~2.5 kb inverted repeat structure located at the end of the piRNA cluster, suggesting that they were siRNAs produced from intra-molecular dsRNAs. Overlaps between piRNA and siRNA clusters were also observed in other cases (seven out of 152 piRNA clusters and 36 siRNA clusters) (Table 2).

siRNAs function through RNAi mechanism

To confirm that the siRNAs are assembled in the RNA-induced silencing complex (RISC), an AGO2-IP small RNA library and a total small RNA library were constructed from ovaries and the abundance of some siRNAs were examined in each libarary. All siRNAs examined were enriched in AGO2-IP library (Fig. 20a), suggesting that these siRNAs are bound to AGO2. In *Ago2* conditional KO oocyte, the

levels of mRNAs that are complementary to the siRNAs and those of retrotransposons that are elevated in *Dicer* KO were increased (Fig. 20b, c), suggesting that siRNAs repress target RNA through RNAi mechanism.

Acknowledgements

I am grateful to Prof. Hiroyuki Sasaki for his help and encouragement throughout this work. I am also grateful to Drs. Takashi Sado, Ken-ichiro Hata, Hiroyasu Furuumi and Kenji Ichiyanagi for advice. I thank Dr. Stomi Miyagawa for discussion and providing me Mili KO mouse samples. I thank Yasushi Totoki for instruction of bioinformatics and his analysis. I thank Atsushi Toyoda and Yoshiyuki Sakaki for 454 sequencing. I thank Dr. Masahiro Kaneda for providing me Dicer KO and Ago2 KO mouse samples. I thank Yayoi Obata for her collection of many growing oocytes. I thank Shinichiro Chuma for providing me many antibodies and his instruction and analysis. I thank Shinichi Tomizawa for his Bisulfite sequencing. I thank Yuko Hoki for her ES cell injection and transplantation. I thank prof. Yumiko Saga, Hitomi Suzuki and Masayuki Oginuma for generation of transgenic mouse. I would like to express my appreciation to all members of the Prof. Sasaki's laboratory.

References

Ambros, V. and X. Chen. 2007. The regulation of genes and genomes by small RNAs. *Development* **134**: 1635-41.

Aravin, A., D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M.J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J.J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, and T. Tuschl. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**: 203-7.

Aravin, A.A., G.J. Hannon, and J. Brennecke. 2007a. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**: 761-4.

Aravin, A.A., R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K.F. Toth, T. Bestor, and G.J. Hannon. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**: 785-99.

Aravin, A.A., R. Sachidanandam, A. Girard, K. Fejes-Toth, and G.J. Hannon. 2007b. Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science*.

Babiarz, J.E., J.G. Ruby, Y. Wang, D.P. Bartel, and R. Blelloch. 2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev* **22**: 2773-85.

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-97.

Brennecke, J., A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G.J. Hannon. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**: 1089-103.

Carmell, M.A., A. Girard, H.J. van de Kant, D. Bourc'his, T.H. Bestor, D.G. de Rooij, and G.J. Hannon. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12**: 503-14.

Chuma, S., M. Hosokawa, K. Kitamura, S. Kasai, M. Fujioka, M. Hiyoshi, K. Takamune, T. Noce, and N. Nakatsuji. 2006. Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. *Proc Natl Acad Sci U S A* **103**: 15894-9.

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

Deng, W. and H. Lin. 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell* **2**: 819-30.

Elbashir, S.M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian

cells. *Nature* **411**: 494-8.

Ghildiyal, M., H. Seitz, M.D. Horwich, C. Li, T. Du, S. Lee, J. Xu, E.L. Kittler, M.L. Zapp, Z. Weng, and P.D. Zamore. 2008. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320**: 1077-81.

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

Gray, T.A., A. Wilson, P.J. Fortin, and R.D. Nicholls. 2006. The putatively functional Mkrn1-p1 pseudogene is neither expressed nor imprinted, nor does it regulate its source gene in trans. *Proc Natl Acad Sci U S A* **103**: 12039-44.

Grivna, S.T., E. Beyret, Z. Wang, and H. Lin. 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* **20**: 1709-14.

Gunawardane, L.S., K. Saito, K.M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M.C. Siomi. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* **315**: 1587-90.

Hamilton, A.J. and D.C. Baulcombe. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950-2.

Hirotsune, S., N. Yoshida, A. Chen, L. Garrett, F. Sugiyama, S. Takahashi, K. Yagami, A. Wynshaw-Boris, and A. Yoshiki. 2003. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* **423**: 91-6.

Hutvagner, G., J. McLachlan, A.E. Pasquinelli, E. Balint, T. Tuschl, and P.D. Zamore. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834-8.

Kato, Y., M. Kaneda, K. Hata, K. Kumaki, M. Hisano, Y. Kohara, M. Okano, E. Li, M. Nozaki, and H. Sasaki. 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**: 2272-80.

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

Kuramochi-Miyagawa, S., T. Kimura, T.W. Ijiri, T. Isobe, N. Asada, Y. Fujita, M. Ikawa, N. Iwai, M. Okabe, W. Deng, H. Lin, Y. Matsuda, and T. Nakano. 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**: 839-49.

Kuramochi-Miyagawa, S., T. Kimura, K. Yomogida, A. Kuroiwa, Y. Tadokoro, Y. Fujita, M. Sato, Y. Matsuda, and T. Nakano. 2001. Two mouse piwi-related genes: miwi and mili. *Mech Dev* **108**: 121-33.

Kuramochi-Miyagawa, S., T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T.W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki, and T. Nakano. 2008. DNA methylation of

retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**: 908-17.

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

Lau, N.C., A.G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D.P. Bartel, and R.E. Kingston. 2006. Characterization of the piRNA complex from rat testes. *Science* **313**: 363-7.

Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-54.

Liu, J., M.A. Carmell, F.V. Rivas, C.G. Marsden, J.M. Thomson, J.J. Song, S.M. Hammond, L. Joshua-Tor, and G.J. Hannon. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**: 1437-41.

Okamura, K., W.J. Chung, J.G. Ruby, H. Guo, D.P. Bartel, and E.C. Lai. 2008. The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453**: 803-6.

Pak, J. and A. Fire. 2007. Distinct populations of primary and secondary effectors during RNAi in C. elegans. *Science* **315**: 241-4.

Pane, A., K. Wehr, and T. Schupbach. 2007. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev Cell* **12**: 851-62.

Peaston, A.E., A.V. Evsikov, J.H. Graber, W.N. de Vries, A.E. Holbrook, D. Solter, and B.B. Knowles. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* **7**: 597-606.

Ruby, J.G., C. Jan, C. Player, M.J. Axtell, W. Lee, C. Nusbaum, H. Ge, and D.P. Bartel. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell* **127**: 1193-207.

Sijen, T., F.A. Steiner, K.L. Thijssen, and R.H. Plasterk. 2007. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* **315**: 244-7.

Soper, S.F., G.W. van der Heijden, T.C. Hardiman, M. Goodheart, S.L. Martin, P. de Boer, and A. Bortvin. 2008. Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev Cell* **15**: 285-97.

Stein, P., P. Svoboda, M. Anger, and R.M. Schultz. 2003. RNAi: mammalian oocytes do it without RNA-dependent RNA polymerase. *Rna* **9**: 187-92.

Svoboda, P., P. Stein, H. Hayashi, and R.M. Schultz. 2000. Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. *Development* **127**: 4147-56.

Tabara, H., M. Sarkissian, W.G. Kelly, J. Fleenor, A. Grishok, L. Timmons, A. Fire, and C.C. Mello. 1999. The rde-1 gene, RNA interference, and transposon silencing in C.

elegans. *Cell* **99**: 123-32.

Tang, F., M. Kaneda, D. O'Carroll, P. Hajkova, S.C. Barton, Y.A. Sun, C. Lee, A. Tarakhovsky, K. Lao, and M.A. Surani. 2007. Maternal microRNAs are essential for mouse zygotic development. *Genes Dev* **21**: 644-8.

Vaucheret, H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev* **20**: 759-71.

Vazquez, F. 2006. Arabidopsis endogenous small RNAs: highways and byways. *Trends Plant Sci* **11**: 460-8.

Watanabe, T., A. Takeda, T. Tsukiyama, K. Mise, T. Okuno, H. Sasaki, N. Minami, and H. Imai. 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* **20**: 1732-43.

Watanabe, T., Y. Totoki, H. Sasaki, N. Minami, and H. Imai. 2007. Analysis of Small RNA Profiles During Development. *Methods Enzymol* **427C**: 155-169.

Wianny, F. and M. Zernicka-Goetz. 2000. Specific interference with gene function by double-stranded RNA in early mouse development. *Nat Cell Biol* **2**: 70-5.

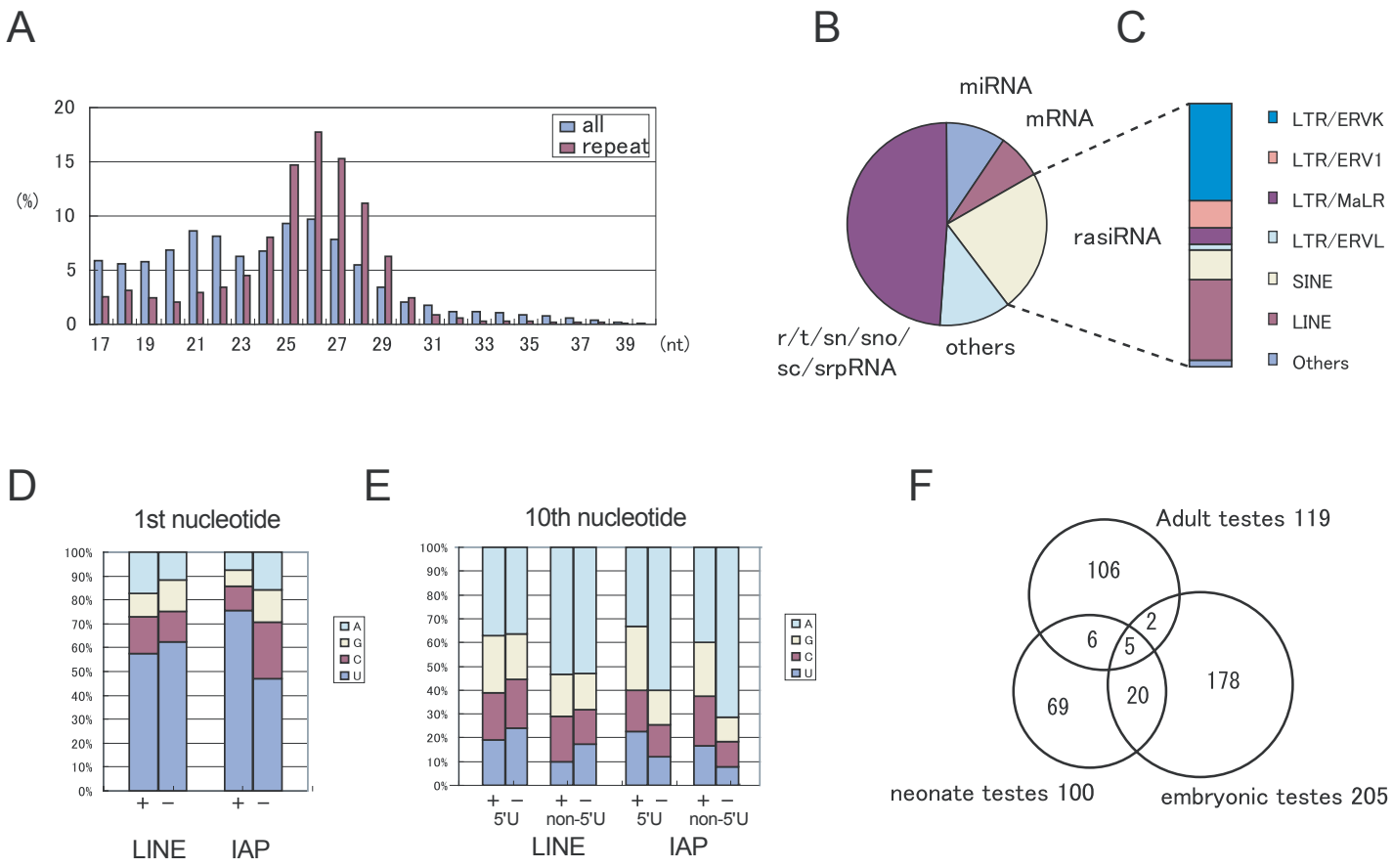
Yan, W., L. Ma, P. Stein, S.A. Pangas, K.H. Burns, Y. Bai, R.M. Schultz, and M.M. Matzuk. 2005. Mice deficient in oocyte-specific oligoadenylate synthetase-like protein OAS1D display reduced fertility. *Mol Cell Biol* **25**: 4615-24.

Yigit, E., P.J. Batista, Y. Bei, K.M. Pang, C.C. Chen, N.H. Tolia, L. Joshua-Tor, S. Mitani, M.J. Simard, and C.C. Mello. 2006. Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**: 747-57.

Zhao, Y., J.A. Stuckey, D.L. Lohse, and J.E. Dixon. 1997. Expression, characterization, and crystallization of a member of the novel phospholipase D family of phosphodiesterases. *Protein Sci* **6**: 2655-8.

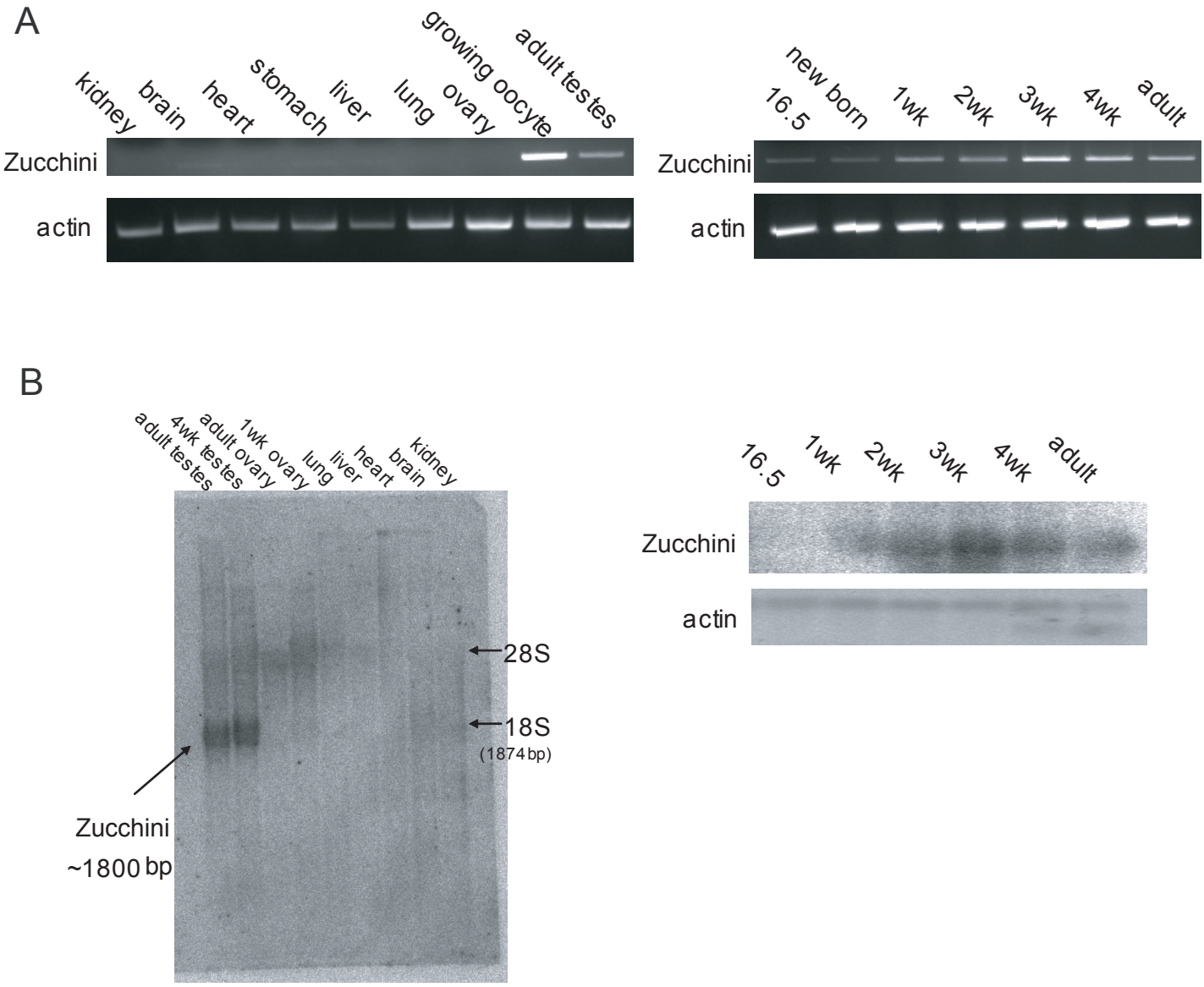
Zilberman, D., X. Cao, and S.E. Jacobsen. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**: 716-9.

Figure 1



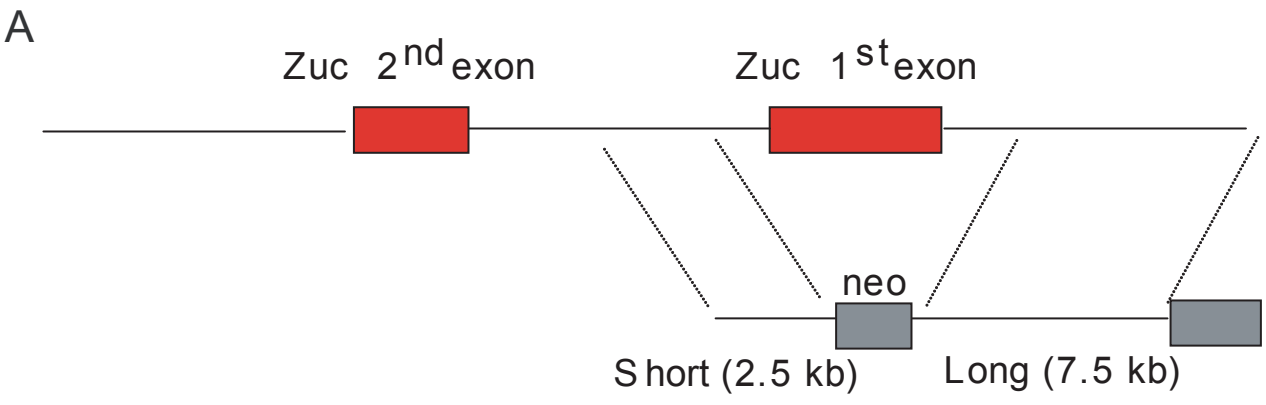
(A) A total of 127,997 small RNAs were sequenced from E12.5-19.5 fetal germ cells. The size distributions (in nucleotides, nt) of the total small RNAs and rasiRNAs are shown by blue and purple bars.
(B, C) Genomic annotation of the small RNAs (B) and the number of piRNA sequences.
(D, E) Comparison of the first (D) and tenth (E) nucleotides in the sense (+) and antisense (−) piRNAs. Nucleotide biases were calculated for the Gf type Line-1 and IAP piRNAs analyzed in Table S2. Those piRNA classes that contain and lack a 5′ U are shown separately in (E).
(F) Venn diagram of the piRNAs in adult(Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Watanabe et al. 2006), neonatal (Aravin et al 2007) and fetal (this study) testes.
(G, H) Distribution of piRNAs corresponding to type Gf Line-1 (G) and IAPIΔ1 (H).

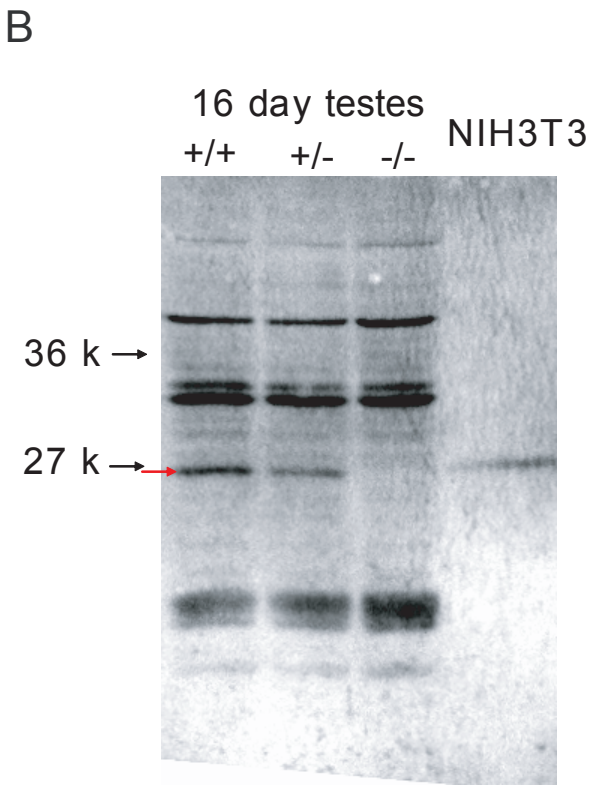
Figure 2

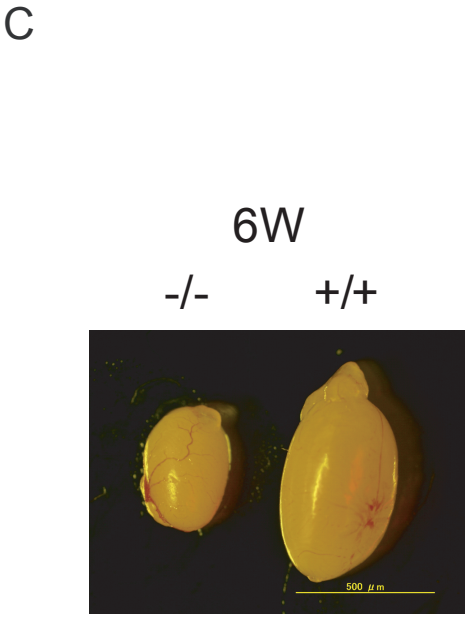


(A) RT-PCR analysis of Zucchini mRNA. Various tissues (left) and testis developmental stages (right) were examined.
(B) Northern Blot analysis of Zucchini mRNA. Tissues (left) and testis developmental stages (right)

Figure 3

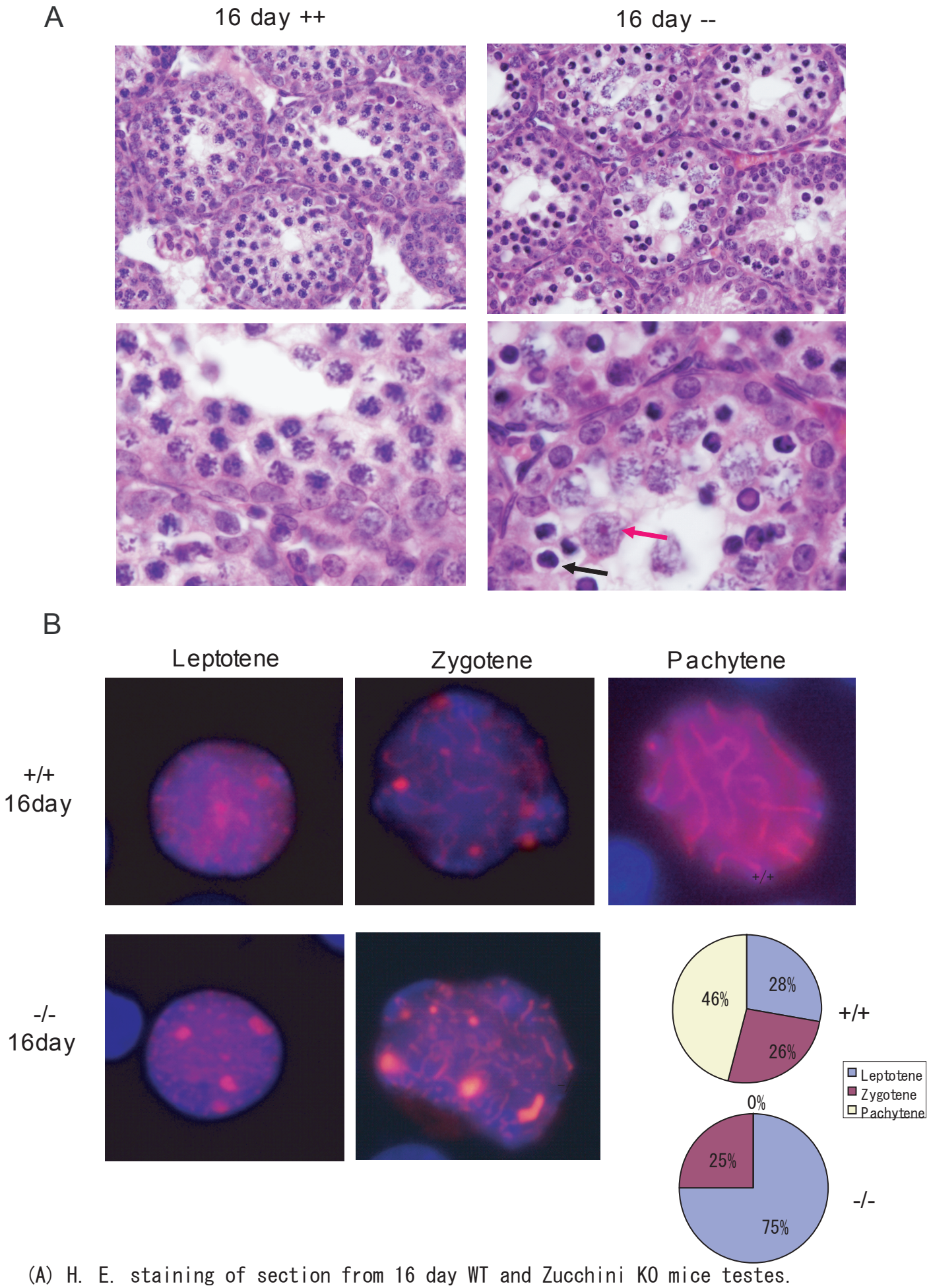






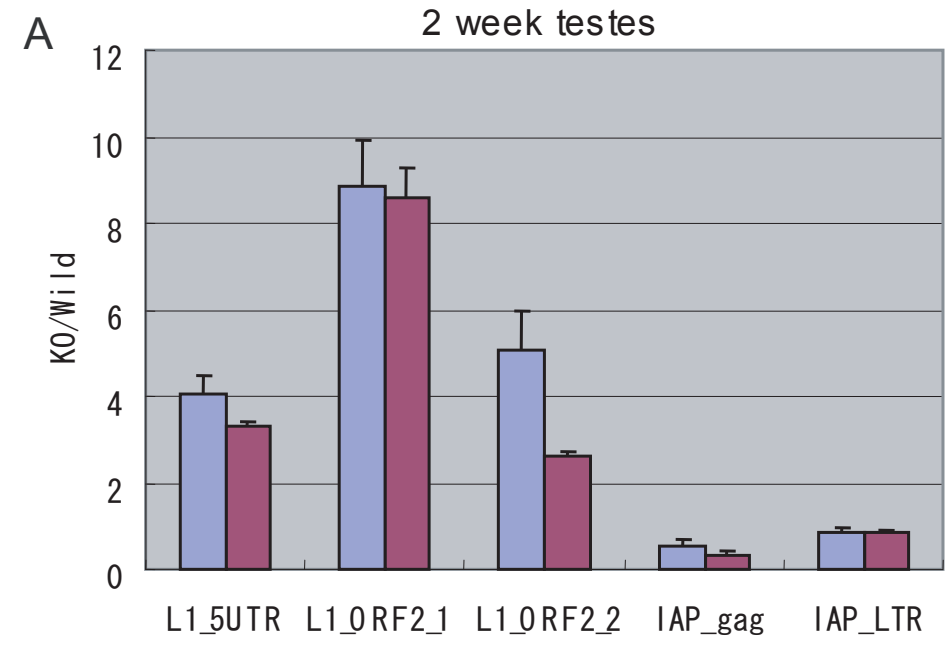
(A) Structure of Zucchini gene. Targeting construct is also illustrated.
(B) Western Blot analysis of Zucchini protein. Lysates of 16 day testis and NIH3T3 transfected with Zucchini expression vector were subjected to 15% acrylamide SDS-PAGE. About 20 μg of testis proteins or 200 ng of NIH3T3 protein were used. Red arrow indicates the band of Zucchini Protein.
(C) 6-week testes from Zucchini KO mouse and wild type mouse.

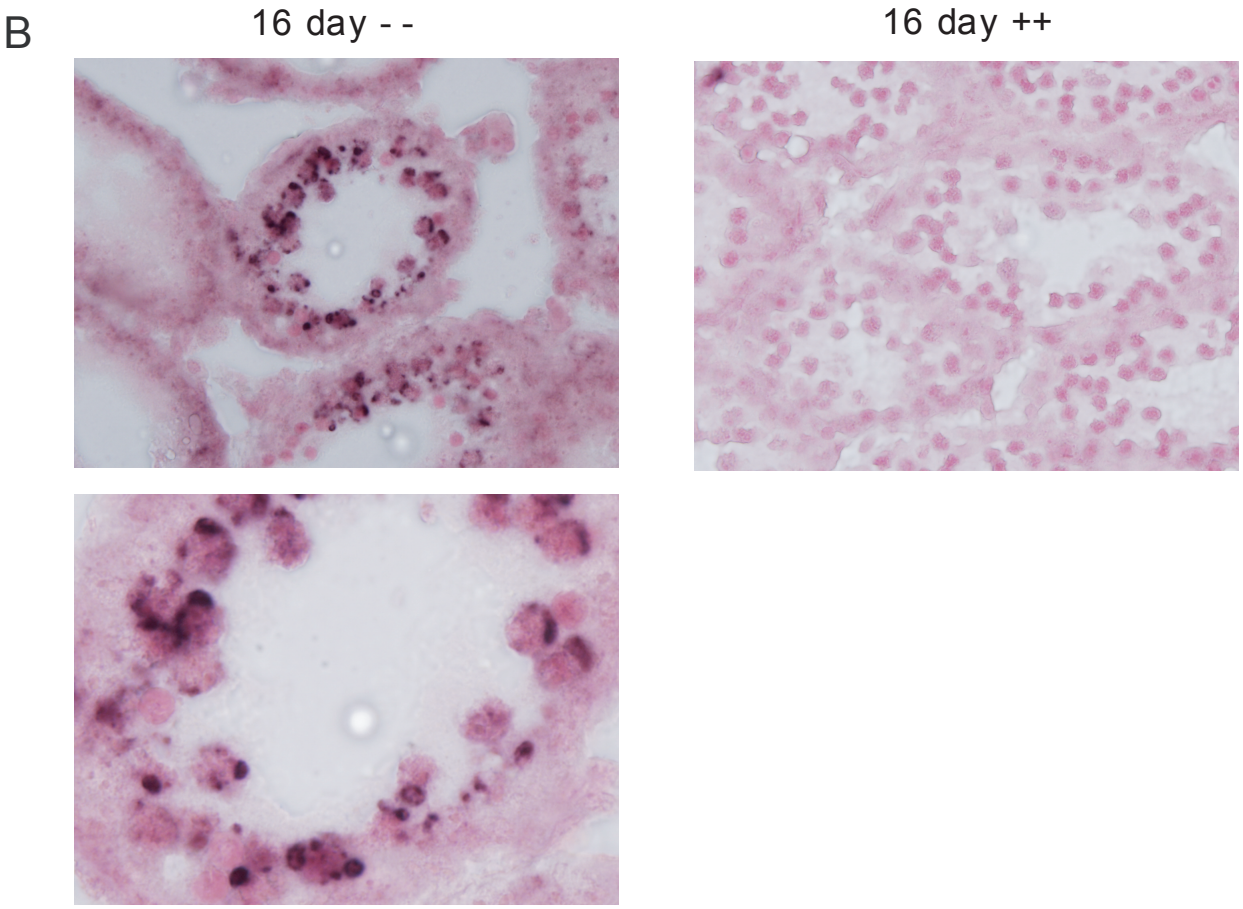
Figure 4



(A) H. E. staining of section from 16 day WT and Zucchini KO mice testes.
(B) SCP3 staining of testicular cells. Graph on the right shows the paercentage of germ cells at each stage of meiosis.

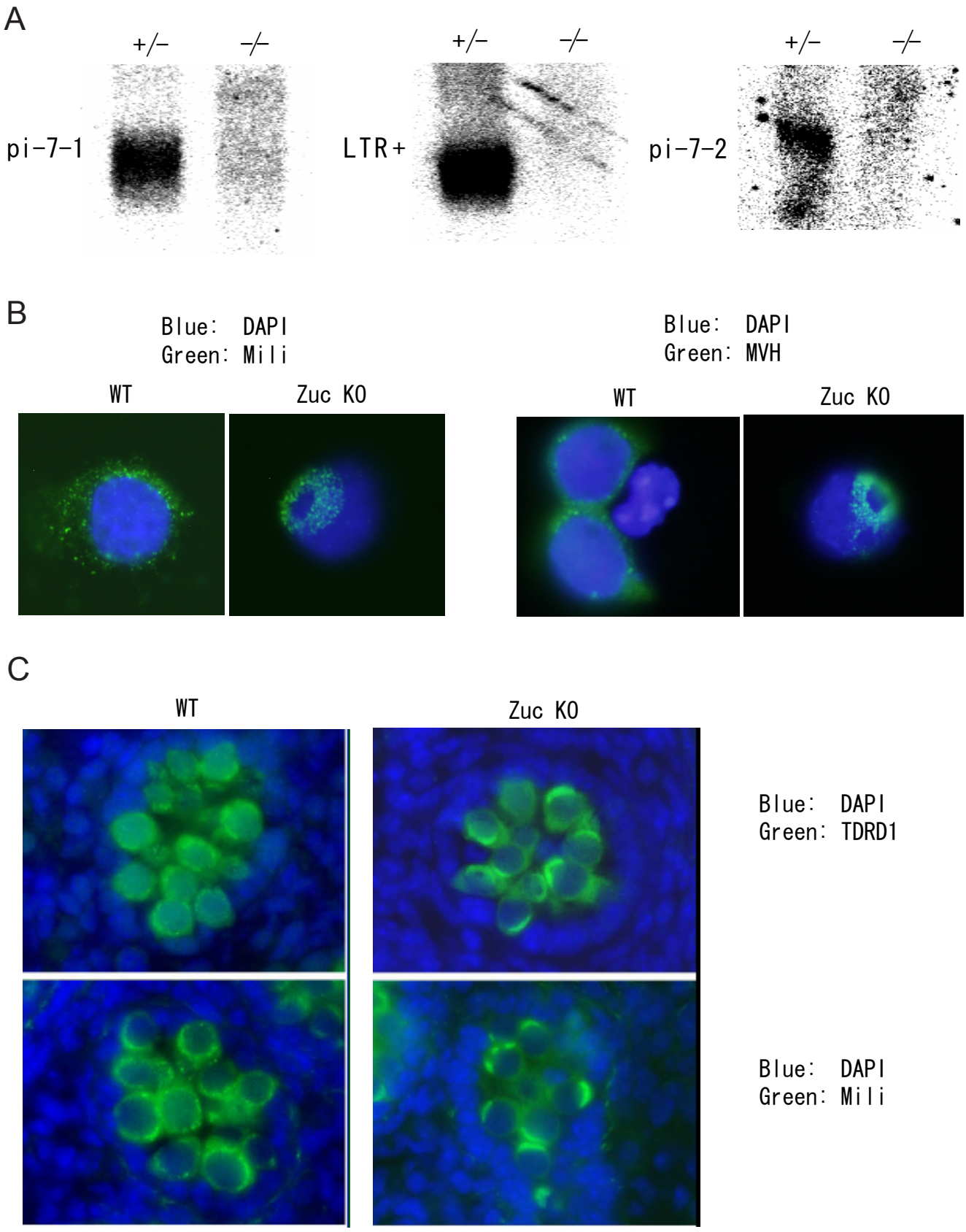
Figure 5





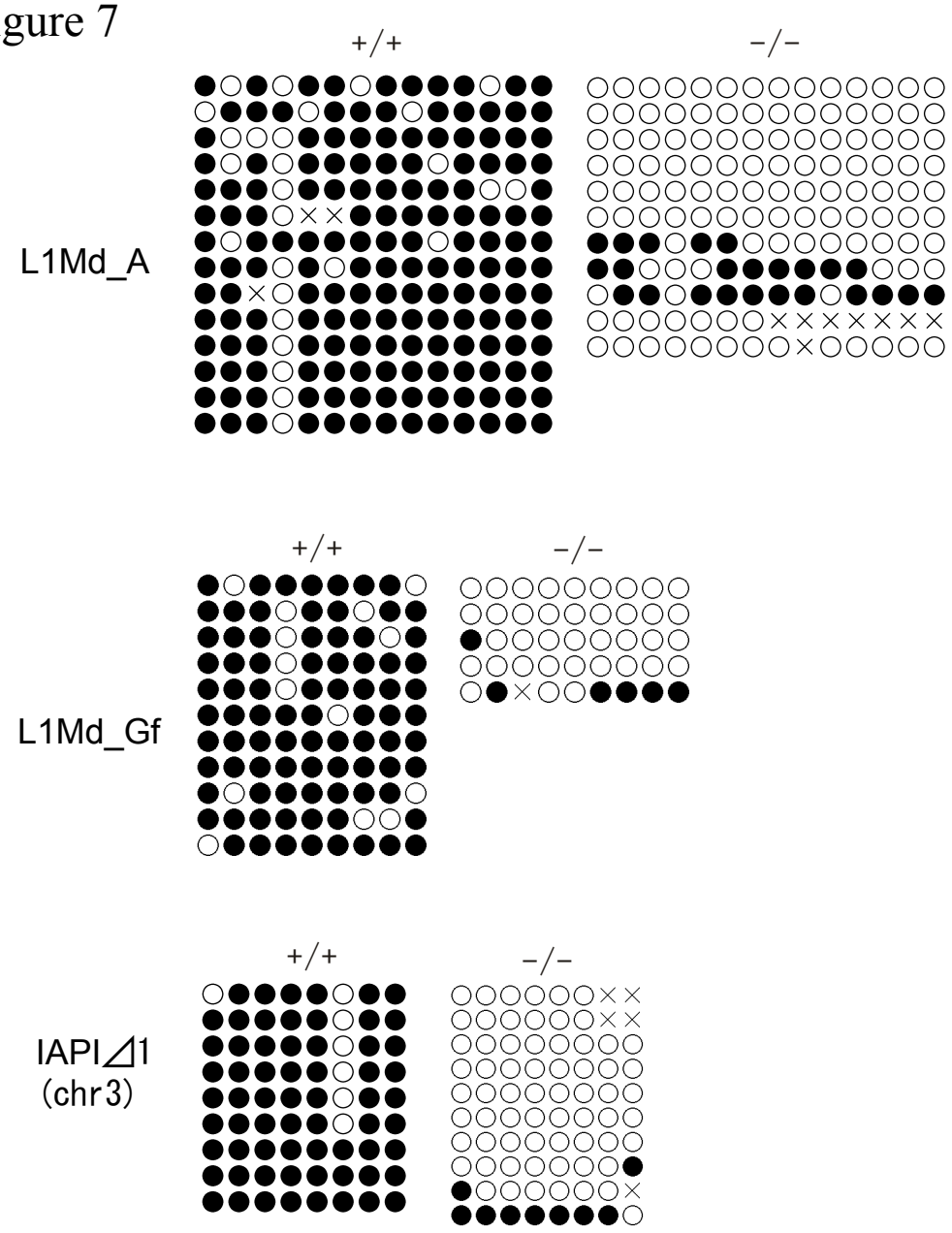
(A) qPCR analysis of L1Md and IAP retrotransposons in 2w Zucchini KO and WT mouse testes. Testes from two individuals (red and blue bars) were analyzed.
(B) In situ hybridization analysis using the probe that detects L1 RNA in 16 day KO and WT testes.

Figure 6



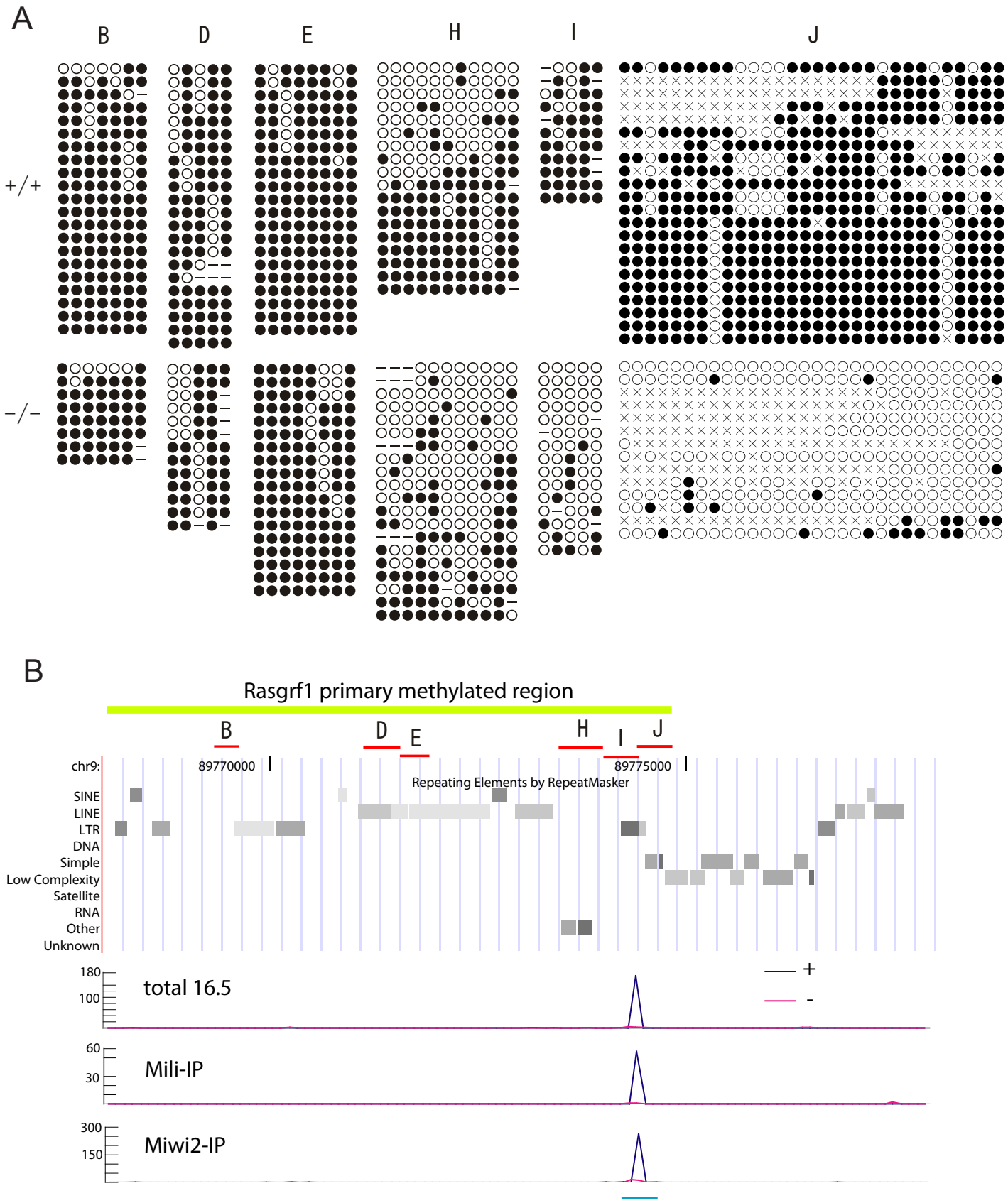
(A) Northern blot analysis of fetal piRNAs in 16.5 dpc Zucchini +/- and -/- testes.
(B) Immunocytochemical analysis of piRNA pathway components in 16.5 dpc Zucchini +/+ and -/- gonocytes. The presence of γ-Tubulin, which is a marker of centrosome, at the center of doughnut shape signal was confirmed (data not shown).
(C) Immunohistochemical analysis of piRNA pathway components in 16.5 dpc Zucchini +/+ and -/- testes.

Figure 7



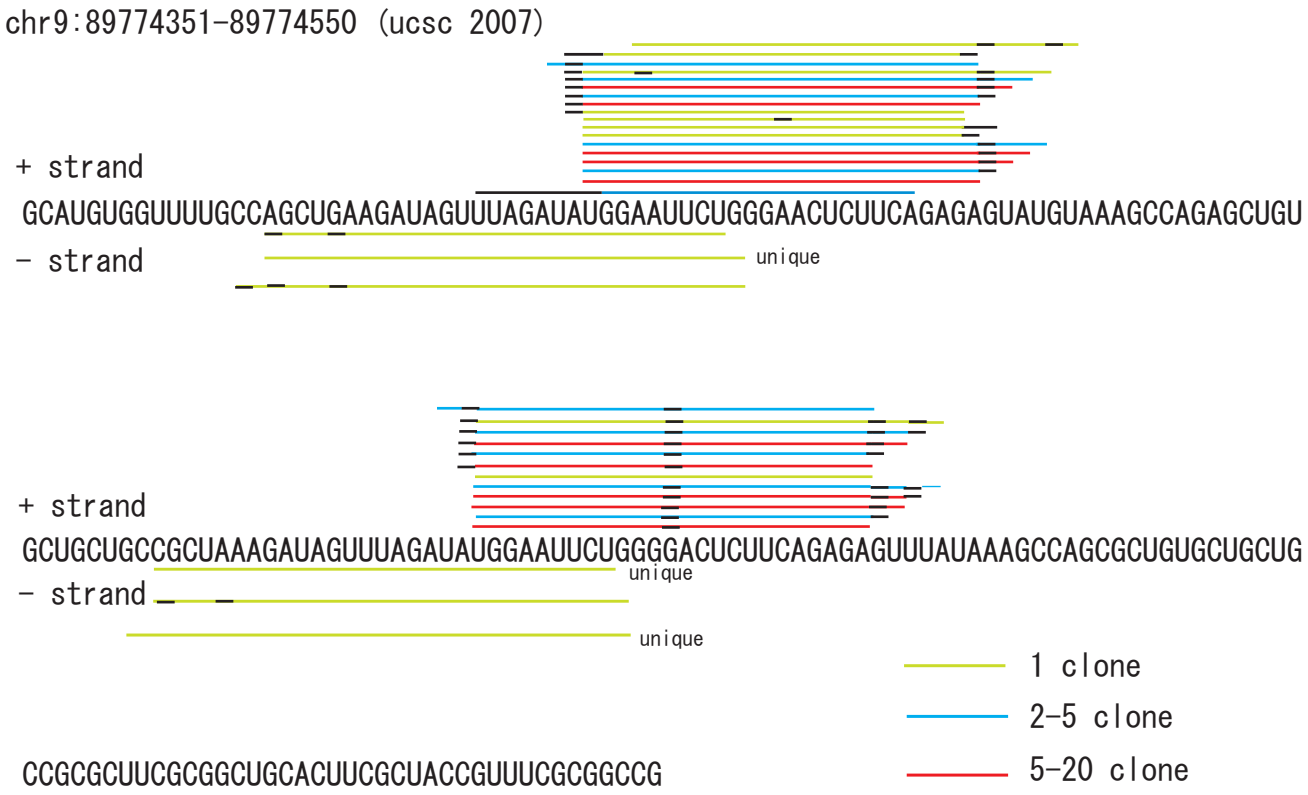
Bisulfite sequencing analysis of retrotransposons in wild type and Zuchini KO spermatogonia.

Figure 8



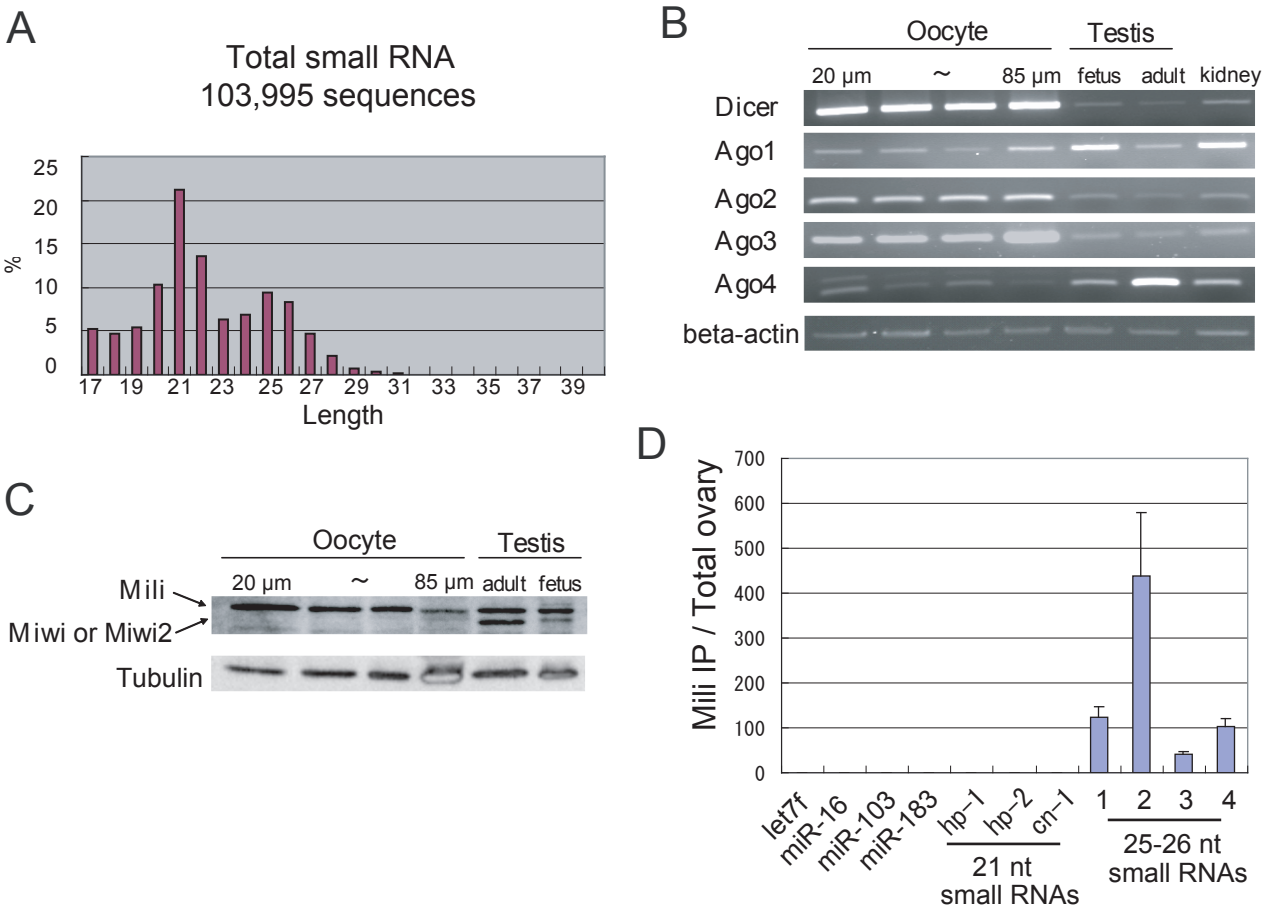
(A) Bisulfite sequencing analysis of Rasgrf1 methylated region in wild type and KO spermatogonia. Positions analyzed are indicated in (B).
(B) piRNAs mapped in Rasgrf1 locus are shown. piRNAs identified from E16.5 testes (Aravin et al. 2008) were blasted against Rasgrf11 locus using criteria that detect 21 nt stretch of perfect match. Y-axis shows the number of small RNAs mapped per 100 bp.

Figure 9



Rasgrf1 region where small RNAs are mapped (indicated by blue bar in Fig.8B). 16.5 dpc total small RNAs are mapped. Small RNAs are represented by bars. Green, blue and red represent number of clone. Black indicate mismatch. Unique small RNAs are indicated.

Figure 10



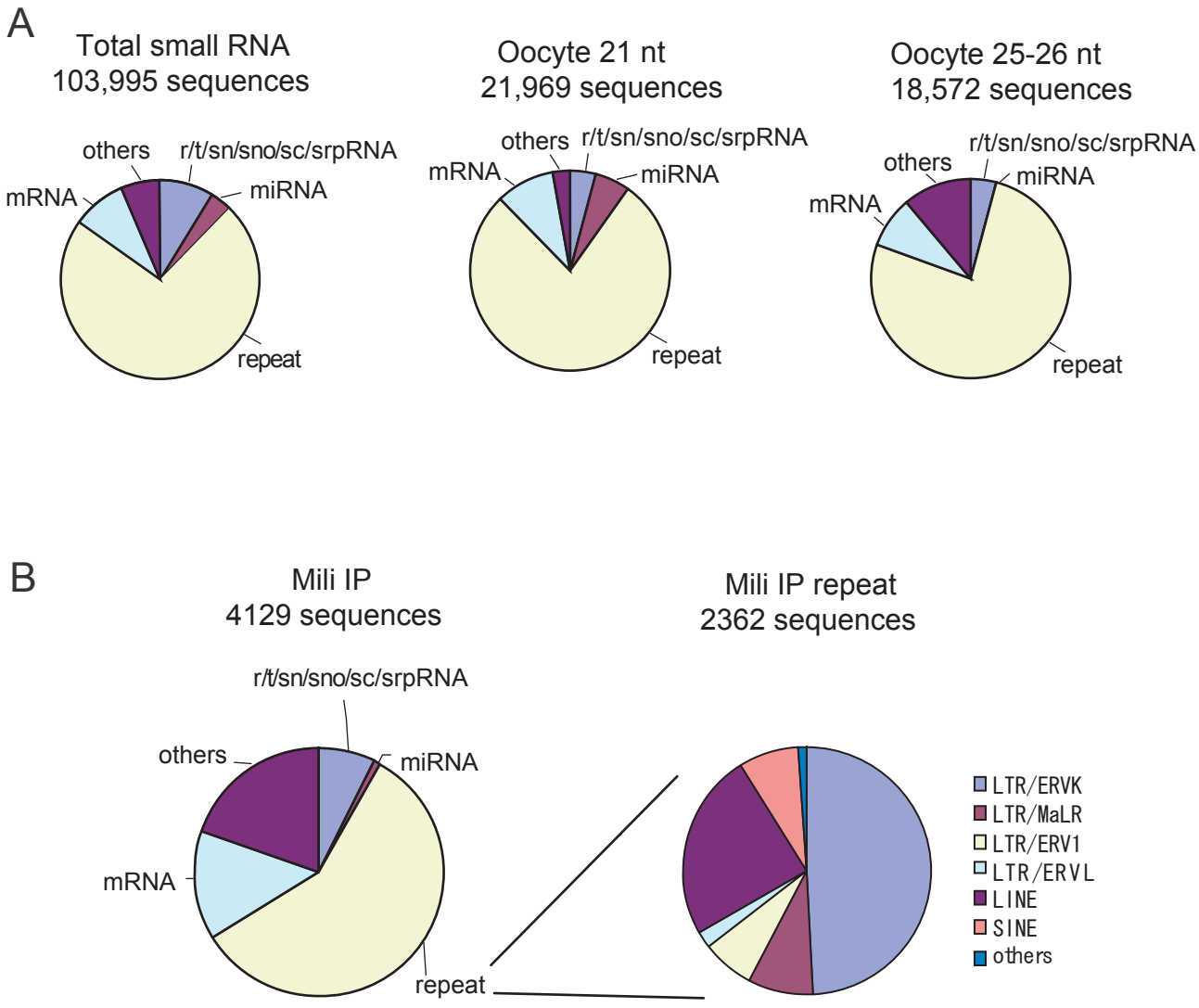
(A) Length distribution of total small RNAs from oocytes.

(B) RT-PCR analysis of genes involved in the small RNA pathway. The oocytes with the diameter of ~20 , ~40, ~60 and ~85 μm (fully grown oocytes) were analyzed. For comparision, mRNAs in testis and kidney were also analyzed.

(C) Western blot analysis of Piwi family proteins in growing oocytes. The antibody used here detects all three Piwi family proteins. Positions of these proteins on the membrane are indicated on the left. The membrane was reprobed with the antibody against tubulin as a loading control.

(D) Quantitative RT-PCR analysis of selected 21 nt and 25-26 nt small RNAs in total and Mili-IP small RNAs from P8 ovaries. For each RNA species tested, the amount in the Mili-IP library was divided by that in the total small RNA library. Error bars represent S.E. (n=3). Some miRNAs were also tested. Amplified products were sequenced and confirmed to have the correct sequences.

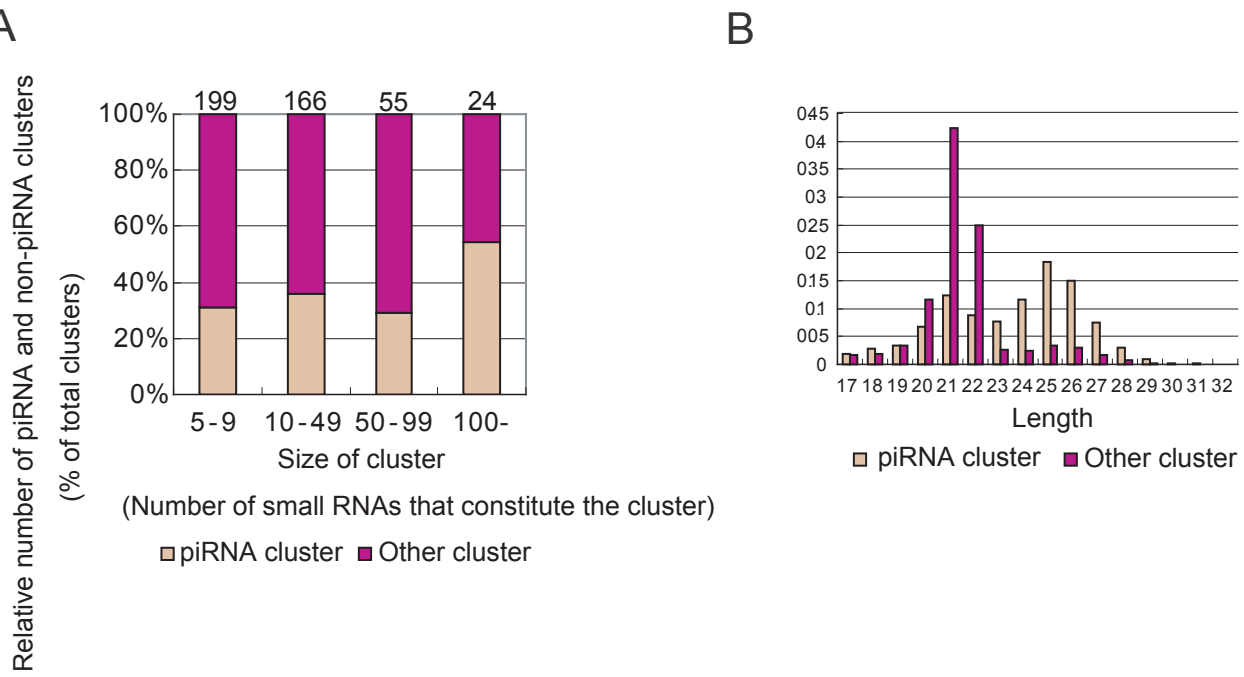
Figure 11



(A) Annotation of total (left), 21 nt (middle) and 25-26 nt (right) small RNAs from growing oocytes.

(B) Annotation of Mili-IP small RNAs from P8 ovaries.

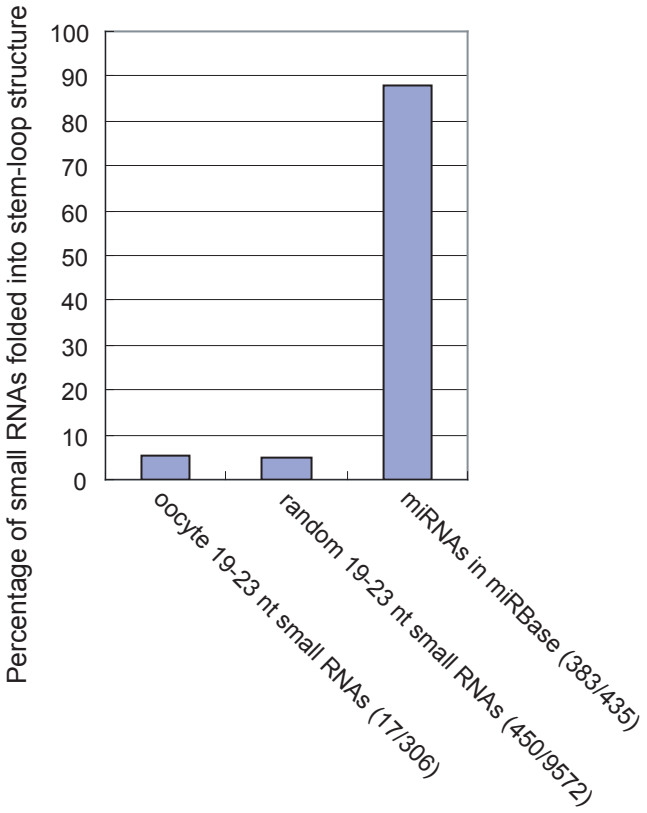
Figure 12



(A) The ratio of the number of piRNA clusters to that of the other small RNA clusters in oocytes. About half of the clusters with more than 100 mapped small RNAs were non-piRNA clusters. The numbers above the bars denote the total number of clusters of the indicated size.

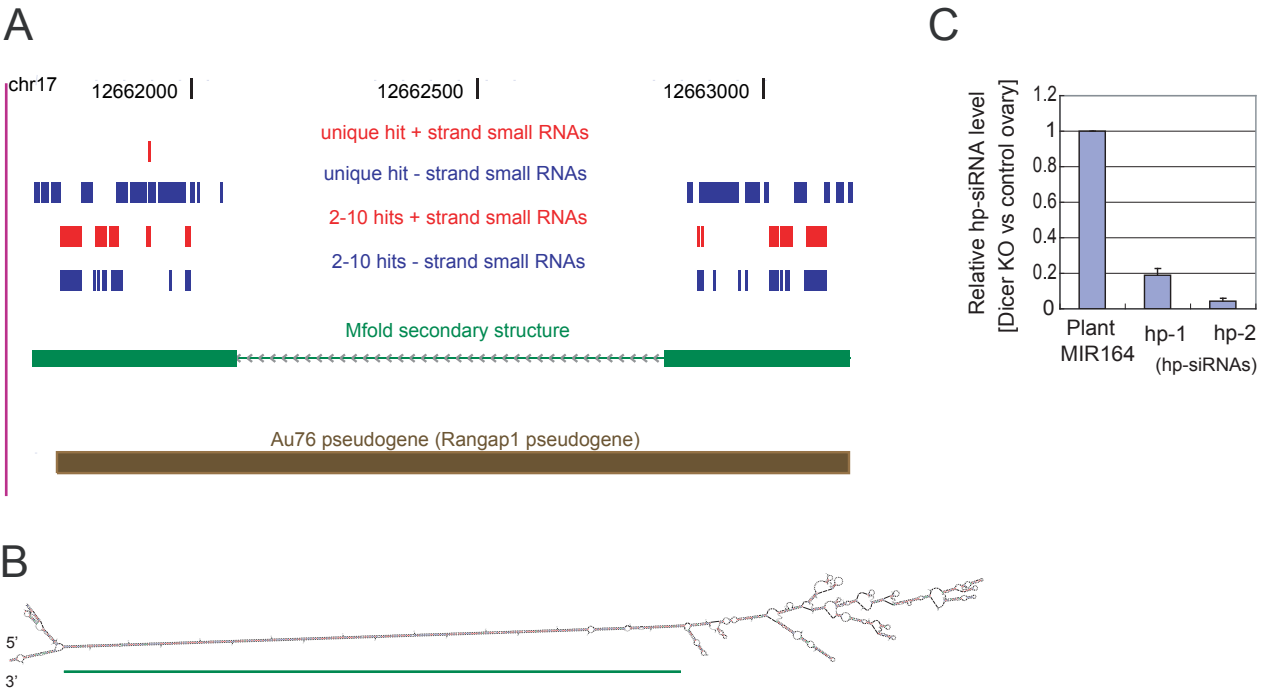
(B) The length distribution of small RNAs that constitute piRNA and other clusters. The small RNAs from the piRNA clusters showed two peaks: the one at 21 nt is due to the fact that some piRNA clusters contain siRNAs (for example, see Figure 19).

Figure 13



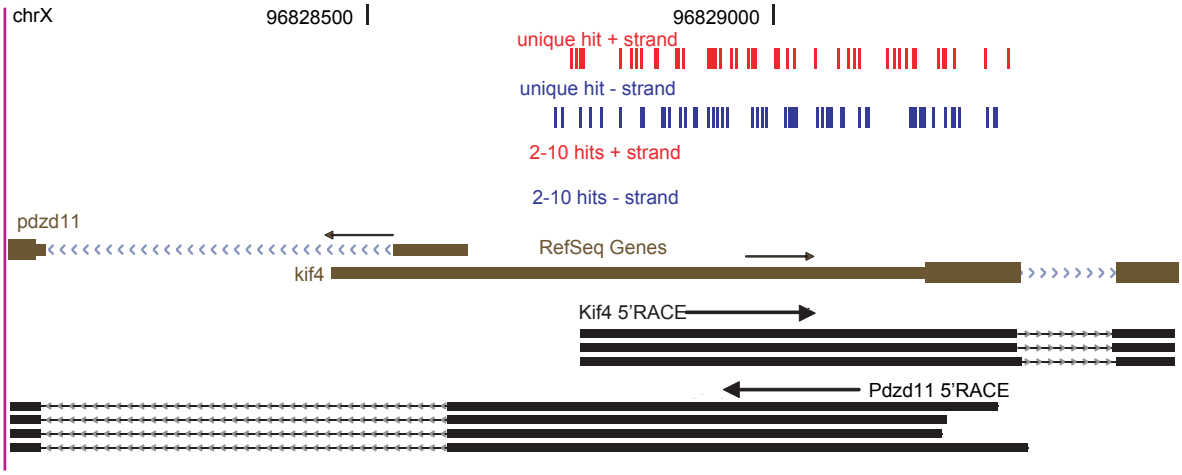
Most of the genomic sequences encompassing 19-23 nt small RNAs from growing oocytes do not form an miRNA-type hairpin structure. The RNAshapes program and randfold program were used to predict whether sequences form an miRNA-type hairpin structure. miRNA sequences registered in miRBase and 19-23 nt sequences randomly retrieved from the mouse genome were also analyzed as positive and negative controls respectively. The number of sequences with a hairpin structure and the total number of analyzed sequences are shown in parentheses.

Figure 14



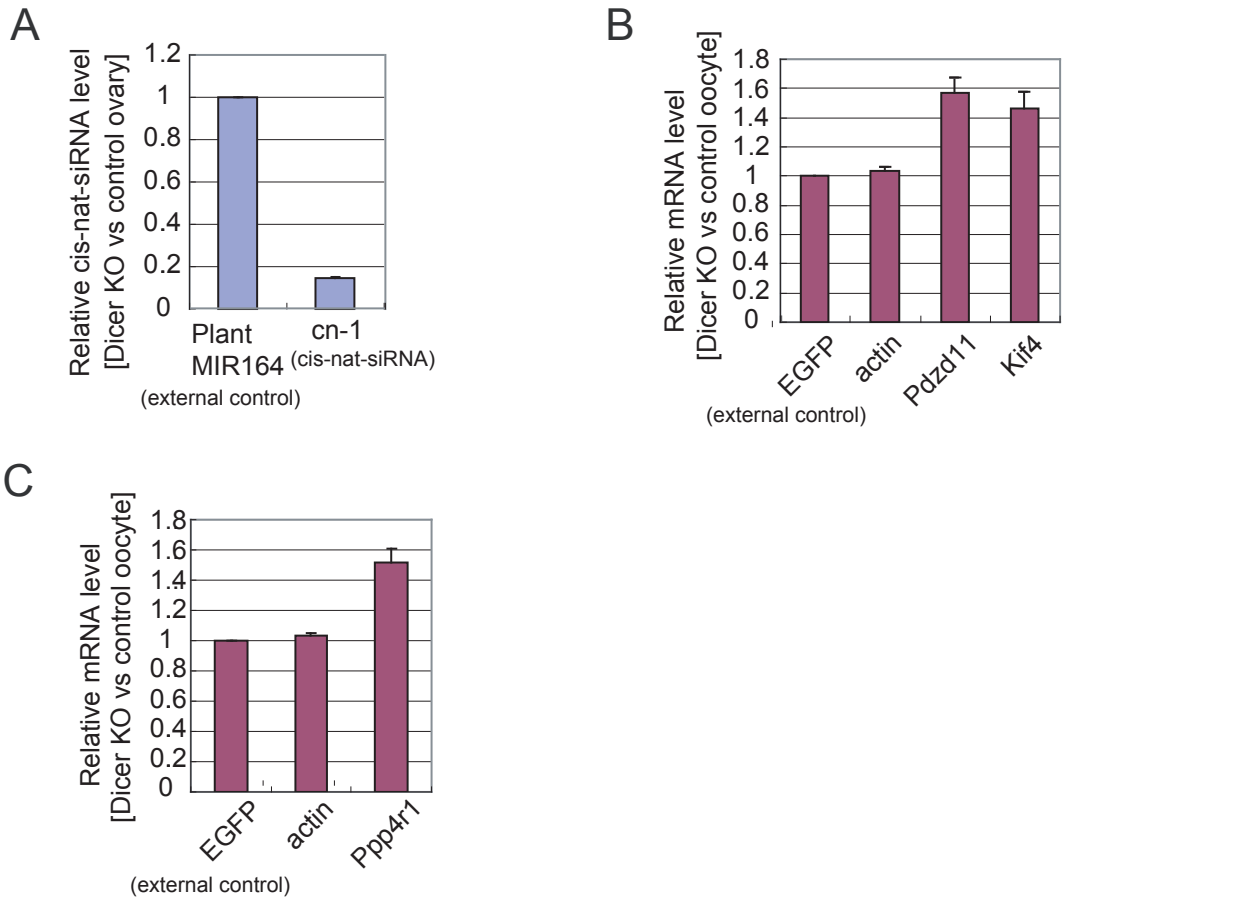
(A) An hp-siRNA cluster at the Au76 locus on chr17. The small RNAs mapped in this region are represented by red (plus-strand) or blue (minus-strand) bars. Small RNAs with unique hit and 2-10 time hits to the genome were indicated in different lines. The thick green bars in Mfold secondary structure represent the portions constituting the stem of the hairpin structure represented in B (green bar in B). Au76, a pseudogene of Rangap1, is indicated in brown.
(B) Secondary structure of the putative transcript derived from Au76. The minus strand of the genomic sequence was folded using Mfold. The most stable structure predicted by Mfold is shown.
(C) Quantitative RT-PCR analysis of two hp-siRNAs derived from this locus in conditional Dicer KO ovaries. The amounts of these RNAs in conditional Dicer KO ovaries relative to those in ovaries that do not express Cre recombinase are shown. To both KO and control samples, the same relative amount of plant MIR164 was added, which served as an external control. Error bars represent S.D. (n=3). Amplified products were sequenced and confirmed.

Figure 15



A cis-nat-siRNA cluster at the Pdzd11/Kif4 locus on chrX. Refseq gene predictions for Pdzd11 and Kif4 (rectangles, exons; arrow head arrays, introns) are shown with the arrows indicating their transcriptional orientations. The transcribed regions determined by 5' RACE are indicated at the bottom (rectangles, exons; arrow head arrays, introns).

Figure 16

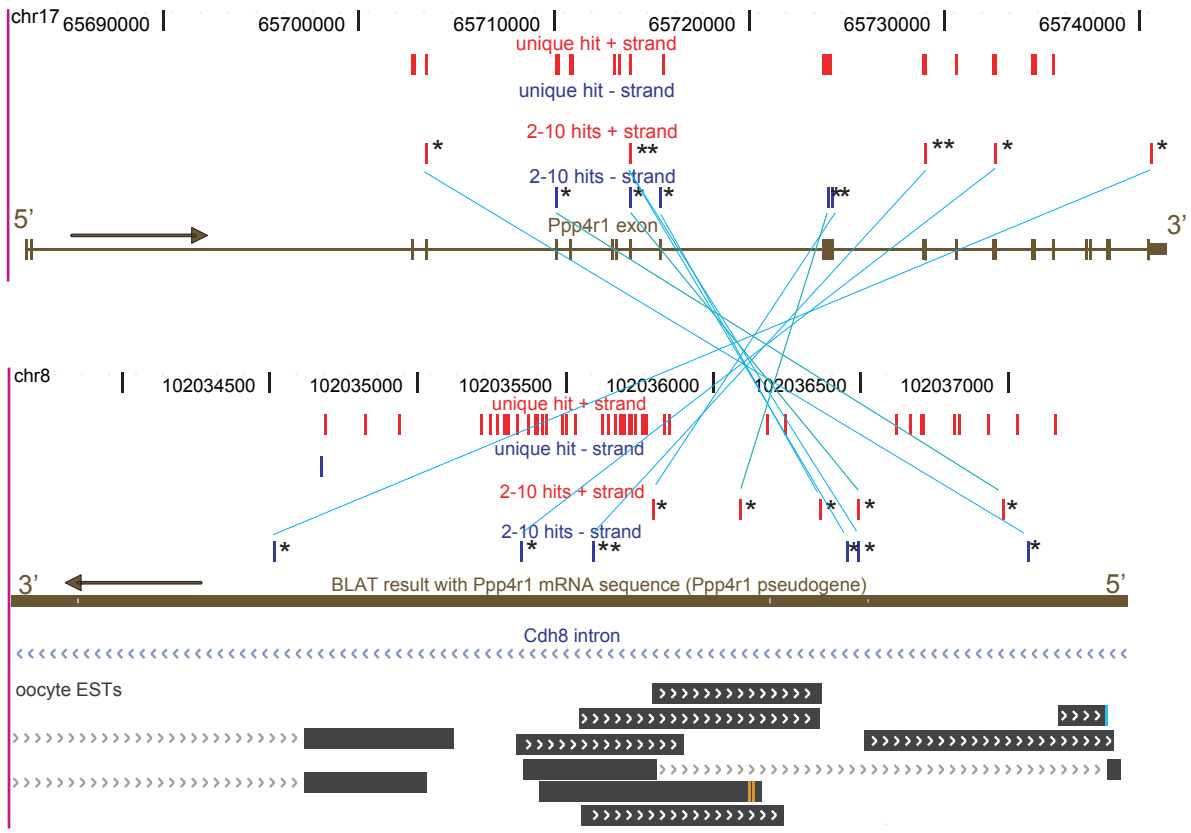


(A) Quantitative RT-PCR analysis of a cis-nat-siRNA derived from Pdzd11/Kif4 locus in conditional Dicer KO ovaries. The amounts of this RNA in conditional Dicer KO ovaries relative to those in ovaries that do not express Cre recombinase are shown. MIR164 served as an external control. Error bars represent S.D. (n=3). Amplified products were sequenced and confirmed.

(B) Quantitative RT-PCR analysis of Pdzd11 and Kif4. Error bars represent S.E. (n=3). The amounts of EGFP (external control), actin (internal control), Pdzd11 and Kif4 mRNAs in conditional Dicer KO oocytes were divided by those in oocytes that do not express Cre recombinase.

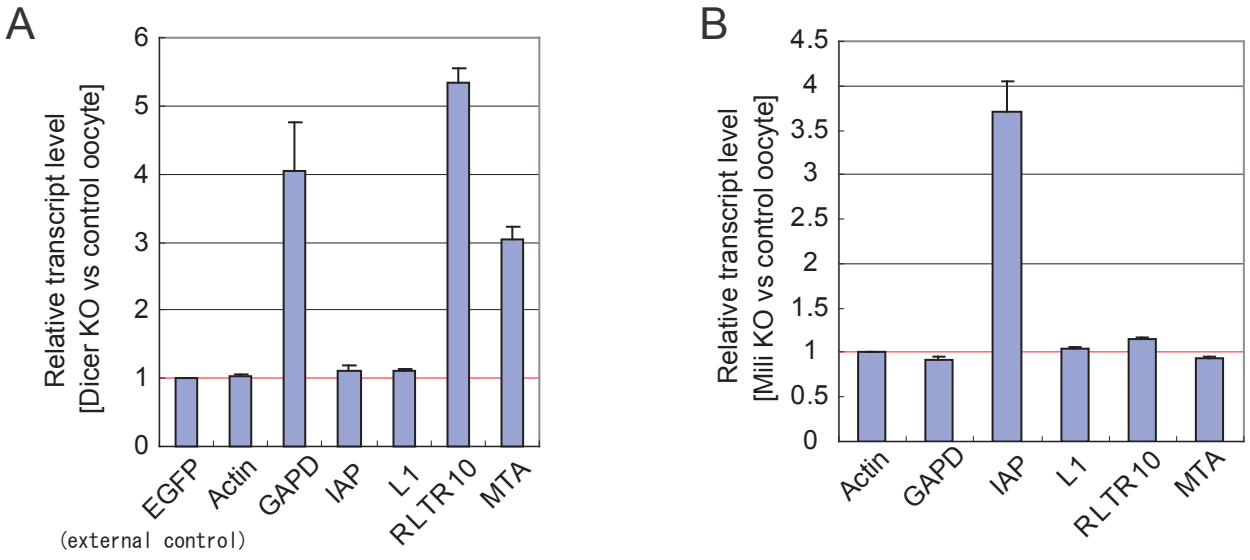
(C) Quantitative RT-PCR analysis of Ppp4r1 mRNA. Error bars represent S.E. (n=3). The amounts of EGFP (external control), actin (internal control) and Ppp4r1 mRNAs in conditional Dicer KO oocytes were divided by those in oocytes that do not express Cre recombinase.

Figure 17



A representative pair of trans-nat-siRNA clusters at the loci of Ppp4r1 on chr17 (top) and its processed psedogene located in a Cdh8 intron on chr8 (bottom). The Ppp4r1 exon/intron structure (top) and the processed pseudogene revealed by BLAT homology search (bottom) are indicated by brown bars. Asterisks indicate the small RNAs mapped to both chr8 and chr17 clusters and lines between the panels connect the locations of the same RNAs. Arrows indicate the 5' to 3' directions of Ppp4r1 and its pseudogene. Black rectangles (bottom) represent oocyte ESTs, all of which are antisense to the Ppp4r1 sequence.

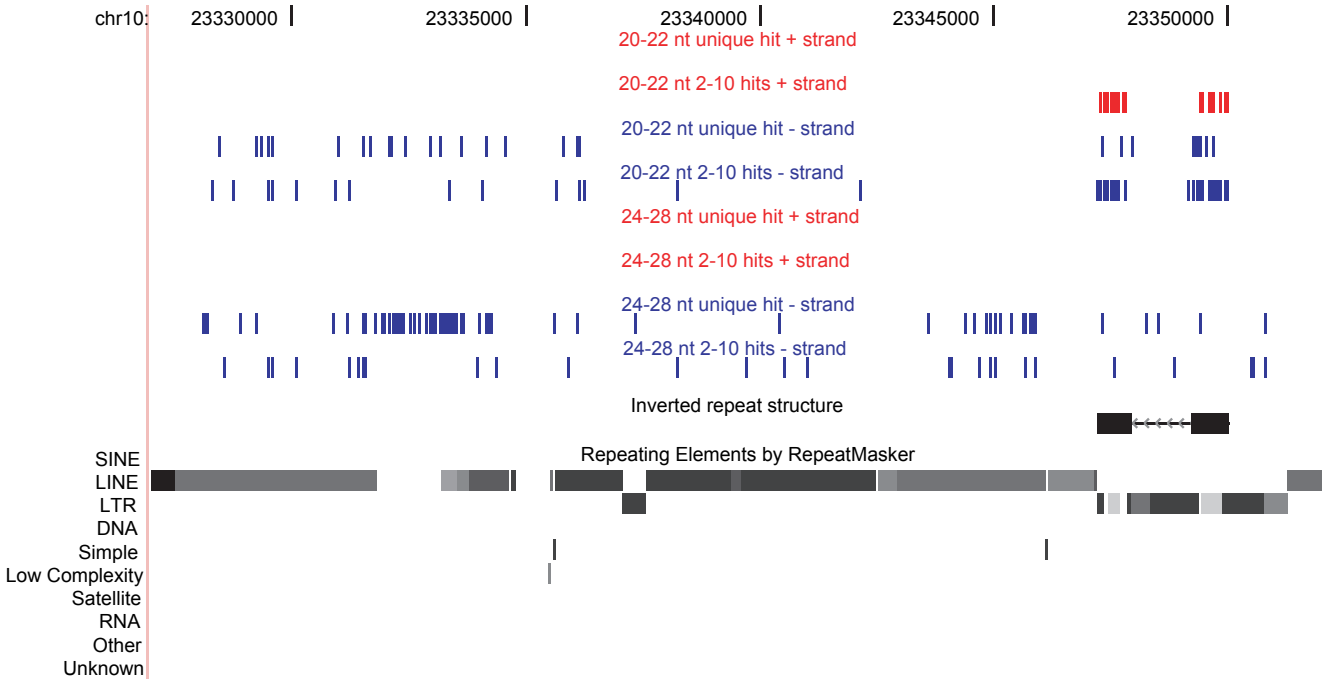
Figure 18



(A) Quantitative RT-PCR analysis of some retrotransposons in Dicer KO oocytes. Error bars represent S.E. (n=4). Oocytes that did not express Cre recombinase were used as controls. As an external control, a certain amount of EGFP mRNA was added per oocyte before isolation of total RNAs. The values were normalized by the relative level of EGFP mRNA. Note that small RNAs corresponding to GAPD were found in the oocyte small RNA library (data not shown).

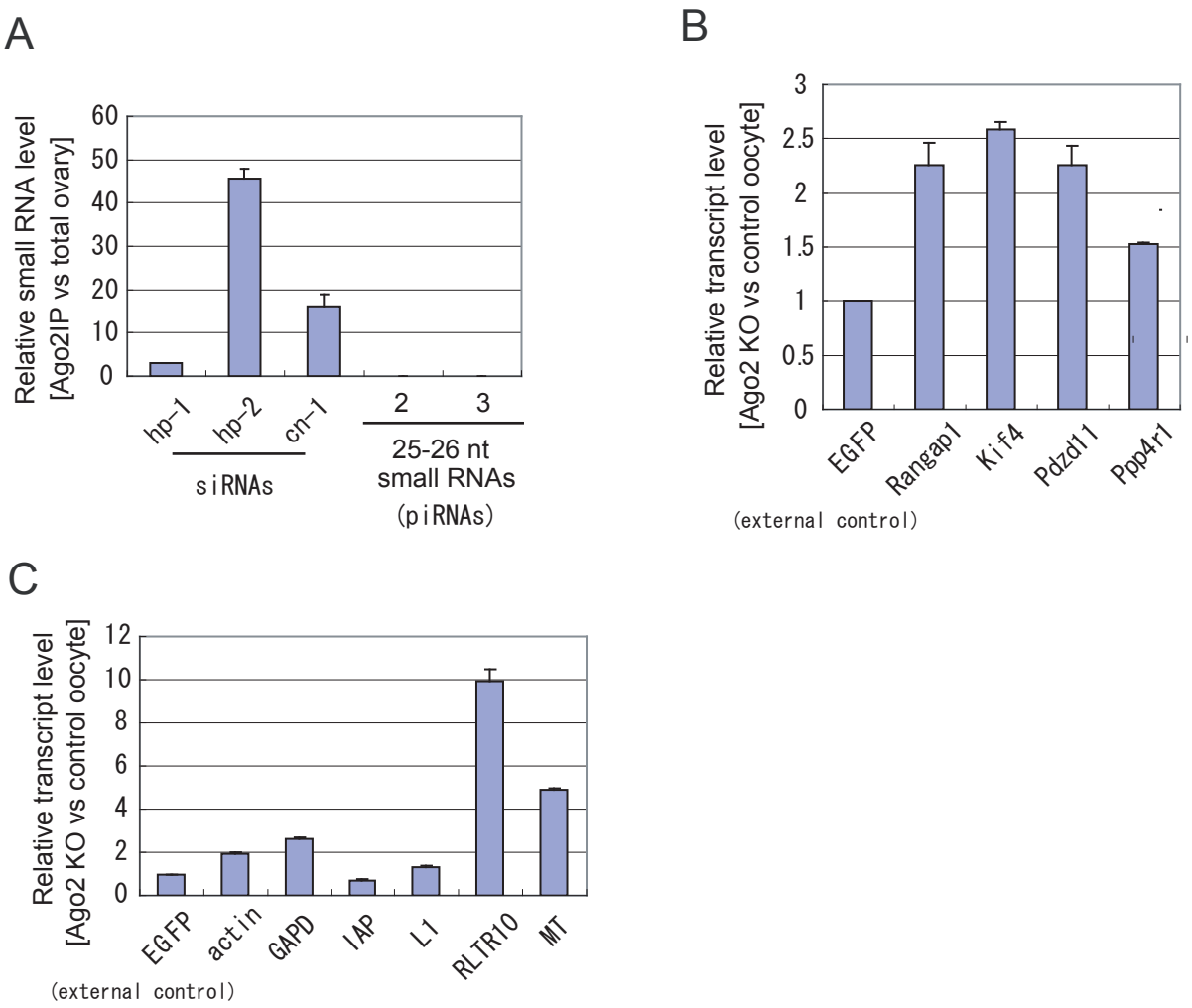
(B) Quantitative RT-PCR analysis of some retrotransposons in Mili KO oocytes. Error bars represent S.E. (n=4). Mili +/- oocytes were used for controls. The values were normalized by the relative level of beta-actin mRNA.

Figure 19



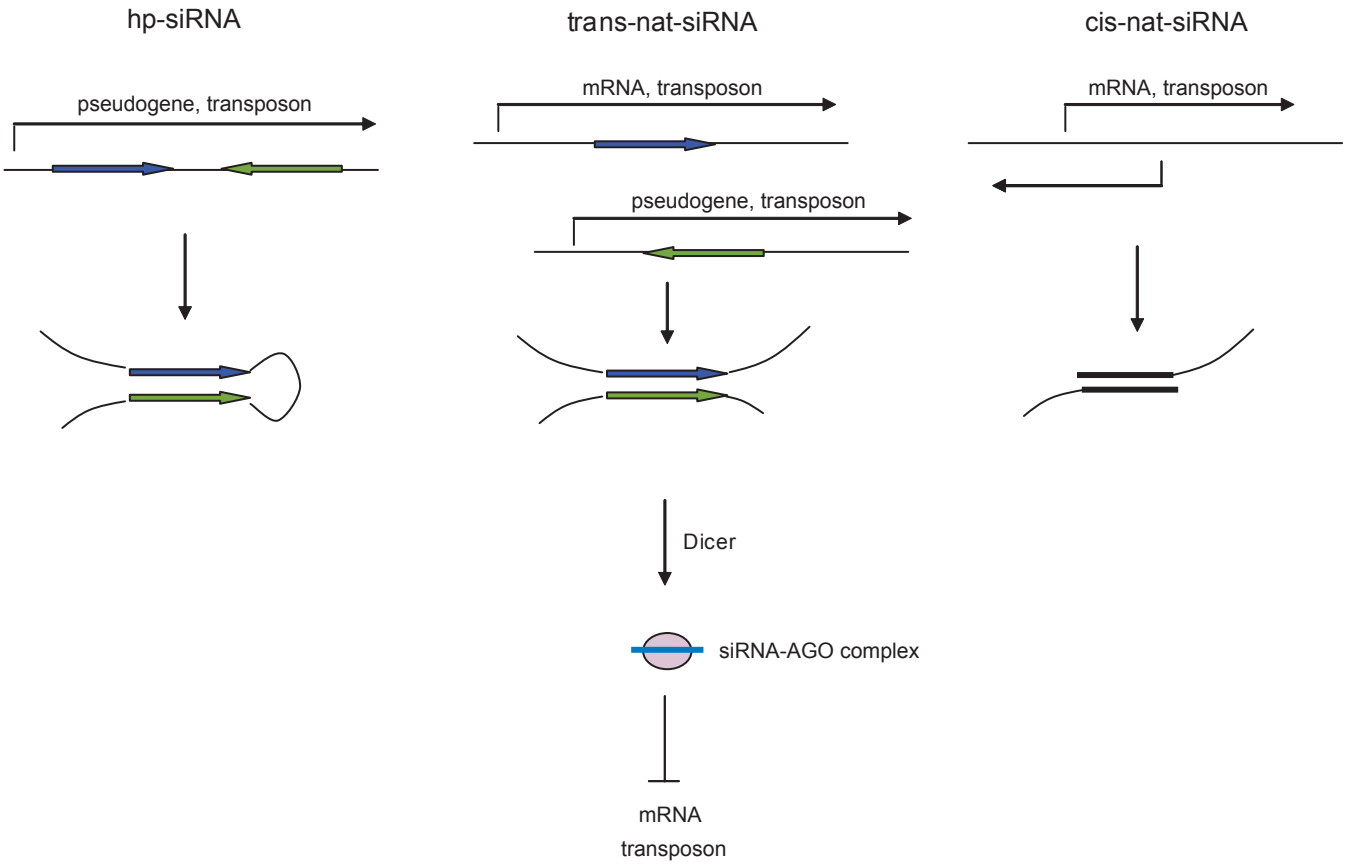
A peculiar cluster in which both siRNAs and piRNAs are mapped. The small RNAs mapped in this region are represented by red (plus-strand) or blue (minus-strand) bars. Small RNAs that hit the genome only once and those that hit the genome 2-10 times are indicated in different rows. The thick black bars represent the inverted repeat structure that constitutes the stem of a hairpin predicted by Mfold. Repeat elements identified by RepeatMasker are also shown.

Figure 20



(A) Quantitative RT-PCR analysis of selected endogenous siRNAs and piRNAs in total and Ago2-IP small RNAs from P8 ovaries. The amount of each RNA species in the Ago2-IP library was divided by that in the total small RNA library. Experiments were repeated three times. Some piRNAs were also tested but were undetectable in the Ago2-IP library. Amplified products were sequenced and confirmed to have the correct sequences. Error bars represent S.D. (n=3)
(B) Quantitative RT-PCR analysis of target mRNAs in Ago2 KO oocytes. Error bars represent S.E. (n=3). The amounts of EGFP (external control), Rangap1, Kif4, Pdzd11 and PPP4r1 mRNAs in conditional Ago2 KO oocytes were divided by those in oocytes that did not express Cre recombinase.
(C) Quantitative RT-PCR analysis of some retrotransposons in Ago2 KO oocytes. Error bars represent S.E. (n=3). The amounts of EGFP (external control), beta-actin and retrotransposons in conditional Ago2 KO oocytes were divided by those in oocytes that did not express Cre recombinase.

Figure 21



Double-stranded RNAs, which are the precursors of siRNAs, are produced in three different ways. Blue and green arrows indicate homologous sequences oriented in opposite directions.

Table 1 Annotaion of fetal testes small RNAs

RNA class	No. of reads (%)
rRNA	51097 (40)
tRNA	8306 (6)
snRNA/snoRNA	2571 (2)
scRNA/srpRNA	696 (0.5)
miRNA	12356 (10)
pachytene piRNA[1]	141 (0.1)
mRNA	9244 (7)
repeat	28809 (23)
unknown[2]	14777 (12)
total	127997

1 This class represents small RNAs that are annotated as piRNA in mouse adult testes (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006b; Lau et al., 2006; Watanabe et al., 2006).

2 This class represent small RNAs that were not classified in any other classes.

Table 2. Annotation of repeat fetal testes small RNAs

RNA class	No. of reads (%)	
LTR/ERVK	10593	(37)
LTR/ERV1	3010	(10)
LTR/MaLR	1790	(6)
LTR/ERVL	623	(2)
SINE	3281	(11)
LINE	8763	(30)
Simple	241	(0.8)
DNA	158	(0.5)
Other	110	(0.4)
Satellite	57	(0.2)
Low-complexity	56	(0.2)
tandem repeat	17	(0.1)
Unknown	108	(0.4)
total	28809	

Table 3. Fetal testes piRNA clusters

A No. of small RNA species
B No. of clone that hit 1-10 times
C No. of clone that are 19-23 nt in length
D No. of clone that are 24-28 nt in length

#	Chr	Genomic position (UCSC mm8)	strand	A	B	C	D
1	chr7	6176657-6262983	-	1397	1767	226	1362
2	chr10	82786798-82915694	+	265	296	46	202
3	chr8	49118636-49140158	+-	135	153	24	108
4	chr5	143076483-143275472	+-	109	119	16	91
5	chr7	18911356-19102745	+-	101	106	16	79
6	chr8	20389854-20404223	+-	95	118	15	85
7	chr3	94667420-94770505	+-	93	106	15	81
8	chr19	5796108-5860502	+-	80	86	18	43
9	chr17	5709359-5737492	+-	77	87	10	67
10	chr7	75637454-75654236	-	75	85	14	63
11	chr2	22439046-22441767	-	62	88	47	15
12	chr11	3049805-3188877	+-	55	56	10	42
13	chr14	64578423-64585091	+-	54	54	4	44
14	chr2	168714644-168930752	+-	47	54	7	37
15	chr1	157824129-157844843	+-	46	47	2	42
16	chr11	120434478-120584670	+-	45	50	8	36
17	chr5	142623457-142838467	+-	40	44	2	38
18	chr3	96390430-96586420	+-	40	47	10	26
19	chr10	60776230-60830430	-	38	44	9	31
20	chr8	120091859-120099134	+	38	42	7	29
21	chr8	73519327-73627094	+-	38	43	7	33
22	chr10	80458907-80592686	+-	34	38	6	31
23	chr4	152581104-152602776	+-	34	36	1	31
24	chr1	192832861-192844758	+-	33	35	2	30
25	chr11	97803931-97888803	+-	33	41	6	31
26	chr1	102603410-102616089	+-	32	35	2	27
27	chr5	148540537-148638826	+-	31	32	8	22
28	chr7	18027006-18155490	+-	30	37	2	33
29	chr1	133845735-133859830	+	30	32	5	26
30	chr7	44846957-44998562	+-	30	34	9	18
31	chr4	134638730-134735478	+-	29	30	6	21
32	chr2	180178750-180253946	+-	28	33	6	27
33	chr7	28246375-28416397	+-	27	30	1	27
34	chr2	167166211-167209527	+-	27	29	4	22
35	chr11	60593864-60605271	+	27	31	1	25
36	chr5	139536154-139703779	-	26	29	5	21
37	chr4	131531301-131586347	+-	26	28	4	19
38	chr15	36591473-36599856	+-	26	26	1	21
39	chr11	77889205-77898353	+	25	188	22	159
40	chr5	148950998-149058888	+-	25	27	5	19
41	chr9	114253782-114416406	+-	25	26	3	21
42	chr2	25954071-26163250	+-	25	26	7	17
43	chr15	79652500-79778594	+-	25	27	5	20
44	chr1	162841613-162874790	+-	24	47	9	27
45	chr7	44462545-44481545	+-	24	27	6	18
46	chr10	67065271-67069909	+-	24	26	5	19
47	chr16	22693980-22700429	+	23	25	2	21

48	chr7	116615667-116616985	-	23	26	3	19
49	chr15	97190594-97219942	+-	23	24	2	22
50	chr11	94475779-94564952	+-	23	25	4	17
51	chr10	76636232-76681938	+-	23	24	3	19
52	chr8	87063680-87102535	+-	22	28	10	13
53	chr13	23754647-23771601	+-	22	26	5	15
54	chr2	154694824-154700066	+	22	23	1	21
55	chr5	135217392-135225489	+-	22	25	5	18
56	chr11	117336831-117343688	+-	22	22	5	15
57	chr17	26003330-26111876	+-	22	26	3	18
58	chr4	94645557-94756169	+-	22	24	4	18
59	chr5	108140901-108147777	+	22	35	11	18
60	chr5	78140623-78161189	+-	22	23	3	19
61	chr6	136428453-136435427	+-	21	23	2	19
62	chr2	129885065-129969486	+	21	23	3	12
63	chr8	8987206-8992042	+-	21	21	3	16
64	chr12	20473570-20488674	+-	20	21	0	18
65	chr11	96504823-96512730	+-	20	26	3	21
66	chr16	90408133-90427068	+-	20	20	1	18
67	chr1	183978975-184067146	+-	20	20	1	19
68	chr9	61711461-61720993	-	20	24	4	17
69	chr10	79971144-80081179	+-	19	20	2	16
70	chr5	23212479-23227779	+-	19	24	2	18
71	chr6	125530174-125561621	+-	19	20	0	19
72	chr7	35462053-35474355	+-	19	21	1	18
73	chr8	124198563-124214969	+-	19	20	0	19
74	chr11	74647566-74662868	+-	19	21	3	16
75	chr10	120661739-120666816	+-	18	18	2	13
76	chr9	108364909-108426915	+-	18	26	7	16
77	chr10	94804633-94810454	+-	18	21	3	16
78	chr8	112950331-112958963	+-	18	18	3	14
79	chr8	87453211-87460072	+-	18	19	1	15
80	chr2	155890665-155908323	+	18	19	2	17
81	chr19	5388966-5525803	+-	18	19	3	14
82	chr2	172661085-172676561	+-	17	18	1	16
83	chr17	47630845-47642705	+-	17	19	3	14
84	chr2	24986240-25010275	+-	17	17	1	15
85	chr5	134387674-134399394	+-	17	17	1	13
86	chr19	10049784-10052038	+	17	17	4	12
87	chr4	146189734-146196210	+-	17	18	2	13
88	chr10	127487094-127493065	+	17	18	3	14
89	chr3	129447923-129464587	+-	17	18	2	14
90	chr6	85410621-85450131	+-	17	19	5	14
91	chr13	23579298-23593078	+-	17	18	6	9
92	chr1	88309882-88357981	+-	17	20	4	11
93	chr10	79195881-79267297	+-	17	17	1	11
94	chr11	105843629-105853614	+-	17	19	4	15
95	chr10	118493339-118555016	+-	17	17	1	14
96	chr3	95268237-95274305	-	16	18	3	15
97	chr3	88216358-88234325	+-	16	16	4	10
98	chr17	56304298-56321739	+-	16	16	0	14
99	chr15	82182287-82194943	+	16	16	4	9
100	chr12	92276189-92285918	+-	15	16	2	13
101	chr11	69193812-69218374	+-	15	15	3	9
102	chr10	126395955-126401432	+	15	16	1	12
103	chr2	164113560-164222088	+-	15	15	1	12
104	chr13	51607787-51650878	+-	15	17	5	10
105	chr7	15455255-15467539	+-	15	17	1	15

106	chr2	122066896-122069503	+-	15	21	2	3
107	chr15	88689088-88696933	+-	15	16	0	15
108	chr9	63972889-63976494	+	15	21	3	9
109	chr5	140847555-140901068	+-	15	15	1	14
110	chr7	11822069-11930884	+-	15	15	3	8
111	chr11	95516621-95588989	+-	15	15	0	13
112	chr7	63759737-63818458	+-	15	18	1	16
113	chr15	103020142-103028880	-	15	15	1	12
114	chr2	164498865-164521893	+-	14	14	0	14
115	chr15	93010686-93017348	+-	14	15	3	11
116	chr16	4528763-4566452	+-	14	14	4	10
117	chr2	30478122-30483412	+-	14	15	1	13
118	chr17	23535754-23556560	+-	14	14	2	8
119	chr14	121225743-121240100	+-	14	14	3	9
120	chr8	125988689-126064171	+-	14	14	6	6
121	chr1	172432565-172439668	+-	14	17	2	12
122	chr6	28872939-28881942	-	14	14	4	10
123	chr7	127561383-127673023	+-	14	15	2	10
124	chr7	30612152-30622519	+-	14	19	3	15
125	chr17	27044495-27064645	+-	14	17	1	11
126	chr8	11005997-11013315	+-	14	17	3	12
127	chr2	94062476-94077944	+-	13	13	1	11
128	chr13	23277878-23317496	+-	13	14	4	8
129	chr3	103367974-103376437	+	13	16	2	10
130	chr11	6312676-6356114	+-	13	15	0	11
131	chr2	32212004-32218023	-	13	14	2	12
132	chr19	38089768-38096694	+-	13	13	2	11
133	chr10	51216500-51222058	+-	13	13	0	12
134	chr17	27280714-27290501	+-	13	15	2	10
135	chr2	27877104-27896489	+-	13	13	2	9
136	chr19	55062066-55069670	+-	13	14	0	11
137	chr19	21779660-21791483	+-	13	14	1	13
138	chr1	135875553-135915662	-	13	14	4	10
139	chr8	122646483-122687422	+-	13	14	1	13
140	chr3	58503457-58505239	+	13	17	0	14
141	chr8	123553900-123560110	+	13	15	0	13
142	chr2	164627597-164634319	+-	13	14	0	11
143	chr11	53598409-53612242	+-	13	17	2	14
144	chr4	56962852-56974978	-	13	13	0	11
145	chr7	45319625-45327802	+-	12	30	0	9
146	chr5	33554290-33559243	+-	12	12	3	9
147	chr17	33513604-33563023	+-	12	13	4	7
148	chr1	158479626-158484926	+	12	12	2	9
149	chr1	167702226-167708202	-	12	13	2	11
150	chr11	21068143-21092452	+-	12	14	1	11
151	chr4	151064732-151079283	+-	12	15	0	13
152	chr8	123414196-123430323	+-	12	13	2	10
153	chr17	35326363-35339943	+-	12	13	0	13
154	chr9	107477847-107491060	+-	12	16	7	8
155	chr1	87712508-87724465	+-	12	14	5	8
156	chr8	124453998-124470995	+-	12	13	1	10
157	chr3	153900157-153906161	+-	12	14	3	11
158	chr13	12313296-12327338	+-	12	14	1	13
159	chr7	33736922-33743756	+-	12	13	2	11
160	chr14	119572115-119576285	+-	12	13	3	8
161	chr17	25149267-25172155	+-	12	13	2	10
162	chr4	153751853-153805922	+-	12	12	0	11
163	chr5	130192063-130206030	+-	12	14	3	10

164	chr7	34179239-34187493	+	12	14	2	11
165	chr7	120295944-120399154	+-	12	13	3	10
166	chr9	62546419-62549081	+-	12	13	1	11
167	chr7	99352859-99360139	+-	12	13	4	7
168	chr7	100712525-100806158	+-	12	12	2	9
169	chr2	71702967-71709549	+-	12	14	2	12
170	chr4	132334431-132369105	+-	12	13	1	11
171	chr7	4610553-4670586	+-	12	13	1	9
172	chr2	153459174-153470334	+-	12	14	0	12
173	chr2	152098345-152105797	+-	12	15	1	13
174	chr7	23016575-23024655	+-	12	12	3	9
175	chr7	127025445-127035443	+-	12	13	1	8
176	chr13	98159714-98167649	+-	12	13	2	11
177	chr14	53556950-53572180	+-	12	14	3	10
178	chr13	100100268-100141415	+	12	13	0	12
179	chr9	106738890-106752969	+-	11	11	1	9
180	chr7	17535510-17548032	+-	11	11	3	7
181	chr19	24351765-24357665	+-	11	11	1	9
182	chr2	172502967-172514061	+-	11	12	1	10
183	chr5	119553296-119563795	+-	11	11	3	7
184	chrX	110082665-110089390	+-	11	11	0	10
185	chr18	80384388-80388790	+	11	12	1	9
186	chr5	113825122-113851251	+-	11	14	0	9
187	chr8	87734834-87746074	+-	11	12	3	9
188	chr17	75460075-75470989	+-	11	11	0	10
189	chr17	18260660-18272824	+-	11	13	1	9
190	chr7	27893222-27904432	+	11	11	2	9
191	chr8	74425123-74437163	+-	11	13	0	13
192	chr9	102968218-103032998	+-	11	11	0	8
193	chr9	12092458-12101484	+-	11	11	2	9
194	chr7	24817940-24821586	+-	11	12	2	9
195	chr5	144889889-144983323	+-	11	13	2	10
196	chr2	4866746-4877023	+-	11	12	2	5
197	chr4	33438530-33452508	+-	11	16	0	16
198	chr13	95445042-95449618	+	11	11	0	7
199	chr17	34559052-34567225	-	11	11	2	7
200	chr10	69720235-69724177	+-	11	11	0	10
201	chr11	48634511-48647686	+	11	11	3	7
202	chr12	31951954-31958775	+-	11	12	2	7
203	chr11	60917792-60977575	-	11	12	2	9
204	chr11	69478766-69488535	+-	11	12	5	6
205	chr11	5905681-5910806	+-	11	11	1	7

Table 4. Annotation of growing oocyte small RNAs

RNA class	No. of reads (%)					
	oocyte total	oocyte 20-22 nt	oocyte 25-27 nt	Mili IP	Mili IP 20-22 nt	Mili IP 25-27 nt
rRNA	5047 (5)	1593 (3)	558 (2)	143 (3)	34 (7)	15 (0.8)
tRNA	3767 (4)	978 (2)	440 (2)	154 (4)	17 (3)	39 (2)
sn_snoRNA	395 (0.4)	128 (0.3)	59 (0.3)	21 (0.5)	3 (0.6)	2 (0.1)
sc_srpRNA	119 (0.1)	43 (0.1)	26 (0.1)	13 (0.3)	2 (0.4)	3 (0.2)
miRNA	3587 (3)	2815 (6)	1 (0)	34 (0.8)	25 (5)	0 (0)
pachytene piRNA[1]	104 (0.1)	50 (0.1)	20 (0.1)	8 (0.2)	0 (0)	6 (0.3)
mRNA	9420 (9)	4445 (11)	1872 (8)	609 (15)	80 (16)	284 (15)
repeat	74989 (72)	35488 (96)	18035 (77)	2362 (57)	272 (56)	1128 (80)
unknown[2]	6567 (6)	1419 (3)	2482 (11)	785 (19)	55 (11)	403 (21)
total	103995 (100)	46959 (100)	23493 (100)	4129 (100)	488 (100)	1880 (100)

1 This class represents small RNAs that are annotated as piRNA in mouse adult testis.

2 This class represents small RNAs that are not classified in any other classes.

Table 5. Small RNA clusters in mouse growing oocytes

#	Chr	Genomic position (UCSC mm8)	strand	No. of small RNAs				Class of cluster
				kind	clone			
					total	19-23 nt	24-28 nt	
1	chr10	82789601-82918207	+	3654	4816	1548	2993	piRNA
2	chr10	117640185-117677987	+-	1970	2630	1101	1376	piRNA
3	chr12	68565038-68604191	+	1327	2220	698	1387	piRNA
4	chr10	23296600-23350790	+-	637	983	715	211	hp-siRNA piRNA
5	chr17	12661727-12663173	+-	485	979	928	4	hp-siRNA
6	chr10	43898893-43918070	+	201	289	84	194	piRNA
7	chr4	56961570-56982731	-	196	225	93	113	piRNA
8	chr9	120636483-120636924	-	167	7306	7215	7	(hp-siRNA)[1]
9	chr17	5728780-5735259	+	151	194	76	103	piRNA
10	chr11	91737385-91742415	+-	144	194	82	95	
11	chr7	75642809-75654236	-	143	167	36	120	piRNA
12	chr9	72892591-72910671	+	137	165	49	100	piRNA
13	chr5	137948929-137951678	+-	126	136	122	4	
14	chrX	96828731-96829310	+-	117	135	126	0	cis-nat-siRNA
15	chr10	26385021-26438068	+	112	146	43	91	piRNA
16	chr2	4936975-4938574	+-	109	152	145	0	hp-siRNA
17	chr7	6177907-6256415	-	94	108	29	72	piRNA
19	chr9	51583810-51592553	+-	86	239	227	5	piRNA
18	chr17	26648484-26664411	+-	86	148	142	2	
20	chr8	72590252-72654897	-	85	98	28	66	piRNA
21	chr13	50650072-50654460	+-	74	103	54	42	piRNA
22	chr12	17091603-17238365	+-	70	87	28	53	piRNA
23	chr8	102034513-102037171	+-	69	77	73	1	trans-nat-siRNA (27)[2]
24	chr3	96347880-96358820	+-	68	81	65	15	
25	chr9	114357265-114364731	+-	68	76	66	4	cis-nat-siRNA piRNA
26	chr9	69155319-69165534	-	67	82	44	35	piRNA
27	chr17	65702757-65789029	+-	63	72	65	4	trans-nat-siRNA (23) piRNA
28	chr14	104076918-104093492	+-	61	69	58	5	trans-nat-siRNA (36) piRNA
29	chr10	11373862-11392299	+-	59	73	36	34	
30	chr5	114103753-114204341	+-	58	60	54	4	cis-nat-siRNA
31	chr9	71397396-71412177	+	56	61	37	22	
32	chr12	81255978-81259024	+-	55	73	37	30	
33	chrX	63037425-63071159	-	54	169	161	1	
34	chr3	51314104-51329858	+	51	60	11	46	piRNA
35	chr1	133805031-133859700	+	51	55	15	39	piRNA
36	chr14	73933479-73963644	+-	47	51	46	1	trans-nat-siRNA (28)
37	chr8	120091309-120100997	+	46	52	11	39	piRNA
38	chr2	152105312-152106270	+-	45	53	52	0	cis-nat-siRNA
39	chr6	94383256-94384938	+-	45	46	42	1	
40	chr18	6545990-6551168	+-	43	59	32	25	
41	chr5	134841245-134845867	+	43	49	7	39	piRNA
42	chr8	26918337-26920176	+-	42	51	46	1	
43	chr6	12062918-12066410	+	41	59	24	35	
44	chr16	55973878-55985107	+-	41	43	21	21	
45	chr9	107809098-107905674	+-	38	49	38	9	trans-nat-siRNA (70) piRNA
46	chr1	21336888-21371325	+-	38	44	39	3	
47	chr7	142706327-142740466	-	38	43	11	30	piRNA
48	chr18	34545947-34546947	+-	36	45	45	0	
49	chr11	40538801-40551528	-	36	45	27	18	
50	chr12	81366581-81378202	-	36	40	19	20	piRNA

51	chr7	46835269-46843697	-	35	40	37	2	
52	chr1	158478810-158485558	+	35	38	10	26	piRNA
53	chr5	30418401-30436754	+-	35	37	34	0	trans-nat-siRNA (98)
54	chr8	13100313-13119892	+	35	36	10	22	
55	chr13	7235082-7235113	+-	33	53	22	25	
56	chr1	87414685-87432169	+-	33	36	32	2	
57	chr1	134811729-134819209	-	33	36	11	25	piRNA
58	chr2	4750393-4755465	+-	33	33	32	1	trans-nat-siRNA (113)
59	chr6	92115112-92137605	+	32	34	10	18	piRNA
60	chr7	4721830-4731169	+-	32	32	26	4	
61	chr18	13101545-13114997	+	29	35	8	22	
62	chr6	5190821-5209596	-	29	33	12	21	
63	chr11	20724935-20730342	+-	29	31	28	1	trans-nat-siRNA (76)
64	chr7	142390136-142391477	+-	28	54	23	20	
65	chr16	15365996-15374973	+-	28	33	23	9	
66	chr12	98618839-98620440	+-	28	30	21	8	
67	chr11	3807801-3903952	+-	28	30	24	4	
68	chrX	119194985-119199429	+-	27	42	26	13	
69	chr12	88704274-88709701	+-	27	40	11	28	
70	chr2	6050173-6051282	+-	27	32	28	0	trans-nat-siRNA (45)
71	chr12	100365867-100381179	+-	27	27	21	5	
72	chr16	94460947-94468144	+-	27	27	24	3	
73	chr6	120852611-120941581	+	26	32	8	21	
74	chr13	47019741-47021917	+-	26	30	29	1	
75	chr10	127092173-127103365	+	26	29	6	22	
76	chr7	67463293-67469209	+-	26	29	25	2	trans-nat-siRNA (63)
77	chr6	83328817-83341495	-	26	29	7	20	
78	chr14	23274157-23289711	-	26	28	7	19	piRNA
79	chr14	48088967-48178294	+-	26	28	24	2	cis-nat-siRNA
80	chr14	99971053-99979397	+-	25	44	40	2	
81	chr17	4355914-4359443	+-	25	29	13	15	
82	chr6	95209508-95276509	+-	25	27	23	3	cis-nat-siRNA piRNA
83	chr2	61664560-61669772	+-	24	33	9	23	
84	chr8	126304234-126307701	+-	24	28	28	0	cis-nat-siRNA
85	chr6	136363113-136365436	+-	24	27	24	2	
86	chr17	20683993-20699062	+-	24	26	10	14	
87	chr15	64993143-65005963	+-	24	26	5	19	
88	chr3	95268019-95291969	-	24	24	9	14	piRNA
89	chr1	178671614-178672755	+-	23	24	23	0	cis-nat-siRNA
90	chr13	10738676-10740678	-	23	24	8	11	
91	chr15	95901456-95903957	-	22	24	16	7	
92	chr6	113675860-113694543	+	22	23	5	18	piRNA
93	chr2	167165313-167176760	+-	22	23	2	19	piRNA
94	chr2	5464350-5465948	-	21	25	22	2	
95	chr5	129335255-129342749	+-	21	23	21	1	
96	chr15	72923976-72933199	-	21	22	10	10	piRNA
97	chr12	31659934-31661154	+-	21	21	21	0	cis-nat-siRNA piRNA
98	chr4	116658019-116664346	+-	21	21	21	0	trans-nat-siRNA (53)
99	chr7	89817431-89817586	+	20	164	148	2	
100	chr19	44616735-44632652	+	20	23	7	12	piRNA
101	chr3	100597893-100601590	-	20	21	7	12	piRNA
102	chr18	61318900-61324893	-	20	21	5	16	piRNA
103	chr9	67451766-67454167	+	19	156	139	3	
104	chr1	120505455-120512436	+	19	22	5	17	piRNA
105	chr4	135812545-135820680	+	19	21	9	12	piRNA
106	chr16	11574160-11597424	+	19	21	10	8	piRNA
107	chr1	34758131-34764196	-	19	21	8	12	piRNA

108	chr5	45509068-45539400	+-	19	20	9	11	
109	chr13	51156805-51163152	+	19	20	5	14	
110	chr17	63160920-63172370	-	19	20	7	11	
111	chr17	13278068-13292672	-	19	20	7	13	piRNA
112	chr6	40331826-40339075	-	19	19	4	14	piRNA
113	chr7	80168083-80185766	+-	19	19	18	0	trans-nat-siRNA (58)
114	chr10	57540697-57545694	+-	18	25	18	6	
115	chr12	99700460-99731966	-	18	22	8	14	piRNA
116	chr3	9413355-9420647	-	18	20	7	12	piRNA
117	chr7	4962602-4964465	+-	18	19	16	0	
118	chr4	142703898-142704269	+-	18	19	18	0	
119	chr16	78133719-78138023	-	18	19	17	1	
120	chr3	103149689-103158115	+-	18	19	17	0	cis-nat-siRNA
121	chr9	114933907-114946738	+	18	18	6	11	
122	chr4	131168448-131195984	+-	18	18	17	0	cis-nat-siRNA
123	chr4	149061674-149061849	+	17	159	156	2	
124	chr18	74783814-74783981	+-	17	21	19	0	
125	chr3	50464302-50466857	+-	17	20	18	1	
126	chr9	3261498-3277798	-	17	20	6	13	piRNA
127	chr1	44994426-44999804	-	17	20	10	10	
128	chr10	116997414-116999574	+-	17	20	19	0	cis-nat-siRNA
129	chr11	22468219-22495726	-	17	19	7	11	
130	chr6	8545951-8547015	+	17	18	15	1	trans-nat-siRNA (210)
131	chr3	138868338-138876679	+	17	18	4	14	
132	chr9	14287825-14300982	-	17	17	7	9	
133	chr10	119869799-119882982	+-	17	17	15	0	
134	chr2	121235120-121235783	+	16	54	52	0	
135	chr6	38619537-38628567	-	16	20	8	10	piRNA
136	chr11	94555502-94566332	+-	16	18	10	6	
137	chr6	81297736-81305583	+	16	18	9	7	
138	chr16	16824580-16840264	+	16	17	3	14	piRNA
139	chr8	54248043-54253836	-	16	17	6	9	
140	chr14	73259745-73270682	-	16	17	3	12	
141	chr9	20658329-20677052	-	16	16	7	3	
142	chr11	45495223-45511026	-	16	16	5	9	piRNA
143	chr4	152484173-152486536	+-	16	16	15	0	hp-siRNA
144	chr18	67526739-67577405	+	15	23	12	11	
145	chr7	81780492-81780533	+-	15	19	9	4	
146	chr5	115559558-115561624	-	15	17	16	1	
147	chr8	13461792-13477199	+	15	16	4	12	piRNA
148	chr1	95406795-95423206	+	15	16	5	9	piRNA
149	chr11	103695150-103746901	-	15	16	1	13	
150	chr5	148619879-148633587	+-	15	15	9	3	piRNA
151	chr8	26616361-26631112	+-	14	20	4	11	
152	chr1	13159784-13175383	+-	14	17	15	2	
153	chr3	96577455-96646117	+-	14	16	7	2	
154	chr19	42703016-42704723	+	14	16	5	10	
155	chr1	185510729-185517260	+	14	16	6	10	
156	chr10	60823950-60829144	-	14	15	3	10	
157	chr8	129295474-129304740	+	14	14	5	8	
158	chrX	88694245-88694342	+-	13	23	23	0	
159	chr18	68375963-68389552	+-	13	18	8	10	piRNA
160	chr11	105870896-105882464	+	13	18	5	12	piRNA
161	chr3	6245265-6248969	-	13	16	1	15	
162	chr13	43206331-43212159	-	13	16	10	6	piRNA
163	chr3	132775264-132780217	+-	13	15	12	3	
164	chr5	114624821-114635985	+-	13	15	3	10	

165	chr8	107238704-107244407	-	13	15	4	9	piRNA
166	chr8	112618024-112624628	-	13	15	7	8	piRNA
167	chr2	161038825-161061882	+	13	15	3	10	piRNA
168	chr18	42771363-42775698	+-	13	14	12	1	cis-nat-siRNA
169	chr2	129491799-129522359	+-	13	14	6	5	
170	chr17	45032004-45035472	-	13	14	12	1	
171	chr3	26802295-26807902	-	13	14	6	7	
172	chr6	39801276-39803027	+-	13	13	9	4	
173	chr1	191154111-191172227	-	13	13	2	11	
174	chr17	28315845-28321841	-	13	13	4	8	
175	chrX	4293190-4303399	-	13	13	2	9	
176	chr5	104082608-104095309	+	13	13	1	12	piRNA
177	chr18	16208720-16215946	-	13	13	6	6	piRNA
178	chr3	97186697-97195680	+	12	16	3	9	
179	chr10	48222016-48224017	+-	12	15	10	4	piRNA
180	chr17	9158916-9161441	+	12	14	12	0	
181	chr7	23903778-23927741	+-	12	13	9	3	
182	chr11	5428058-5434381	-	12	13	3	7	piRNA
183	chr10	41870533-41874493	-	12	13	3	9	piRNA
184	chr11	6049181-6050516	-	12	13	5	7	
185	chr11	95582191-95588104	+	12	12	3	8	piRNA
186	chr5	115404297-115411357	-	12	12	2	10	piRNA
187	chrX	17234869-17242153	-	12	12	5	5	
188	chr1	167702458-167706714	-	12	12	5	7	piRNA
189	chr18	65748663-65761591	+	12	12	4	5	piRNA
190	chr2	100920049-100927042	+-	12	12	4	7	
191	chr5	21320263-21320558	-	11	13	7	5	
192	chr2	28644737-28662243	-	11	13	3	9	piRNA
193	chrX	63882324-63887081	+	11	13	4	7	
194	chr1	37384478-37387632	-	11	13	4	9	piRNA
195	chr9	43894126-43903893	-	11	13	1	10	piRNA
196	chr4	139999570-140014688	-	11	12	3	8	piRNA
197	chr9	63445448-63450576	-	11	12	1	11	
198	chr16	96097712-96109333	-	11	12	3	9	
199	chr11	53260720-53281062	+	11	12	7	2	
200	chr14	75149422-75152155	+-	11	12	11	1	cis-nat-siRNA
201	chr2	24771332-24783870	+-	11	11	5	4	piRNA
202	chr5	135230763-135231375	+-	11	11	11	0	cis-nat-siRNA
203	chr15	83182267-83192941	-	11	11	3	7	
204	chr14	28860563-28869129	+	11	11	3	8	piRNA
205	chr15	93068290-93072348	-	11	11	3	8	piRNA
206	chr4	151146856-151150721	+	11	11	6	5	
207	chr13	64202446-64205517	-	11	11	7	4	piRNA
208	chr13	38352627-38353981	+-	10	13	13	0	cis-nat-siRNA
209	chr1	74913195-74918326	+	10	12	12	0	cis-nat-siRNA
210	chr10	4540955-4546987	+-	10	11	10	0	trans-nat-siRNA (130)
211	chr2	112370289-112374057	+	10	11	1	7	
212	chr5	78340488-78343421	+-	10	11	7	1	
213	chr10	126400198-126402431	+	10	11	9	2	piRNA
214	chr5	136194266-136273055	+-	10	10	5	5	piRNA
215	chr5	90691748-90711649	+	10	10	4	4	piRNA
216	chr7	5678553-5678593	+-	10	10	5	4	
217	chr5	72859514-72862801	+-	10	10	9	1	
218	chr16	24897620-24906135	+	10	10	1	8	piRNA
219	chr11	115401202-115420852	-	10	10	3	7	
220	chr4	120389015-120393077	+	10	10	3	7	piRNA
221	chr2	160450912-160462914	-	10	10	2	8	piRNA

222	chr13	51987033-51999657	+	9	12	3	9	
223	chr11	88781089-88788871	-	9	11	10	0	
224	chr15	34562755-34571836	-	9	11	0	10	
225	chr12	87403420-87410427	+	9	11	2	9	
226	chr4	138165033-138169671	-	9	11	4	7	piRNA
227	chr3	9233185-9234811	+-	9	10	9	0	
228	chr15	34220753-34228625	+	9	10	1	9	
229	chr9	27097633-27106063	+	9	10	5	4	piRNA
230	chr14	45001894-45007212	+	9	10	3	5	
231	chr15	80759962-80766473	+	9	10	2	7	piRNA
232	chr11	79501642-79513861	+	9	9	3	5	piRNA
233	chr3	33615089-33619227	+-	9	9	7	2	
234	chr5	108731683-108733445	-	9	9	7	2	
235	chr7	121863096-121872704	+	9	9	4	5	piRNA
236	chr1	49238587-49241102	+-	9	9	5	4	piRNA
237	chr15	98690860-98700158	+-	9	9	8	0	
238	chr1	91411467-91412653	+-	9	9	7	1	
239	chr4	127374823-127379960	+-	9	9	3	6	
240	chr19	5388341-5388997	+-	9	9	7	1	
241	chr16	10614394-10622645	+-	9	9	4	5	piRNA
242	chr4	47122099-47127933	+	8	10	2	8	
243	chr6	83771865-83774455	-	8	10	1	9	
244	chr18	44543508-44555328	+-	8	10	3	6	piRNA
245	chr9	76900901-76903575	+	8	9	1	7	
246	chr8	126133046-126141010	+-	8	9	7	1	
247	chr13	96436922-96438213	+	8	9	5	1	piRNA
248	chr11	120441863-120445572	-	8	9	5	4	
249	chr11	61499101-61501230	+-	8	9	8	1	
250	chr5	143804666-143818933	+-	8	8	3	4	
251	chr6	89517913-89527260	-	8	8	4	4	
252	chr15	98361734-98370374	+-	8	8	7	0	piRNA
253	chr8	119843540-119853024	+-	8	8	3	4	
254	chr2	151999779-152005308	+	8	8	1	6	piRNA
255	chr7	52448319-52452629	+-	8	8	3	5	
256	chr9	110102368-110108278	-	8	8	0	8	
257	chr11	97211493-97214410	+	8	8	2	6	
258	chr8	128934772-128950365	+	8	8	3	4	piRNA
259	chr9	8004651-8004985	+-	8	8	8	0	
260	chrX	6311562-6322912	-	8	8	0	7	piRNA
261	chr16	93562473-93562692	+	8	8	4	4	
262	chr5	111509922-111515929	-	8	8	4	4	
263	chr8	119356863-119366277	-	8	8	1	7	
264	chr18	33326187-33328502	-	8	8	2	6	piRNA
265	chr19	41637815-41653825	+-	8	8	1	6	piRNA
266	chr1	174213609-174219938	+	8	8	0	8	piRNA
267	chr7	139986307-139988398	+	8	8	2	6	
268	chr19	46932520-46939165	+-	8	8	2	5	piRNA
269	chr4	117254902-117258913	-	8	8	2	5	
270	chr5	36808213-36816089	-	8	8	1	7	
271	chr5	110721206-110727475	+-	8	8	6	1	
272	chr13	24995115-25000178	-	8	8	4	4	
273	chr2	33272500-33276347	-	8	8	1	7	piRNA
274	chr3	134792701-134794491	+-	8	8	4	3	
275	chr1	183038958-183038981	-	7	16	15	0	
276	chr15	42185408-42185433	+	7	11	8	2	
277	chr2	152952822-152961368	+	7	10	3	7	
278	chr5	141012034-141013417	+-	7	10	5	4	

279	chr17	5120193-5134102	+	7	10	3	7	piRNA
280	chr4	62619564-62623539	+	7	9	3	6	
281	chr19	15709187-15721356	-	7	9	1	8	piRNA
282	chr1	74631966-74634487	+	7	9	1	8	piRNA
283	chr18	87604937-87613299	+-	7	9	5	0	
284	chr10	14336030-14345739	+-	7	9	6	3	
285	chr3	8557865-8568358	+-	7	8	0	7	
286	chr2	28507994-28512110	+	7	8	0	7	piRNA
287	chr15	76159740-76166157	+-	7	8	7	1	
288	chr5	113729437-113734901	+	7	8	1	7	piRNA
289	chr1	172033076-172035869	-	7	8	5	3	piRNA
290	chr6	54570452-54574287	+	7	8	3	5	piRNA
291	chr4	45417642-45421236	-	7	8	2	6	piRNA
292	chr17	38152580-38164996	-	7	8	1	7	
293	chr1	90095231-90097373	-	7	8	3	5	piRNA
294	chr9	7184560-7184854	+-	7	7	6	0	
295	chr17	27617719-27629066	+-	7	7	0	6	piRNA
296	chr4	146309698-146315785	+-	7	7	3	2	
297	chr12	18317514-18321029	+	7	7	2	5	
298	chr11	60712602-60723112	-	7	7	3	4	
299	chr6	134822441-134824157	+	7	7	3	3	piRNA
300	chr4	126814167-126816934	+-	7	7	5	1	
301	chr2	3775123-3781227	+-	7	7	3	3	
302	chr6	128147654-128152146	-	7	7	2	5	piRNA
303	chr18	35392471-35402296	-	7	7	0	6	
304	chr1	105032976-105041994	+-	7	7	3	3	
305	chr18	80467071-80471925	+-	7	7	6	1	
306	chr16	8574827-8579722	-	7	7	1	5	
307	chr11	59271657-59278844	+	7	7	1	6	piRNA
308	chr4	140950037-140956836	-	7	7	0	6	piRNA
309	chr12	88049245-88060501	+-	7	7	3	3	piRNA
310	chr9	59284219-59284554	+-	7	7	6	0	
311	chr11	77331551-77338484	+	6	9	1	8	
312	chr8	113884363-113892285	-	6	8	6	2	piRNA
313	chr10	80993410-80999321	+-	6	8	4	4	
314	chr14	25927114-25932205	+	6	8	1	6	piRNA
315	chr15	62920430-62925880	-	6	7	3	2	
316	chr14	63769408-63772627	-	6	7	0	7	piRNA
317	chr14	83953963-83954918	+-	6	7	3	3	
318	chr4	3793047-3798960	+-	6	7	3	4	
319	chr16	18414509-18422102	+	6	7	0	7	piRNA
320	chr7	99795112-99800393	-	6	7	2	5	piRNA
321	chr12	82635238-82643169	-	6	7	5	2	piRNA
322	chr3	76055232-76060954	+-	6	7	4	3	
323	chr11	4765167-4772453	+	6	7	7	0	
324	chr2	164511521-164517325	+-	6	7	4	3	
325	chr16	32197122-32197924	+-	6	7	6	0	
326	chr12	105433762-105434028	+-	6	7	6	1	
327	chr2	153088831-153096012	+	6	7	0	5	
328	chr15	62198518-62207195	+-	6	7	4	3	
329	chr4	44132697-44137830	+	6	7	0	7	
330	chr6	26704111-26704153	+-	6	7	4	1	
331	chr12	8936286-8939367	+	6	7	2	2	piRNA
332	chr2	143631434-143634165	+-	6	6	2	3	
333	chr6	142968662-142972444	-	6	6	2	2	
334	chr5	147045545-147047971	-	6	6	0	4	piRNA
335	chr5	110872541-110883104	+	6	6	3	3	piRNA

336	chr7	81476972-81483420	-	6	6	1	5	
337	chr13	111459564-111461024	+-	6	6	5	1	
338	chr8	43042317-43048284	+	6	6	1	4	
339	chr8	124482114-124491255	+-	6	6	5	1	
340	chr9	35106703-35108480	+-	6	6	6	0	
341	chr4	59864700-59866047	-	6	6	1	5	piRNA
342	chrX	14035030-14039295	+	6	6	2	3	
343	chr4	100596727-100605257	+	6	6	0	6	
344	chr9	119284785-119288559	+	6	6	1	5	piRNA
345	chr7	121428230-121437711	-	6	6	2	4	piRNA
346	chr5	33980708-33988920	+	6	6	6	0	
347	chr1	35832995-35842811	+-	6	6	3	3	
348	chr10	36287103-36291897	+	6	6	1	5	
349	chr10	87500596-87509391	-	6	6	2	3	
350	chr1	180173816-180174451	+-	6	6	5	0	
351	chr16	98214356-98218563	-	6	6	1	5	
352	chr12	105166117-105171063	-	6	6	3	3	piRNA
353	chr13	56480018-56491622	+	6	6	2	4	
354	chr16	14908733-14911890	-	6	6	4	2	
355	chr12	3901604-3911961	+	6	6	1	5	piRNA
356	chr4	150779848-150787415	+-	6	6	2	1	
357	chr10	5988630-5991767	-	6	6	4	1	
358	chr4	125836721-125844131	-	6	6	2	3	piRNA
359	chr4	78320998-78322438	+	6	6	4	1	
360	chr11	97605437-97613251	+	5	35	2	0	
361	chr13	115316811-115316831	-	5	28	26	0	
362	chr9	69999182-69999201	-	5	24	2	0	
363	chr7	139809052-139809072	-	5	18	15	0	piRNA
364	chr8	98213653-98221198	-	5	12	6	6	
365	chr5	82107595-82107614	+	5	12	11	1	
366	chr5	136552505-136556146	-	5	8	1	7	
367	chr1	186523196-186523276	-	5	8	8	0	
368	chr4	118903961-118904197	-	5	7	7	0	
369	chr13	29347957-29355599	+-	5	7	0	7	piRNA
370	chr5	147264217-147269361	+-	5	7	4	3	
371	chr11	118844500-118847128	+	5	6	0	3	
372	chr15	31920753-31927669	+	5	6	0	6	
373	chr15	81512032-81515042	+	5	6	2	4	
374	chr5	143132237-143136224	-	5	6	2	3	piRNA
375	chr10	75483331-75487605	+-	5	6	2	4	
376	chr13	43313523-43315005	+-	5	6	4	2	
377	chr1	173177956-173178961	+-	5	6	6	0	
378	chr3	36756981-36760554	+-	5	6	5	1	
379	chr11	3093081-3093843	+-	5	6	5	1	
380	chr6	85389576-85397467	+	5	6	4	2	
381	chr5	34492245-34494538	-	5	6	4	2	
382	chr18	34896583-34899483	+	5	6	3	3	
383	chr3	11164754-11169061	-	5	6	2	4	
384	chr1	60421864-60425159	+	5	6	1	4	
385	chr1	162841119-162849022	-	5	6	3	3	piRNA
386	chr18	44787043-44787518	+-	5	6	2	3	piRNA
387	chr13	56745618-56751271	+	5	6	0	5	piRNA
388	chr19	35550807-35554633	+	5	5	1	4	
389	chr8	73281801-73286707	+-	5	5	5	0	
390	chr6	36578038-36579473	+	5	5	1	4	
391	chr5	136387678-136395935	+-	5	5	3	2	
392	chr1	152898423-152904530	+	5	5	2	3	

393	chr18	75293767-75302159	+	5	5	1	4	
394	chr10	69585523-69588019	+-	5	5	1	1	piRNA
395	chr5	135601446-135610329	-	5	5	2	3	piRNA
396	chr4	37556395-37556418	+-	5	5	3	2	
397	chr19	47314742-47316298	-	5	5	1	3	piRNA
398	chr4	134684500-134686381	-	5	5	1	3	piRNA
399	chr4	132469647-132470934	+	5	5	1	4	piRNA
400	chr9	21658122-21660328	-	5	5	1	4	
401	chr1	136571507-136574950	-	5	5	0	5	piRNA
402	chr4	115586752-115594523	+-	5	5	4	1	
403	chr6	118538604-118538644	+	5	5	3	2	
404	chrX	16663535-16669345	-	5	5	2	3	piRNA
405	chr9	70871324-70880973	+	5	5	0	4	
406	chr3	138046044-138046063	-	5	5	1	0	
407	chr9	63554665-63556113	+-	5	5	4	1	
408	chr6	116658718-116660694	-	5	5	1	3	
409	chr6	116367342-116373718	+	5	5	3	2	piRNA
410	chr6	147031958-147037907	+-	5	5	1	4	
411	chr1	46207886-46210345	+	5	5	2	3	
412	chr7	130413235-130414520	+	5	5	1	4	
413	chr7	67600806-67610020	+	5	5	2	3	
414	chr11	60353966-60355183	+-	5	5	3	1	
415	chr2	38546330-38549443	-	5	5	1	4	piRNA
416	chr2	25757145-25761954	-	5	5	3	0	
417	chr12	92038250-92044184	+-	5	5	2	3	
418	chr13	58137185-58137856	+-	5	5	5	0	
419	chr16	94918792-94919700	+	5	5	1	4	
420	chr9	65863575-65864527	+-	5	5	4	1	
421	chr11	97899431-97901848	-	5	5	3	2	piRNA
422	chr11	117824890-117831372	+	5	5	3	2	
423	chr12	85531341-85540000	+	5	5	1	3	
424	chr13	11550798-11555216	-	5	5	5	0	
425	chr13	23941242-23942569	+-	5	5	3	2	
426	chr15	27541614-27550415	-	5	5	1	4	
427	chr2	76541713-76550299	-	5	5	2	3	piRNA
428	chr12	111711571-111714318	+	5	5	2	3	
429	chr13	73976093-73982302	+-	5	5	3	1	
430	chr13	12668452-12676580	-	5	5	0	5	piRNA
431	chr2	163308159-163309840	+	5	5	2	2	
432	chr2	131996629-132001619	+	5	5	2	2	piRNA
433	chr2	21280255-21286211	+-	5	5	0	5	
434	chr4	25598518-25603682	+-	5	5	1	4	
435	chr11	57201351-57202592	-	5	5	2	3	
436	chr3	62691658-62692423	+-	5	5	4	1	
437	chr3	19341644-19349828	-	5	5	0	4	
438	chr10	33971755-33972395	+-	5	5	4	0	
439	chr5	66190136-66191498	+	5	5	2	2	piRNA
440	chr19	3171146-3172891	+	5	5	1	4	
441	chr14	25746223-25752890	-	5	5	1	3	
442	chr17	9027586-9032570	+	5	5	1	4	
443	chr2	26727417-26730230	-	5	5	0	4	
444	chr15	81650780-81656011	+-	5	5	2	3	

[1] This hp-siRNA cluster was found by observation, not by the computational search.

[2] The number in the parenthesis after "trans-nat-siRNA" represents the partner cluster.