Global characterization and target identification of piRNAs and endo-siRNAs in mouse gametes and zygotes.

*Jesús García-López, Juan de Dios Hourcade, Lola Alonso[1], David B. Cárdenas and Jesús del Mazo**

Department of Cellular and Molecular Biology; [1]Bioinformatics Service. Centro de Investigaciones Biológicas (CSIC), Ramiro de Maeztu 9, 28040 Madrid, Spain.

*Corresponding author: Phone: +34 91 837 3112. Fax: +34 91 536 0432. E-mail: jdelmazo@cib.csic.es

Running title: rasRNAs of mouse gametes and zygotes

Keywords: piRNAs, endo-siRNAs, deep sequencing, transposable elements, germ cells, zygote.

ABSTRACT

A set of small RNAs known as rasRNAs (repeat-associated small RNAs) have been related to the down-regulation of Transposable Elements (TEs) to safeguard genome integrity. Two key members of the rasRNAs group are piRNAs and endo-siRNAs. We have performed a comparative analysis of piRNAs and endo-siRNAs present in mouse oocytes, spermatozoa and zygotes, identified by deep sequencing and bioinformatic analysis. The detection of piRNAs and endo-siRNAs in spermatozoa and revealed also in zygotes, hints to their potential delivery to oocytes during fertilization. However, a comparative assessment of the three cell types indicates that both piRNAs and endo-siRNAs are mainly maternally inherited. Finally, we have assessed the role of the different rasRNA molecules in connection with amplification processes by way of the "ping-pong cycle". Our results suggest that the ping-pong cycle can act on other rasRNAs, such as tRNA- and rRNA-derived fragments, thus not only being restricted to TEs during gametogenesis.

HIGHLIGHTS

- The expression profiles of rasRNAs are different in oocytes and spermatozoa.
- The major piRNA group found in gametes was derived from tRNA and rRNA processing, suggesting alternative roles of this small RNAs in translation regulation.
- The ping-pong mechanisms of amplification of rasRNAs are not restricted to male gametogenesis and could have an important role in early embryogenesis.

1. INTRODUCTION

Mammalian genomes contain an abundant number of repeated DNA sequences, more than 50% in humans [1]. Highly repetitive DNA is normally localized in the genome in tandem clusters of multiple copies of untranslated regions, as for example satellite DNA, as well as in clusters of transcribed genes such as histones, ribosomal RNA (rRNAs) or transfer RNAs (tRNAs) [2-3]. However, a large number of DNA sequence repeats correspond to transposable elements (TEs) categorized in three main groups as long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) or long terminal repeats (LTRs). Since repetitive DNA regions can be transcribed into RNA molecules, the transcripts from LINEs, SINEs or LTRs can be self-propagated via RNA mediated transposition. TEs have played decisive roles in the evolution of genomes having accumulated in large portions of eukaryotic genomes. However, their transposition can induce, among other alterations, insertion mutations in protein-coding genes, enhancer or promoter modifications, abnormal chromatin rearrangements and also chromosome breakage [4].

Maintaining genomic integrity is essential during germ cell formation and subsequently in the early morphogenetic development of multicellular organisms. Mechanisms involved in genome protection evolved to guard from potentially deleterious transcription effects. Some groups of small RNAs, acting through RNA interference, have been related to the

posttranscriptional regulation of genes encoded by highly repeated DNA [5-6]. These repeat-associated small RNAs (rasRNAs) include piRNAs and endo-siRNAs [7-9]. piRNAs and endo-siRNAs are generated by different pathways, yet both can act through RNAi mechanisms that control transposon repression and TE mobility [6, 10-12]. The interaction between piRNAs/endo-siRNAs and repeat elements of genomes is dual, on the one hand repeat transcripts can be the source of piRNA and endo-siRNA biogenesis, and on the other hand transcripts of repeat elements are targeted by rasRNAs.

In mouse, piRNAs are ≈ 23-32 nt long RNAs that act in association with PIWI proteins: MILI, MIWI and MIWI2 [13-15]. The biogenesis of piRNAs is not completely understood, nonetheless, two different generation pathways seem to be involved. The first way comprises the expression of specific areas of the genome encoding large RNA precursor molecules, which subsequently must be cleaved into smaller fragments [16-17]. The second however is more related to their function. It has been established that piRNAs bind to TEs [18-19] and that the binding between each piRNA and their TE target RNA takes place at the 5′ end of the piRNA sequence. Upon binding, poorly characterized proteins participate in the cleavage of TE transcripts after nucleotide position 10 at the 5′end of the piRNA binding site. Cleavages generate new piRNAs derived from the RNA of TEs [18-19] in a process known as the "ping-pong model" or "ping-pong cycle" [14, 20-21]. The ping-pong cycle generates characteristic cleavage specific piRNA sequence signatures [14, 20] in this manner, piRNAs whose targets are the TE transcripts show a bias for uridine at the 5′ end (*5U*), whereas the piRNA sequences generated from TE cleavage show a bias for nucleotide adenosine at position 10 (*10A*) [14, 20]. These characteristics have been conserved in such diverse organisms as the mouse, *Drosophila*, zebrafish and silkworm [9-10, 22-23].

Mutated PIWI family genes have been related to severe defects in male fertility [24-25], on the contrary, female fertility was noticed to be unaffected by PIWI protein inactivation [24-25]. Female fertility and early development progression seem to be associated to the presence and activity of DICER [26-30] and AGO2 [31-34]. These observations have been related to the

functions of endogenous small interference RNAs (endo-siRNAs) present in the female germline.

Endo-siRNAs are ≈ 21-24 nt long sequences derived both from sense-antisense transcripts yet mainly proceed from transposon transcripts [6, 18, 35]. Contrary to piRNAs, endo-siRNAs have been determined as specific of the female germline and preimplantation embryos [18, 35-36]. DICER and AGO2 proteins participate in endo-siRNA biogenesis and silencing activity respectively, while *Dicer* and *Ago2* null mice displayed lethality during early embryogenesis [37-39]. The regulatory role of endo-siRNA during mammalian gametogenesis is still poorly understood and endo-siRNA inheritance from gametes to embryo has not been clarified yet.

The piRNAs and endo-siRNAs are also known to hold crucial roles in transposon silencing in order to avoid epigenetic disorders that could compromise embryo viability at early developmental stages. In the present study, we have used high throughput sequencing technology to perform a global characterization of piRNA and endo-siRNA sequence populations present in mouse spermatozoa, oocytes and zygotes. Based on the RNA targets of piRNAs and endo-siRNAs we have evaluated their potential roles with regard to gamete differentiation and fertilization. Finally, we have tracked ping-pong cycle events from gametes to early zygote formation.

Our findings, suggest a vital role of maternally inherited piRNAs and endo-siRNAs in connection with the repression of transposons. However, transposable element regulation appears not to be the exclusive role performed by rasRNAs. We have also detected an elevated ratio of piRNAs and endo-siRNAs derived from rRNA or tRNA molecules which could consequently be involved in translation regulatory pathways.

2. MATERIALS AND METHODS

2.1. Animals

All procedures relating to the care and handling of the animals used in the present study, were carried out in the CIB-CSIC bioterium under specific pathogen-free (SPF), temperature (22±1ºC) and humidity-controlled (50-55%) conditions. All animals were housed on a 12h light/dark cycles with *ad libitum* access to food and water. Animal care and handling was carried out in accordance with the regulations of the Bioethics Committee of the *Consejo Superior de Investigaciones Científicas (CSIC)* that approved the study, and adhering to the European Commission guidelines.

2.2. Spermatozoa collection

Cauda epididymis and vasa deferentia from mature CD-1 males were collected in 500 µl of M2 medium after which adipose tissue and blood vessels were removed. Cleaned structures were placed in a new 200 µl drop of M2 medium covered with mineral oil. Epididymal fluid was squeezed out and sperm was suspended in M2 medium. Concentrations were determined with a Neubauer hemocytometer. The sperm suspension was layered in a 15 ml conical tube on a discontinuous 90%/45% Percoll gradient in a 1:1:1 ratio (cell sample: 45% Percoll: 90% Percoll) and centrifuged at 700 *g* during 20 min. The bottom pellet was washed in PBS and centrifuged at 350 *g* during 5 min. In order to entirely remove contaminant cells, a hiposmotic shock was induced by resuspension of spermatozoa in H_2O-DEPC. Finally, the pellet was resuspended and homogenized in 100 µl of TRIzol® Reagent (Invitrogen) and frozen at -80ºC until use.

2.3. Oocyte and zygote collection

Fully-grown oocytes and zygotes were collected from the oviducts of mice as has been described previously [40]. Briefly, superovulation was induced in 4-5 week old C57BL6 female mice by intraperitoneal injection of 5 IU pregnant mare serum gonadotropin (PMSG), followed 48 hours later by 5 IU of human chorionic gonadotropin (HCG). After HCG administration, female mice were mated with DBA/6J males. Fully-grown oocytes and zygotes were treated with hyaluronidase (300 µg/ml, Sigma H3884) in order to remove cumulus cells by passage

through several drops of M2 medium (Sigma M7167). To carry out the high-throughput sequencing a total of 15,210 fully-grown oocytes and 15,416 zygotes were collected. In all cases, the zona pellucida was removed by incubation in 30 μl of prewarmed acidic Tyrode solution under mineral oil during approximately 15 seconds, followed by careful washing in three drops of M2 medium under a stereomicroscope. Samples were stored in TRIzol® Reagent (Invitrogen) and frozen at -80ºC until use.

2.4. RNA purification and sequencing

Total RNA from 15,200 metaphase II oocytes, 500 million of spermatozoa and 15,400 zygotes was isolated using TRIzol® Reagent (Invitrogen) according to the instructions provided by the manufacturer. RNA concentrations were quantified measuring absorbance (A260/280 ratio) on a NanoDrop Spectrophotometer ND-1000 (NanoDrop). RNA integrity, considering the particular profile of spermatozoa, was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). Illumina protocols (available on the Illumina website at www.illumina.com/support) were followed to prepare the small RNA libraries from total RNA. Briefly, 3´and 5´adapters were ligated to each RNA molecule existing in total RNA previously isolated from spermatozoa, oocytes and zygotes. After adapter ligations an RT reaction was performed to cDNA synthesis. In order to avoid any bias, the cDNA was PCR amplified using common primers that were designed against adapter sequences. After cDNA amplification, the small RNA fraction was isolated by gel purification. Small RNA fraction was sequenced by high-throughput deep sequencing using an Illumina® Hiseq2000 sequencer.

2.5. Data analysis

Adaptor sequences were identified and removed (DNAVision SA, Belgium). Sequencing using the Illumina platform yielded sequences with a length range of 18-32 nucleotides. Using bioinformatics approaches we proceeded to classify all the small RNA sequences according to their identity with the annotated small RNA sequences found in public databases. The remaining sequences that had not identity with any small RNA sequence,

ncRNA sequence or mRNA sequence were classified as unannotated (see Supplemental Fig. S5). The term "read" will strictly refer to the number of times that any RNA sequence has been detected, the term “sequence” will imply the specific succession of nucleotides sequenced, and the term "tag" will stand for the identifier of the defined element to which each sequence belongs. Therefore, the sum of all reads concerning a defined tag represents its expression value. For example, regarding the sequences: *TAAAACTGTATTTGAATTTGGGGC*; *TAAAACTGTATTTGAATTTGGGGCGA (read twice)* and *TAAAACTGTATTTGAATTTGGGGCGAGC*; this case represents 4 reads and 3 sequences although only 1 tag, the underlying reason being that all these sequences which differ in length were previously identified as the piRNA PIR222563 of the “RNAdb 2.0” [41-43] . Therefore, the sum of reads represents the expression value of the piRNA PIR222563. In order to compare the tag representation of different samples, we applied the DeSeq tool of the R/Bioconductor software package [44-46].

Tags with only one read were excluded from the analysis, while the remaining tags were classified following the pipeline shown in Supplemental Figure S5, and annotation was undertaken using the Bowtie (0.12.9 release) [47] to align the sequences to the miRBase 19, piRNA database and Deepbase, allowed for up to three mismatches, whereas alignments to the Rfam and ncRNA databases were performed allowing for no mismatches. This permissive alignment was due to the potential editing of double-stranded small RNA precursor molecules [48-49]. The total number of reads of each sample was different, therefore, in order to compare the reads among different samples, reads were normalized using a negative binomial distribution based method employing the DESeq tool of the R/Bioconductor software package [44].

2.6. **Ping-pong analysis**

Ping-pong signatures were searched in all sequences classified as piRNAs. Although piRNA biogenesis is not yet fully understood, the ping-pong model could be accepted as a possible process that encompasses self-amplification and produces a recognizable signature.

Overlapping regions between sense and antisense piRNA sequences were observed which presented a perfect match complementarity along the first 10 nucleotides at the 5' end. As detailed in results, we searched for overlapping regions between the sequences that corresponded to the primary piRNA group and the secondary piRNA group. The same type of search was also performed with sequences clustered in groups 5U10A and other. Moreover, an *inter* and *intra* sample search was performed, seeking for putative piRNAs in oocytes or spermatozoa that could be the source of amplification rates in zygotes. The results of this quest were classified as *match* or *no match*. The complementarity between uncoupled tags (*no match*) and mRNA was also evaluated using the BLAST aligner program (blastall; -1e -8; -W11) and the *cDNA Mus_musculus.NCBIM37.64* database. High throughput sequencing generated several sequences that shared the same seed but differed in length; hence in the alignment against the *cDNA Mus_musculus.NCBIM37.64* database we always made use of the longest sequence of each tag.

2.7. **Data availability**

Public data base accessions are provided for all raw datasets and all data obtained as result of ping-pong analysis have been deposited in the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE45983.

3. RESULTS

Small RNAs expressed in spermatozoa, oocytes and zygotes were size selected (18-32) from total RNA and sequenced by high-throughput sequencing (Illumina) (see experimental procedures). After removal of the low quality 3´ ends and linker sequences, the analysis of small RNA libraries returned 34,111,418 reads derived from spermatozoa, 53,616,426 reads from oocytes and 53,064,988 reads from zygote samples. These reads corresponded to 2,325,500

unique sequences of spermatozoa, 1,466,161 of oocytes and 1,671,944 of zygotes. Although the total number of reads was lower in spermatozoa than those of oocytes or zygotes, the spermatozoa sample contained a far higher degree of sequence diversity (Table 1 and Supplemental Fig. S1).

By means of bioinformatics approaches we were able to classify all the small RNA sequences according to their reference identities retrieved from the public databases. The terms used as "read", “sequence” and "tag" are defined in Material and Methods.

3.1. Identification of piRNA and endo-siRNA

The first challenge of the present study was the characterization of small RNAs identified in oocytes, zygotes and spermatozoa into miRNAs, piRNAs and endo-siRNAs. The miRNAs are derived from double stranded precursor molecules which are contained in “miRBase” database. However, due to the piRNA and endo-siRNA populations have been less studied the discrimination of both populations entails certain difficulties. Both piRNA and endo-siRNA sequences present structural modifications at their 3´ termini. According to the Illumina protocol used in the present work, the method can target small RNAs with 3´ hydroxyl group as endo-siRNAs have. In fact, Illumina sequencing have been employed to confirm the presence of piRNAs that present 2’ and 3´ hydroxyl groups which are characteristics of piRNA sequences [50].

In order to accomplish piRNA annotation and taking into account the above premises, the sequences which were not annotated as miRNAs were aligned against piRNA homologues retrieved from the “RNAdb 2.0” [51], a database of mammalian non-coding RNAs, (http://research.imb.uq.edu.au/rnadb/FastaDownloads/default.aspx). The “RNAdb 2.0” contains set of small RNA sequences classified as piRNAs which were obtained after inmunoprecipitation assays using PIWI proteins [41-43]. The inmunoprecipitation assays were carried out using total RNA isolated from human, rat and mouse testis samples. So the first requirement which we considered to classify a sequence as piRNA, was to present more than 84% of sequence homology with the piRNA sequences contained in “RNAdb 2.0”.

However, the sequences identified in the present analysis could contain embryo-specific or gamete-specific piRNA sequences that were not represented in "RNAdb 2.0". It has been reported that both piRNAs and endo-siRNAs can be associated to transposons and other genome repeated sequences (reviewed in [52-55]). To test this, we compared the sequences not aligned to "RNAdb 2.0" with the sequences contained in "Deepbase" database (http://deepbase.sysu.edu.cn/download.php), which is one of the major sources of repeat associated RNAs. We can distinguish piRNAs and endo-siRNAs due to their sequence length differences. So, the sequences identified as rasRNAs after "Deepbase" alignment were further divided in two categories: piRNAs and endo-siRNA depending on sequence length, i.e. sequences annotated as piRNA consisted of more than 24 nucleotides, while endo-siRNA population comprised sequences with a length between 18-24 nucleotides. Finally, piRNA population contained sequences that are homologues to piwi-interacting sequences plus sequences that have more than 24 nt and were associated to genome repeat elements as transposons, pseudogenes, tRNAs or rRNAs.

3.2. Small non-coding RNAs of oocytes, sperm and zygotes are mainly derived from processed transcripts of highly repeated sequences present in the genome

Approximately 50% of the total RNA sequences detected in oocytes and zygotes were identified as piRNAs or endo-siRNAs (rasRNA category) (Table 1). These sequences encompassed 80% of the total reads detected in oocytes and zygotes, considerably differing from the values observed in spermatozoa, in which case less than 60% of the total reads corresponded to sequences identified as rasRNAs (Table 1). Despite of the existing differences detected among the three samples, our data nonetheless strongly suggest that rasRNAs play important roles during gametogenesis and fertilization. Surprisingly, sequences identified as miRNAs represented less than 6% of the total reads obtained from all samples. This data contrasts with a previous analysis reported in echinoderms where the miRNA population reached proportions of up to 60% [56]. At the moment, a detailed analysis of the miRNA from these samples is in progress (manuscript in preparation).

In order to establish the origin and role of the corresponding rasRNAs detected in each sample, we performed a multiple alignment of the rasRNA sequences to repetitive element annotated classes of the mouse genome. The multiple alignments revealed important differences between the spermatozoa *versus* the oocyte and zygote derived samples (Fig. 1). While in oocytes and zygotes respectively 72% and 85% of rasRNAs were associated to TEs, in the spermatozoa only 10% of all the rasRNAs detected was found to be associated to TEs (Fig. 1B). Most rasRNAs were associated to LINE and LTR sequences in both samples of oocytes and zygotes, while in the spermatozoa sample rasRNAs were for the most part associated to tRNAs and rRNAs (Fig. 1B). These differences were even more conspicuous when the population of rasRNAs was divided into piRNAs and endo-siRNAs and the same type of analysis was carried out (Fig. 2). In oocytes and zygotes, endo-siRNAs were primarily associated to LINE and LTRs in contrast to the endo-siRNAs of spermatozoa which displayed heterogeneous associations with different genome elements (Figs. 2D-2F). In a curious manner, even though sperm piRNAs were also heterogeneously associated with different genome elements (Fig. 2B), in this case their read distribution was biased to those piRNAs that aligned with rRNAs. Considering the overall biogenesis of piRNAs and endo-siRNAs, these results suggest pronounced activities of LTRs and LINEs during female gametogenesis and in early zygote formation emphasizing the significance that piRNAs and endo-siRNAs hold to prevent the free transposition of TEs.

Overall, 19% (45,145) of the total endo-siRNAs were exclusively detected in one sole cell type and were thus classified as cell specific (see Supplemental Fig. S2). In spite of the fact that endo-siRNAs have been previously described as specific of the female germline, we were able to classify up to 71,780 sequences as endo-siRNAs present in spermatozoa, whilst some of them could even be classified as spermatozoa-specific. Of all the spermatozoa specific endo-siRNAs detected, one of the most expressed was associated to satellite genomic regions, this endo-siRNA (endo_mmur001237608) showed over 10,000 normalized reads and was concretely associated to one specific type of satellite DNA: GSAT.

Our comparative results suggest that the mechanisms of action controlling genome

repeats, mediated by piRNAs and endo-siRNAs, are of a very different nature with regard to the developmental progress of gametogenesis in both sexes as well as to the particular functions carried out by piRNAs or endo-siRNAs. However, the presence of TE-derived endo-siRNAs in spermatozoa suggests that the endo-siRNA regulatory activity controlling transposons is not restricted to the female germline [34, 57].

Collectively, the alignment results of the rasRNAs sequences to the genome were similar in oocyte and zygote samples. This fact entails a prominent maternal inheritance of piRNAs and endo-siRNAs contributing to transposon mobility regulation in zygotes.

3.3. tRNA and rRNA-derived fragments as new classes of PIWI-interacting RNAs or endo-siRNAs

Sequences classified as piRNAs or endo-siRNAs were mapped to the mouse genome. As results, we identified in all three samples a large group of piRNAs derived from tRNA and rRNA cleavage (Figs. 2A-2C). About one third of all piRNA reads detected in gametes of both sexes in addition to 20% of the zygote piRNA reads corresponded to tRNA-derived fragments (also known as tRFs) generated through mature tRNA precursor processing [58]. piRNAs derived from tRNA-Gly and tRNA-Glu were the most abundant in all three sample types (Fig. 3A). Upon analysis of piRNA sequences derived from tRNAs, we observed a strong bias in their 5′ and 3′ ends (Fig. 3B). Over 90% of all piRNAs derived from tRNAs had a 3′ end consisting of the nucleotide triplets CCT, CGC, GCC and GGC (Fig. 3B). Conversely, 5′ ends were mainly represented by CCC and GCA nucleotide triplets (Fig. 3B). Considering these 5′ and 3′ end biases, it results difficult to argue that piRNAs derived from tRNAs are a mere RNA degradation product [59]. Moreover, the tRNAs-derived piRNAs reported in the present study were of a quite different nature from the tRNA fragments that have been found as result of stress induction [60], mainly due to differences detected at the 5′and 3′ ends of the piRNA sequences. All together, this suggests that tRNAs-derived piRNAs might encompass specific roles in germ cell development and zygotic activation.

Interestingly, as yet another "unconventional" small RNA with regulatory functions [61] we also found piRNAs and endo-siRNAs generated from rRNA cleavage. This rRNA-derived group was particularly abundant in spermatozoa. In sperm cells more than 50% of the total reads of piRNAs and 37% of endo-siRNAs belonged to rasRNAs derived from rRNA sequences. In particular, the highest represented sequence of all samples (CGACTCTTAGCGGTGGATCACTC) annotated as PIR64428 corresponds to the cleavage product of the 5.8S ribosomal RNA (Rs5-8s1). This sequence was detected in over 4 million reads in the spermatozoa sample. The expression profiles of this piRNA sequence along with the expression of other rasRNAs were validated by RT-qPCR (data not shown). At present, we still do not know the possible role of this abundant piRNA.

3.4. Maternal and paternal contributions of piRNAs and endo-siRNAs to the zygote.

A comparison of the piRNAs and endo-siRNAs detected in oocytes, spermatozoa and zygotes, allowed to assess the origin of the rasRNAs detected in zygotes.

3.4.1. piRNAs.

By means of bioinformatics approaches, we globally annotated 154,734 different tags as piRNAs. Of these 18% (28,069) were detected in all three samples while 46% (71,914) were only detected in one of the three samples and were thus classified as specific piRNAs. The remaining 36% (54,571) of the piRNAs detected were present in two of the three samples (Fig. 4A).

The main contribution of maternal piRNAs to the zygote became more evident after undertaking an analysis of the levels of the highest represented tags (≥ 700 normalized reads) predominant in each cell type. By non-supervised hierarchical clustering (Fig. 4A) we detected similar expression patterns in oocytes and zygotes. Nevertheless, the piRNAs associated with LINEs and LTR transposons increased their expression levels in the zygote with respect to both gamete types (Fig. 4B), whereas the expression of piRNAs derived from tRNA and rRNA cleavage showed a relative reduction with respect to the rates detected in gametes of both sexes

(Fig. 4B). This suggests that while the piRNAs associated to tRNAs and rRNAs could exhibit specific roles during gametogenesis, the piRNAs related to LINEs and LTRs were mainly activated after fertilization.

The MA-plots which represent expression values from shared tags in the oocyte/zygote and spermatozoa/zygote samples shed light on the dynamic of piRNAs changes as a result of fertilization along with the putative parental contribution (Fig. 4F, D). Of 48,631 piRNAs shared by oocytes and zygotes, 60% (29,336) showed similar expression levels before and after fertilization (Fig. 4C). In contrast, 31% increased their expression levels over a twofold in zygotes with respect to the levels detected in oocytes (Fig. 4C). These after fertilization upregulated piRNAs were mainly associated to LINE and LTR transposons (Fig. 4C). The preponderance of piRNAs derived from processing of LTR transcripts, suggests an increase of LTR activity shortly after fertilization. Finally, our analysis showed that 14% of the piRNAs decreasing a twofold as a consequence of fertilization were derived from rRNA cleavage (Fig. 4C).

The paternal piRNA contribution to the zygote can only be assessed by a comparative analysis of the piRNAs detected in spermatozoa and in the zygote but not present in oocyte derived samples. MA-plots indicated that potential piRNAs of paternal origin are quantitatively represented by very low levels as shown by the average expression (normalized reads) detected in shared piRNAs of spermatozoa and zygotes (Fig. 4D). Furthermore, the number of tags corresponding to such shared piRNA populations was lower (2% of total tags) in comparison to those detected in the oocytes/zygotes (31%) (Fig. 4A). Of these, 53% (1,517 piRNAs) decreased their expression levels after fertilization, while only 14% (411 piRNAs) were found to be upregulated (Fig. 4D). These upregulated piRNAs were mainly derived from LINEs, LTRs and SINEs (Fig. 4D), which agrees with the activation of these types of TEs as has been suggested above. A total of 26% of piRNAs showing lower levels after fertilization corresponded to piRNAs of unknown targets (Fig. 4D). Most probably, these piRNAs had not been eliminated during spermiogenesis and therefore could have been released into the oocyte at fertilization.

Such putative piRNAs had low read levels in spermatozoa as well as very low read levels in zygotes. Yet, if these piRNAs had been supplied to the zygote by the sperm cells, they were not amplified in the zygote by ping-pong mechanisms. All together, these results do not exclude the possibility that paternal piRNAs were transmitted to zygotes through sperm cells, but even if so, their contribution does not seem to be functionally significant, at least quantitatively.

Finally, with regard to the specific piRNAs detected in each cell type, in oocytes and zygotes they were mainly associated to SINE, LINE and LTR transposons while in spermatozoa the specific piRNAs were derived from piRNA clusters comprising unknown targets (see Supplemental Fig. S2).

3.4.2. endo-siRNAs.

The analysis carried out on endo-siRNA populations classified them in 234,364 tags, of which 34% were shared by the three samples (Fig. 5). These endo-siRNAs were heterogeneously derived from TE and non-TE sequences (Fig. 5B). Similar to the results found for the piRNA analysis, we observed that the endo-siRNA expression profiles detected were almost identical in oocytes and zygotes and radically different from the sperm profile. This became evident after undertaking a hierarchical cluster analysis (Fig. 5A) and upon analysis of the read distribution of shared tags (Fig. 5B). Furthermore, 99,316 tags (56.3% of the total tags detected), were shared by oocyte and zygote samples (Fig. 5A). This strongly suggests that the endo-siRNAs detected in zygotes corresponded mainly with a maternal inheritance. For most of the endo-siRNAs identified in zygotes and oocytes an association to LTR transposons was observed. More than 50% of the tags that increased over a twofold in the zygote were LTR-associated. In contrast, endo-siRNAs that showed a relative decrease in their expression after fertilization were predominantly associated to LINEs (Fig. 5C).

An analysis performed on the endo-siRNAs shared by spermatozoa and zygotes revealed that they only represented 2.4% (5,683) of all the endo-siRNAs detected (Fig. 5A) and in addition that they were preferentially aligned to LINE and SINE repeats (Fig. 5D). In

agreement with data reported previously [62], we detected a group of endo-siRNAs derived from SINE B1 that were highly expressed in the zygote. As these endo-siRNAs were not detected in oocytes we therefore proceeded to examine their inheritance potential by paternal contribution. However, at read evaluation of these endo-siRNAs in spermatozoa and zygotes, we considered that the presence of SINE B1-derived endo-siRNAs in zygotes should be attributed to Zygotic Genome Activation (ZGA).

The further analysis of other endo-siRNAs such as those derived from GSAT sequences provided similar conclusions about the limited potential contribution of sperm endo-siRNAs to the zygote. One of the highly expressed endo-siRNAs in sperm (endo_mmur000317426) was also detected in zygotes but not in oocytes. However, its expression level could be considered as residual in zygotes compared to the level observed in spermatozoa (from more than 12,000 normalized reads in spermatozoa compared to 2 normalized reads in zygotes).

Finally, we identified specific endo-siRNAs in samples obtained from spermatozoa, oocytes and zygotes. While the specific endo-siRNAs of oocytes and zygotes were associated to LINE and LTR transposons, the specific endo-siRNAs detected in sperm showed heterogeneous associations with TE and Non-TE sequences (see Supplemental Fig. S2).

3.5. Functional dynamics and characterization of piRNAs detected in female and male gametes and the zygote.

Initially, according to their biogenesis during spermatogenesis piRNAs were grouped into pachytene and pre-pachytene classes. However, this classification refers to the particular stage of spermatogenesis in which piRNAs have been isolated [63]. Alternatively, due to the presence of piRNAs in somatic cells [64-65] together with their sequence and genomic location heterogeneity, a more functional classification based on sequence features has been proposed [9, 20, 54, 57, 66-67]. Attending to sequence features we classified piRNAs in four groups: *primary piRNAs* (5U), *secondary piRNAs* (10A), *5U10A piRNAs* and *other piRNAs* (see Supplemental Fig. S3). The primary and secondary groups contain sequences with uridine in the

first nucleotide from the 5´ end, and adenine at position 10, respectively. The 5U10A group includes sequences that enclose both features: uridine at the 5´ end, and adenine at position 10. The group named "other" includes those sequences that contain none of the above features (see Supplemental Fig. S3A, S3B).

Analyzing the distribution of these four types of piRNAs we found that the oocyte and zygote samples showed a similar distribution pattern. However, in spermatozoa the ratio among the four groups of piRNAs was different (see Supplemental Fig. S3A). The piRNAs classified as "*other"* of the sperm sample contained a larger number of sequences (see Supplemental Fig. S3A). In contrast, in oocyte and zygote samples, primary and 5U10A piRNAs were found to be predominant. This indicates that the biogenesis of piRNAs varied according to the gamete type in question. The maternally inherited pattern for piRNA biogenesis is maintained in the zygote with a slight increase of both primary and secondary as well as 5U10A groups.

One of the purposes of our study was to evaluate the dynamics of piRNA amplification by the ping-pong cycle in relation to each cell type and as a consequence of the maternal to zygotic transition. For that matter, we first analyzed the rate of primary piRNAs that showed an overlapping and perfect match of the first 10 nt at the 5' end with secondary piRNA sequences (see Supplemental Fig. S3B). Initially, the quest was performed among primary and secondary piRNA groups because pairing rules of the ping-pong cycle were initially described for piRNAs that presented 5U and 10A sequence features (reviewed in [18]). Alternatively, despite of the fact that no matching rules have been previously described for both of these types of piRNAs, we asked whether the piRNA included in the *5U10A* and *"other"* groups were also subject to amplification by the ping-pong process. Thus, in order to search for potential pairings, the sequences contained in the *5U10A* and *"other"* groups were matched internally with themselves. Following this criterion and considering all groups together, the piRNAs classified as 5U10A were the most active as to forming ping-pong pairs (see Supplemental Fig. S3C and GEO deposited data), especially in both the oocyte and zygote samples, suggesting that this specific sequence feature could facilitate the amplification process of piRNAs. In spite of the

fact that the ping-pong cycle was initially described in *Drosophila* male germ cells, in this study, we identified more active amplification processes in oocytes and in zygotes than in spermatozoa (see Supplemental Fig. S3C). Consequently, our results indicate that: a) the amplification based processes of the ping-pong model operate differentially in mouse oocytes, zygotes and spermatozoa; b) piRNAs considered as primary and secondary as well as those classified as 5U10A and "other" could all actively participate in the amplification at a similar level; and c) the piRNAs participating in the ping-pong cycle were not only associated to TEs elements as we also detected piRNAs associated to tRNAs and rRNAs (see "ping-pong analysis" in the corresponding GEO data file).

Finally, to identify potential piRNAs that could be acting as microRNAs, the piRNA sequences without a ping-pong partner were aligned in a reverse complementary manner with those sequences contained in the *Mus_musculus*.NCBIM37.64 cDNA database. We found that 3%, 2.8% and 5.5% of the sequences with these features were respectively matched to different mRNAs in oocytes, zygotes and spermatozoa. Globally, 657 mRNAs can be identified as potential target to piRNAs distributed amongst the three cell samples, mostly localized on oocytes (Supplemental figure S6). The identification of target mRNAs suggested the participation of these particular piRNAs in relevant biological functions associated to the germ cell differentiation and fertilization. For example, those piRNAs targeting shared mRNAs between spermatozoa and zygotes but not presented in oocytes could indicates putative piRNAs spermatozoa-specific potentially delivered to the oocyte by fertilization. But from our results, this group of piRNAs targeted only 15 mRNAs, showed low number of reads and no represented relevant ontological pathways involved in early development. Nevertheless, a group of genes with high expression in spermatogenic cells but incompatible with early postzygotic development were potential targets of specific oocyte piRNAs. Among them, the most relevant genes were: *Ctcfl* (Ctcf like), *Tcp11* (T-complex protein 11), *Strbp* (Spermatid perinuclear RNA-binding protein) and *Zscan2* (zinc finger and SCAN domain containing 2). All these genes are specifically expressed in spermatogenesis. In particular, *Ctcfl* (also named BORIS for

brother of the regulator of imprinted sites) is a testis-specific paralog of *Ctcf* directly involved in the parental imprinting mechanisms, specifically on the paternal imprinting of *H19* [68-69]. The presence in the oocyte of piRNAs that potentially target male specific transcripts might prevent the translation of these mRNAs that could be incompatible with the postzygotic development.

4. Discussion

A comparative characterization of mouse oocytes, zygotes and sperm cells, based on a large number of small RNAs identified by means of deep sequencing and bioinformatic analysis, has so far not been carried out. This study has allowed us to deepen our insight into the biogenesis of piRNAs and endo-siRNAs, as well as to discern their expression differences in germ cells of both sexes and to quantify their potential contributions to zygote formation.

Studies have revealed that piRNAs are mainly involved in TE silencing, epigenetic modifications and in the development associated to male germ cells [67, 70-72]. However, their presence has also been reported in oocytes and inclusive in brain tissues [64-65, 73-75]. TE mobility control during gametogenesis and zygotic genome activation determines reproductive fitness and embryo viability. Our data strongly suggest that the regulation of TEs during the maternal to zygotic transition relied primarily on the maternally inherited piRNAs and endo-siRNAs. Nevertheless, our identification of specific cell type piRNAs and endo-siRNAs suggests that in the transition from oocyte to zygote some maternally inherited rasRNAs are eliminated in the zygote while at the same time some specific piRNAs and endo-siRNAs are expressed after fertilization. The replacement of maternal transcripts by those that are zygotically expressed is a well-known process that although evidenced in the mouse at the 2-cell embryo stage is already initiated in the 1-cell embryo phase (reviewed in [76]). Our results are in concordance with the results previously obtained by two different analyses of zygotic gene expression which describe that about 20% of the genes detected corresponded to repetitive sequences [77-78].

The specific piRNAs and endo-siRNAs we have detected in spermatozoa could be involved in spermatogenesis and remain as a result of spermatozoa differentiation. However,

their almost complete absence in the zygote evidenced the limited, if any, transference of these RNAs to the oocyte as a consequence of fertilization. Our comparative study of rasRNA populations has raised doubts about whether stored rasRNAs in the spermatozoa could contribute to the zygote as result of fertilization or on the contrary they consist of sheer remnant products of piRNA degradation that occurs during spermiogenesis as has been recently reported [79]. That is, from the 48,978 sperm specific piRNAs detected, the alignment to the genome showed that 75% of these piRNAs corresponded to unknown targets similar to those previously reported in a pachytene piRNA population (reviewed in [18]).

As shown in Supplemental Figure S3A in Additional data file 1, the main group of piRNA detected in spermatozoa lacks any ping-pong signatures and was therefore classified as "*other"* piRNAs. The alignment of these sequences to the genome showed that the majority of tags that we classified as "*other"* piRNAs were derived from tRNA and rRNA molecules (see Supplemental Fig. S4). This high amount of processing of tRNA and rRNA as a source of piRNAs has so far not been reported. One of the key features of piRNAs is their interaction with PIWI proteins [6, 80-82]. We have identified as potential piRNAs those rasRNAs showing at least 84% identity with sequences contained in the piRNA database. Since sequences included in the piRNA database have been isolated using immunoprecipitation assays with PIWI proteins [41-42, 83-84], the tRNA derived fragments (tRFs) [58] and the rRNA derived fragments could be considered as a subtype of the PIWI interacting RNAs. In this sense, it has been previously proposed that tRFs are involved in transposon silencing, based on the complementary nature of their 3' end [59, 85]. However, this could possibly not be the sole role of tRFs. It has also been suggested that the tRFs possess a role in translation inhibition and could even target certain mRNAs for degradation [86].

Despite of their relative abundance in the three cell types studied here (Fig. 2), the particular role of piRNAs derived from tRNAs and rRNAs is at present unknown during gametogenesis and early development. Presence of tRNAs and rRNAs in sperm nuclei has been previously reported [87]. In the present study, the most abundant tag detected in spermatozoa consisted of a piRNA derived from 5.8S rRNA: Rs5-8s1 cleavage. The 5.8S rRNA subunit

participates in the ribosome translocation binding to ribosome-associated tRNA, *in vitro* studies have suggested that it retains an essential role in protein synthesis [88]. Altogether, our findings suggest an additional potential role of these piRNAs with regard to translation regulation through interaction with the translation machinery elements.

The substitution of maternally inherited mRNAs by newly formed zygotic transcripts implies a maternal mRNA cleavage [89-92]. In zebrafish, miRNAs participate in such maternal mRNA decay [93]. However, in mouse, miRNA activity seems to decrease after fertilization [94-96]. With the loss of activity of miRNAs, the partaking of tRFs in connection with maternal mRNA clearance could be crucial during the early stages of embryonic development. In fact, the overexpression of tRFs induces a decrease of both abundance and activity of the miRNAs [97-98].

Recently, the piRNA pathway has been also involved in mRNA decay [99-100]. In that sense, the piRNAs identified in the oocyte as potential regulators of male specific transcripts could be a safeguard system developed in the female germ cells to protect the early embryo of potential deleterious effect from male mRNAs. For instance, the putative presence in the zygote of functional *Ctcfl/Boris* mRNA, involved in the paternal imprinting of *H19* [68-69] could alter the maternal imprinting programme of the reciprocal imprinting locus *H19/Igf2* with aberrant developmental consequences. In *D. melanogaster* piRNAs expressed in early embryo can act over dual target types of RNAs (TEs and protein-coding RNAs) [100]. In general, the primary mechanisms of action of piRNAs are not an impediment to such potential duality. In fact, a crosstalk between different pathways of miRNAs, endo-siRNAs and piRNAs has been suggested [101]. In that sense, depletion of DICER and DROSHA proteins affects to rRNA processing and, in particular, 5.8s rRNA precursors were accumulated in cells depleted of such proteins [102]. Considering this crosstalk, it is tempting to speculate that these diverse types of small non-protein-coding RNAs could participate in shared functions.

Our analysis has evidenced that the expression patterns of both piRNAs and endo-siRNAs consisted of almost identical profiles in oocytes and zygotes which in turn resulted

radically different to the rasRNAs populations detected in spermatozoa. The preponderance in spermatozoa of piRNAs derived from non-TEs agrees with a recent analysis of piRNA identified after MIWI and MILI immunoprecipitation in prepuberal and adult testis of mice. Interestingly, the piRNA derived from non-TE elements significantly increased from prepuberal testis (absence of mature sperm cells) to adult testis (highly abundant in mature sperm cells) [103], hence suggesting that our data focused on the final stage of cell differentiation in spermatozoa which in turn could represent the final fate of these small RNAs during spermatogenesis. Nonetheless, our results do not support the total absence of ping-pong mechanisms during spermatogenesis, the deep sequence analysis of more than 200.000 piRNA sequences in sperm cells indicates a high level of molecules potentially involved in ping-pong mechanisms. Moreover, our data suggests that this amplification process could contemplate in addition to transposon mobility defense mechanisms also other possible piRNA-associated regulatory functions.

Specific endo-siRNA expression in female germ cells has been previously suggested [57, 75]. Many studies have so far identified different types of RNAs localized in sperm [104]. However, the presence of endo-siRNAs in male germ cells has not been previously described. As a very relevant fact, in this study we have identified endo-siRNAs in sperm corresponding to satellite DNA sequences, concretely GSAT (Fig. 2E and Supplemental Fig. S2). Such specificity suggests a potential role of these endo-siRNAs with regard to spermatogenesis due to the nature of satellite DNA in relation to the particular organization of sperm chromatin.

4.1. piRNA and endo-siRNA biogenesis.

The ping-pong mechanism has been established as an efficient mechanism to amplify piRNAs in *Drosophila* and silkworm [9, 105-106]. In this study, we have observed ping-pong signatures present in piRNA populations of female and male gametes as well as in zygotes (see Supplemental Fig. S3C and GEO deposited data). However, our results indicate that the ping-pong model-based amplification process could be considered as a wider mechanism of rasRNA

regulation which is not only restricted to defend genome integrity of germ and stem cells from TEs. Nevertheless, the starting point of primary piRNAs to generate secondary piRNAs could not be followed uniquely by the amplification process. Moreover, the existence of "degenerated pairing" (one primary piRNA sequence could potentially have several secondary piRNA partners), also points towards a potential amplification by multiple targets.

ACKNOWLEDMENTS

We thank M. Quesada and M. Moreno for technical assistance, and R. M. Fratini for revision of the manuscript. This work was funded by grants from The European Chemical Industry Council Long-range Research Initiative (CEFIC-LRi), from the MEDDTL (*11-MRES-PNRPE-9-CVS-072*), France, from the CSIC (*PIE 201020E016*), Spain. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interests.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

TABLES

Table 1. Total sequences and reads of RNA classes classified. The asterisks signalize the two rasRNA classes.

FIGURE LEGENDS

FIGURE 1. Expression of rasRNAs and their association with transposable (TE) and non-transposable (Non-TE) elements. Fractional distribution of rasRNAs tags detected in oocytes (A), spermatozoa (B) and zygotes (C) and their normalized expression levels.

FIGURE 2. Analysis of piRNA and endo-siRNA populations. The circular diagrams show the percentage of tags associated with different genomic elements. The bar diagrams show the read distribution from classified tags. Panels A, B and C show the data corresponding to piRNAs detected in oocytes, spermatozoa and zygotes, respectively. Panels D, E and F show the data corresponding to endo-siRNAs detected in oocytes, spermatozoa and zygotes, respectively. Data normalized using the DeSeq tool of the R/Bioconductor software package.

FIGURE 3. Analysis of piRNAs derived from tRNA transcript cleavage. A) The heat map represents the normalized expression reads of all piRNA tags associated with tRNA molecules. B) The bar diagrams represent the most abundant 3´ and 5´ ends identified from sequences associated with tRNAs.

FIGURE 4. piRNAs detected in oocytes, spermatozoa and zygotes. (A) Hierarchical clustering analysis of common piRNAs detected in all three cell types. The heat maps show the expression of shared piRNAs considering highly represented tags (≥ 700) of each cell type. The Venn diagram represents the shared and specific tags detected in each cell type. (B) Tag and read distribution for the 28,069 sequences that were detected as shared by all three cell types. (C and D) Expression analysis of piRNAs shared by female and male gametes and zygotes. The MA plots show a comparison of tag expression before and after fertilization. The horizontal axis represents the average expression (normalized reads) for the same tag in each sample, whereas the vertical axis represents the ratio of normalized reads for each tag between different cell types. Colored spots (red and green) represent tags that were up or downregulated at least a twofold after fertilization. (C) Common piRNAs detected in oocytes and zygotes. (D) Common piRNAs detected in spermatozoa and zygotes.

FIGURE 5. Common endo-siRNAs detected in oocytes, spermatozoa and zygotes. (A) Hierarchical cluster analysis. The Venn diagram represents the common and specific tags detected in each cell type. The heat maps show the comparison among the highly expressed tags (≥ 700 normalized reads) of each cell type. (B) Tag and read distribution for the 78,891

sequences that were detected in the three cell types. (C and D) Expression analysis of endo-siRNAs shared between female and male gametes and zygotes. (C) Common endo-siRNAs detected in oocytes and zygotes. (D) Common endo-siRNAs detected in spermatozoa and zygotes.

REFERENCES

[1] G.F. Richard, A. Kerrest, B. Dujon, Comparative genomics and molecular dynamics of DNA repeats in eukaryotes, Microbiol Mol Biol Rev, 72 (2008) 686-727.

[2] J. Jurka, V.V. Kapitonov, O. Kohany, M.V. Jurka, Repetitive sequences in complex genomes: structure and evolution, Annu Rev Genomics Hum Genet, 8 (2007) 241-259.

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

[4] L. Girard, M. Freeling, Regulatory changes as a consequence of transposon insertion, Dev Genet, 25 (1999) 291-296.

[5] R. Almeida, R.C. Allshire, RNA silencing and genome regulation, Trends Cell Biol, 15 (2005) 251-258.

[6] T.U. Banisch, M. Goudarzi, E. Raz, Small RNAs in Germ Cell Development, Curr Top Dev Biol, 99 (2012) 79-113.

[7] J.H. Yang, P. Shao, H. Zhou, Y.Q. Chen, L.H. Qu, deepBase: a database for deeply annotating and mining deep sequencing data, Nucleic Acids Res, 38 (2010) D123-130.

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

[9] S. Kawaoka, Y. Arai, K. Kadota, Y. Suzuki, K. Hara, S. Sugano, K. Shimizu, Y. Tomari, T. Shimada, S. Katsuma, Zygotic amplification of secondary piRNAs during silkworm embryogenesis, Rna, 17 (2011) 1401-1407.

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

[11] A.I. Kalmykova, M.S. Klenov, V.A. Gvozdev, Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline, Nucleic Acids Res, 33 (2005) 2052-2059.

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

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

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

[15] T. Thomson, H. Lin, The biogenesis and function of PIWI proteins and piRNAs: progress and prospect, Annu Rev Cell Dev Biol, 25 (2009) 355-376.

[16] J.J. Ipsaro, A.D. Haase, S.R. Knott, L. Joshua-Tor, G.J. Hannon, The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis, Nature, 491 (2012) 279-283.

[17] H. Nishimasu, H. Ishizu, K. Saito, S. Fukuhara, M.K. Kamatani, L. Bonnefond, N. Matsumoto, T. Nishizawa, K. Nakanaga, J. Aoki, R. Ishitani, H.

Siomi, M.C. Siomi, O. Nureki, Structure and function of Zucchini endoribonuclease in piRNA biogenesis, Nature, 491 (2012) 284-287.
[18] V.N. Kim, J. Han, M.C. Siomi, Biogenesis of small RNAs in animals, Nat Rev Mol Cell Biol, 10 (2009) 126-139.
[19] C. Klattenhoff, W. Theurkauf, Biogenesis and germline functions of piRNAs, Development, 135 (2008) 3-9.
[20] L.S. Gunawardane, K. Saito, K.M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, M.C. Siomi, A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila, Science, 315 (2007) 1587-1590.
[21] K.A. O'Donnell, J.D. Boeke, Mighty Piwis defend the germline against genome intruders, Cell, 129 (2007) 37-44.
[22] S. Houwing, E. Berezikov, R.F. Ketting, Zili is required for germ cell differentiation and meiosis in zebrafish, EMBO J, 27 (2008) 2702-2711.
[23] S. Kawaoka, N. Hayashi, S. Katsuma, H. Kishino, Y. Kohara, K. Mita, T. Shimada, Bombyx small RNAs: genomic defense system against transposons in the silkworm, Bombyx mori, Insect Biochem Mol Biol, 38 (2008) 1058-1065.
[24] S. Kuramochi-Miyagawa, 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, T. Nakano, DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes, Genes Dev, 22 (2008) 908-917.
[25] M. Reuter, P. Berninger, S. Chuma, H. Shah, M. Hosokawa, C. Funaya, C. Antony, R. Sachidanandam, R.S. Pillai, Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing, Nature, 480 (2011) 264-267.
[26] M. Otsuka, M. Zheng, M. Hayashi, J.D. Lee, O. Yoshino, S. Lin, J. Han, Impaired microRNA processing causes corpus luteum insufficiency and infertility in mice, J Clin Invest, 118 (2008) 1944-1954.
[27] L. Lei, S. Jin, G. Gonzalez, R.R. Behringer, T.K. Woodruff, The regulatory role of Dicer in folliculogenesis in mice, Mol Cell Endocrinol, 315 (2010) 63-73.
[28] L.J. Luense, M.Z. Carletti, L.K. Christenson, Role of Dicer in female fertility, Trends Endocrinol Metab, 20 (2009) 265-272.
[29] E.P. Murchison, P. Stein, Z. Xuan, H. Pan, M.Q. Zhang, R.M. Schultz, G.J. Hannon, Critical roles for Dicer in the female germline, Genes Dev, 21 (2007) 682-693.
[30] A.K. Nagaraja, C. Andreu-Vieyra, H.L. Franco, L. Ma, R. Chen, D.Y. Han, H. Zhu, J.E. Agno, P.H. Gunaratne, F.J. DeMayo, M.M. Matzuk, Deletion of Dicer in somatic cells of the female reproductive tract causes sterility, Mol Endocrinol, 22 (2008) 2336-2352.
[31] K. Okamura, A. Ishizuka, H. Siomi, M.C. Siomi, Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways, Genes Dev, 18 (2004) 1655-1666.
[32] A.S. Pepper, R.W. Beerman, B. Bhogal, T.A. Jongens, Argonaute2 suppresses Drosophila fragile X expression preventing neurogenesis and oogenesis defects, PLoS One, 4 (2009) e7618.
[33] X.H. Shen, Y.J. Han, X.S. Cui, N.H. Kim, Ago2 and GW182 expression in mouse preimplantation embryos: a link between microRNA biogenesis and GW182 protein synthesis, Reprod Fertil Dev, 22 (2010) 634-643.
[34] T. Watanabe, Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M.A. Surani, Y. Sakaki, H. Sasaki, Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes, Nature, 453 (2008) 539-543.
[35] D.E. Golden, V.R. Gerbasi, E.J. Sontheimer, An inside job for siRNAs, Mol Cell, 31 (2008) 309-312.
[36] E.M. Lucchetta, R.W. Carthew, R.F. Ismagilov, The endo-siRNA pathway is essential for robust development of the Drosophila embryo, PLoS One, 4 (2009) e7576.

[37] E. Bernstein, S.Y. Kim, M.A. Carmell, E.P. Murchison, H. Alcorn, M.Z. Li, A.A. Mills, S.J. Elledge, K.V. Anderson, G.J. Hannon, Dicer is essential for mouse development, Nat Genet, 35 (2003) 215-217.
[38] S. Morita, T. Horii, M. Kimura, Y. Goto, T. Ochiya, I. Hatada, One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation, Genomics, 89 (2007) 687-696.
[39] D. O'Carroll, I. Mecklenbrauker, P.P. Das, A. Santana, U. Koenig, A.J. Enright, E.A. Miska, A. Tarakhovsky, A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway, Genes Dev, 21 (2007) 1999-2004.
[40] B.R. Hogan B, Constantini F, Lacy E, Manipulating the Mouse Embryo: A Laboratory Manual, 2rd ed., Cold Spring Harbor Laboratory Press, 1994.
[41] A. Aravin, 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, T. Tuschl, A novel class of small RNAs bind to MILI protein in mouse testes, Nature, 442 (2006) 203-207.
[42] S.T. Grivna, E. Beyret, Z. Wang, H. Lin, A novel class of small RNAs in mouse spermatogenic cells, Genes Dev, 20 (2006) 1709-1714.
[43] N.C. Lau, A.G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D.P. Bartel, R.E. Kingston, Characterization of the piRNA complex from rat testes, Science, 313 (2006) 363-367.
[44] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol, 11 (2010) R106.
[45] V.M. Kvam, P. Liu, Y. Si, A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data, Am J Bot, (2012) 248-256.
[46] J.Q. Yao, F. Yu, DEB: A web interface for RNA-seq digital gene expression analysis, Bioinformation, 7 (2011) 44-45.
[47] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol, 10 (2009) R25.
[48] J. Garcia-Lopez, D. Hourcade Jde, J. del Mazo, Reprogramming of microRNAs by adenosine-to-inosine editing and the selective elimination of edited microRNA precursors in mouse oocytes and preimplantation embryos, Nucleic Acids Res, 41 (2013) 5483-5493.
[49] K. Nishikura, Functions and regulation of RNA editing by ADAR deaminases, Annu Rev Biochem, 79 (2010) 321-349.
[50] M. Kawano, H. Kawaji, V. Grandjean, J. Kiani, M. Rassoulzadegan, Novel small noncoding RNAs in mouse spermatozoa, zygotes and early embryos, PLoS One, 7 (2012) e44542.
[51] K.C. Pang, S. Stephen, M.E. Dinger, P.G. Engstrom, B. Lenhard, J.S. Mattick, RNAdb 2.0--an expanded database of mammalian non-coding RNAs, Nucleic Acids Res, 35 (2007) D178-182.
[52] G. Cecere, A. Grishok, A nuclear perspective on RNAi pathways in metazoans, Biochim Biophys Acta, (2013) doi: 10.1016/j.bbagrm.
[53] H. Ishizu, H. Siomi, M.C. Siomi, Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines, Genes Dev, 26 (2012) 2361-2373.
[54] A.A. Aravin, G.J. Hannon, Small RNA silencing pathways in germ and stem cells, Cold Spring Harb Symp Quant Biol, 73 (2008) 283-290.
[55] N. Zamudio, D. Bourc'his, Transposable elements in the mammalian germline: a comfortable niche or a deadly trap?, Heredity, 105 (2010) 92-104.
[56] S. Kadri, V.F. Hinman, P.V. Benos, RNA Deep Sequencing Reveals Differential MicroRNA Expression during Development of Sea Urchin and Sea Star, PLoS One, 6 (2011) e29217.
[57] O.H. Tam, A.A. Aravin, P. Stein, A. Girard, E.P. Murchison, S. Cheloufi, E. Hodges, M. Anger, R. Sachidanandam, R.M. Schultz, G.J. Hannon, Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes, Nature, 453 (2008) 534-538.

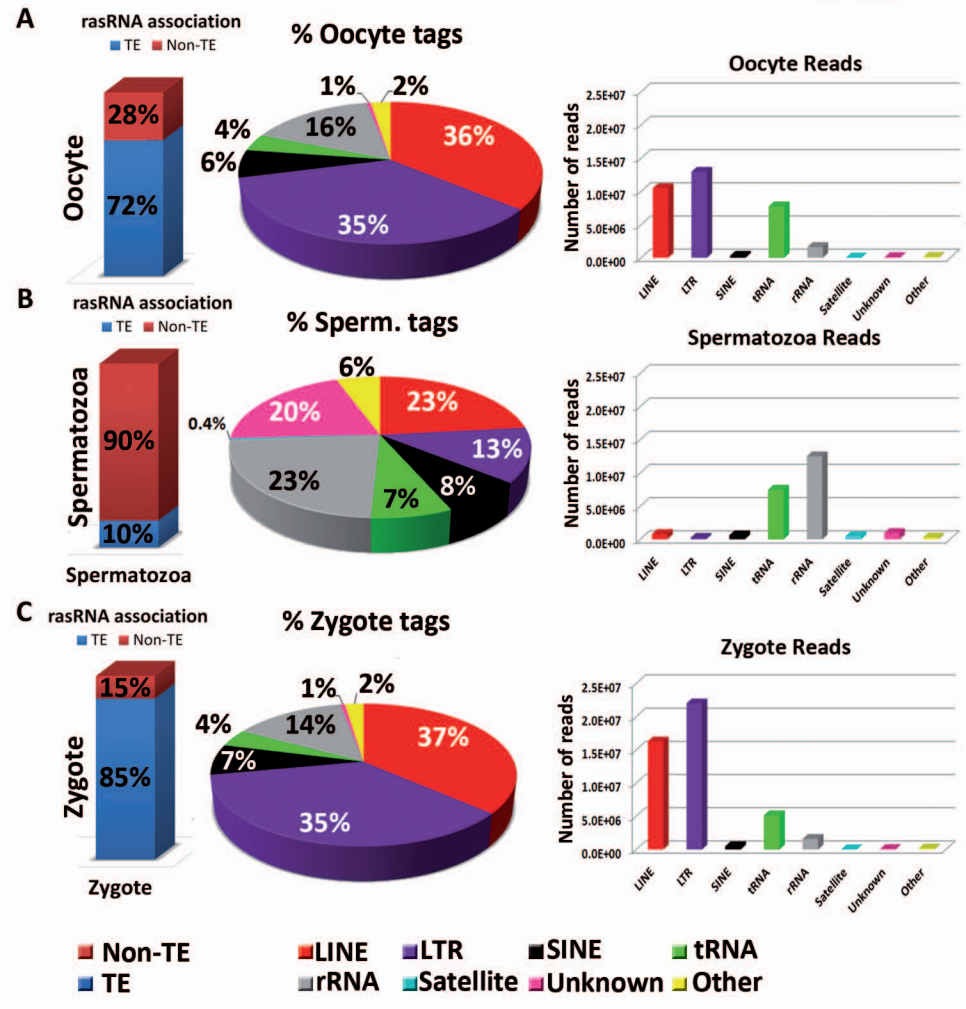
[58] Y.S. Lee, Y. Shibata, A. Malhotra, A. Dutta, A novel class of small RNAs: tRNA-derived RNA fragments (tRFs), Genes Dev, 23 (2009) 2639-2649.
[59] Z. Li, C. Ender, G. Meister, P.S. Moore, Y. Chang, B. John, Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs, Nucleic Acids Res, 40 (2012) 6787-6799.
[60] D.M. Thompson, C. Lu, P.J. Green, R. Parker, tRNA cleavage is a conserved response to oxidative stress in eukaryotes, Rna, 14 (2008) 2095-2103.
[61] T. Pederson, Regulatory RNAs derived from transfer RNA?, RNA, 16 (2010) 1865-1869.
[62] Y. Ohnishi, Y. Totoki, A. Toyoda, T. Watanabe, Y. Yamamoto, K. Tokunaga, Y. Sakaki, H. Sasaki, H. Hohjoh, Active role of small non-coding RNAs derived from SINE/B1 retrotransposon during early mouse development, Mol Biol Rep, 39 (2012) 903-909.
[63] A.A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, G.J. Hannon, Developmentally regulated piRNA clusters implicate MILI in transposon control, Science, 316 (2007) 744-747.
[64] T. Esposito, S. Magliocca, D. Formicola, F. Gianfrancesco, piR_015520 belongs to Piwi-associated RNAs regulates expression of the human melatonin receptor 1A gene, PLoS One, 6 (2011) e22727.
[65] P. Rajasethupathy, I. Antonov, R. Sheridan, S. Frey, C. Sander, T. Tuschl, E.R. Kandel, A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity, Cell, 149 (2012) 693-707.
[66] J. Brennecke, A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G.J. Hannon, Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila, Cell, 128 (2007) 1089-1103.
[67] A.A. Aravin, D. Bourc'his, Small RNA guides for de novo DNA methylation in mammalian germ cells, Genes Dev, 22 (2008) 970-975.
[68] P. Jelinic, J.C. Stehle, P. Shaw, The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation, PLoS Biol, 4 (2006) e355.
[69] P. Nguyen, H. Cui, K.S. Bisht, L. Sun, K. Patel, R.S. Lee, H. Kugoh, M. Oshimura, A.P. Feinberg, D. Gius, CTCFL/BORIS is a methylation-independent DNA-binding protein that preferentially binds to the paternal H19 differentially methylated region, Cancer Res, 68 (2008) 5546-5551.
[70] R.J. Frost, F.K. Hamra, J.A. Richardson, X. Qi, R. Bassel-Duby, E.N. Olson, MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs, Proc Natl Acad Sci U S A, 107 (2010) 11847-11852.
[71] Z. He, M. Kokkinaki, D. Pant, G.I. Gallicano, M. Dym, Small RNA molecules in the regulation of spermatogenesis, Reproduction, 137 (2009) 901-911.
[72] K. Ichiyanagi, Y. Li, T. Watanabe, T. Ichiyanagi, K. Fukuda, J. Kitayama, Y. Yamamoto, S. Kuramochi-Miyagawa, T. Nakano, Y. Yabuta, Y. Seki, M. Saitou, H. Sasaki, Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development, Genome Res, 21 (2011) 2058-2066.
[73] N.C. Lau, T. Ohsumi, M. Borowsky, R.E. Kingston, M.D. Blower, Systematic and single cell analysis of Xenopus Piwi-interacting RNAs and Xiwi, EMBO J, 28 (2009) 2945-2958.
[74] S.E. Linsen, E. de Wit, E. de Bruijn, E. Cuppen, Small RNA expression and strain specificity in the rat, BMC Genomics, 11 (2010) 249.
[75] T. Watanabe, H. Imai, N. Minami, Identification and expression analysis of small RNAs during development, Methods Mol Biol, 442 (2008) 173-185.
[76] R.M. Schultz, Regulation of zygotic gene activation in the mouse, Bioessays, 15 (1993) 531-538.
[77] M.S. Ko, J.R. Kitchen, X. Wang, T.A. Threat, A. Hasegawa, T. Sun, M.J. Grahovac, G.J. Kargul, M.K. Lim, Y. Cui, Y. Sano, T. Tanaka, Y. Liang, S. Mason, P.D. Paonessa, A.D. Sauls, G.E. DePalma, R. Sharara, L.B. Rowe, J. Eppig, C. Morrell, H. Doi, Large-scale cDNA analysis reveals phased gene expression

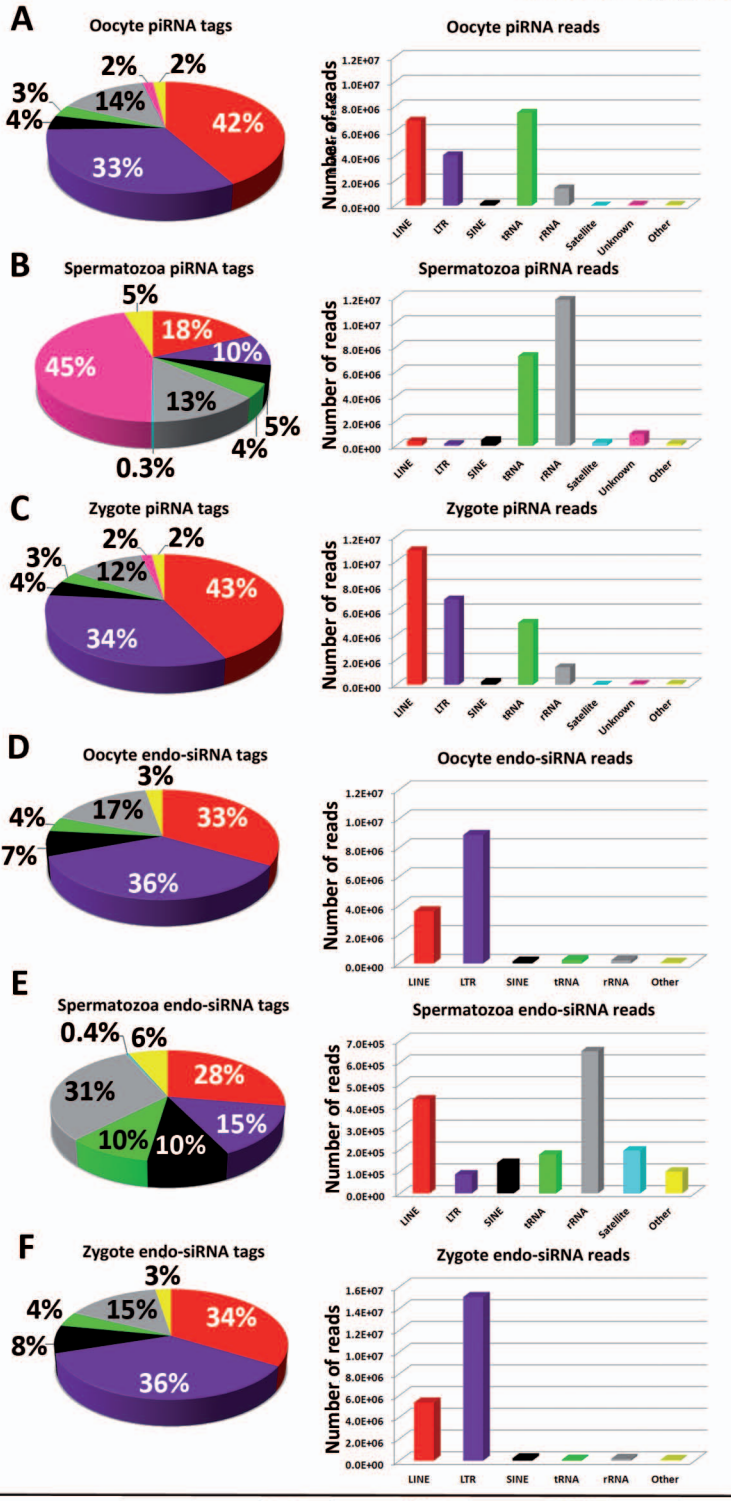
patterns during preimplantation mouse development, Development, 127 (2000) 1737-1749.
[78] J. Ma, P. Svoboda, R.M. Schultz, P. Stein, Regulation of zygotic gene activation in the preimplantation mouse embryo: global activation and repression of gene expression, Biol Reprod, 64 (2001) 1713-1721.
[79] S. Zhao, L.T. Gou, M. Zhang, L.D. Zu, M.M. Hua, Y. Hua, H.J. Shi, Y. Li, J. Li, D. Li, E.D. Wang, M.F. Liu, piRNA-triggered MIWI ubiquitination and removal by APC/C in late spermatogenesis, Dev Cell, 24 (2013) 13-25.
[80] C.Y. Chu, T.M. Rana, Small RNAs: regulators and guardians of the genome, J Cell Physiol, 213 (2007) 412-419.
[81] J.V. Hartig, Y. Tomari, K. Forstemann, piRNAs--the ancient hunters of genome invaders, Genes Dev, 21 (2007) 1707-1713.
[82] J. Hock, G. Meister, The Argonaute protein family, Genome Biol, 9 (2008) 210.
[83] A. Girard, R. Sachidanandam, G.J. Hannon, M.A. Carmell, A germline-specific class of small RNAs binds mammalian Piwi proteins, Nature, 442 (2006) 199-202.
[84] V.N. Kim, Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes, Genes Dev, 20 (2006) 1993-1997.
[85] H. Kawaji, M. Nakamura, Y. Takahashi, A. Sandelin, S. Katayama, S. Fukuda, C.O. Daub, C. Kai, J. Kawai, J. Yasuda, P. Carninci, Y. Hayashizaki, Hidden layers of human small RNAs, BMC Genomics, 9 (2008) 157.
[86] D.M. Thompson, R. Parker, Stressing out over tRNA cleavage, Cell, 138 (2009) 215-219.
[87] C.A. Pessot, M. Brito, J. Figueroa, Concha, II, A. Yanez, L.O. Burzio, Presence of RNA in the sperm nucleus, Biochem Biophys Res Commun, 158 (1989) 272-278.
[88] S. Abou Elela, R.N. Nazar, Role of the 5.8S rRNA in ribosome translocation, Nucleic Acids Res, 25 (1997) 1788-1794.
[89] L. Li, P. Zheng, J. Dean, Maternal control of early mouse development, Development, 137 (2010) 859-870.
[90] J.Y. Nothias, S. Majumder, K.J. Kaneko, M.L. DePamphilis, Regulation of gene expression at the beginning of mammalian development, J Biol Chem, 270 (1995) 22077-22080.
[91] A.F. Schier, The maternal-zygotic transition: death and birth of RNAs, Science, 316 (2007) 406-407.
[92] C. Wrenzycki, D. Herrmann, H. Niemann, Messenger RNA in oocytes and embryos in relation to embryo viability, Theriogenology, 68 Suppl 1 (2007) S77-83.
[93] A.J. Giraldez, Y. Mishima, J. Rihel, R.J. Grocock, S. Van Dongen, K. Inoue, A.J. Enright, A.F. Schier, Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs, Science, 312 (2006) 75-79.
[94] J. Garcia-Lopez, J. del Mazo, Expression dynamics of microRNA biogenesis during preimplantation mouse development, Biochim Biophys Acta, 1819 (2012) 847-854.
[95] J. Ma, M. Flemr, P. Stein, P. Berninger, R. Malik, M. Zavolan, P. Svoboda, R.M. Schultz, MicroRNA activity is suppressed in mouse oocytes, Curr Biol, 20 (2010) 265-270.
[96] N. Suh, L. Baehner, F. Moltzahn, C. Melton, A. Shenoy, J. Chen, R. Blelloch, MicroRNA function is globally suppressed in mouse oocytes and early embryos, Curr Biol, 20 (2010) 271-277.
[97] D. Haussecker, Y. Huang, A. Lau, P. Parameswaran, A.Z. Fire, M.A. Kay, Human tRNA-derived small RNAs in the global regulation of RNA silencing, Rna, 16 (2010) 673-695.
[98] A. Sobala, G. Hutvagner, Transfer RNA-derived fragments: origins, processing, and functions, Wiley Interdiscip Rev RNA, 2 (2011) 853-862.
[99] B. Barckmann, M. Simonelig, Control of maternal mRNA stability in germ cells and early embryos, Biochim Biophys Acta, 1829 (2013) 714-724.

[100] C. Rouget, C. Papin, A. Boureux, A.C. Meunier, B. Franco, N. Robine, E.C. Lai, A. Pelisson, M. Simonelig, Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo, Nature, 467 (2010) 1128-1132.
[101] M. Ghildiyal, P.D. Zamore, Small silencing RNAs: an expanding universe, Nat Rev Genet, 10 (2009) 94-108.
[102] X.H. Liang, S.T. Crooke, Depletion of key protein components of the RISC pathway impairs pre-ribosomal RNA processing, Nucleic Acids Res, 39 (2011) 4875-4889.
[103] E. Beyret, N. Liu, H. Lin, piRNA biogenesis during adult spermatogenesis in mice is independent of the ping-pong mechanism, Cell Res, 22 (2012) 1429-1439.
[104] G.D. Johnson, C. Lalancette, A.K. Linnemann, F. Leduc, G. Boissonneault, S.A. Krawetz, The sperm nucleus: chromatin, RNA, and the nuclear matrix, Reproduction, 141 (2011) 21-36.
[105] A. Nagao, T. Mituyama, H. Huang, D. Chen, M.C. Siomi, H. Siomi, Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis, RNA, 16 (2010) 2503-2515.
[106] J.S. Tushir, P.D. Zamore, Z. Zhang, SnapShot: Fly piRNAs, PIWI proteins, and the ping-pong cycle, Cell, 139 (2009) 634, e631.

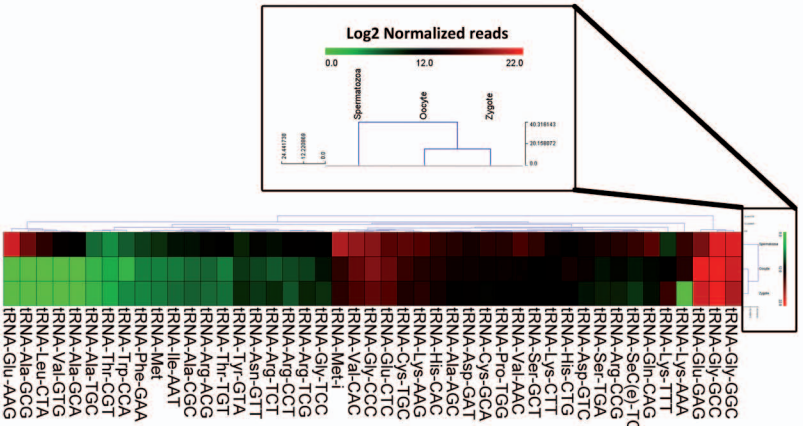
Garcia-Lopez_Table1

	SPERMATOZOA				OOCYTE				ZYGOTE			
	Total sequences		Total Reads		Total sequences		Total Reads		Total sequences		Total Reads	
miRNAs	25,028	**(1.1%)**	1,441,261	**(4.2%)**	21,012	**(1.4%)**	2,424,083	**(4.5%)**	23,907	**(1.4%)**	2,839,459	**(5.4%)**
piRNAs*	208,060	**(8.9%)**	18,728,064	**(54.9%)**	356,403	**(24.3%)**	24,756,486	**(46.2%)**	385,371	**(23.0%)**	22,075,046	**(41.6%)**
endo-siRNA*	71,780	**(3.1%)**	1,555,582	**(4.6%)**	360,213	**(24.6%)**	16,074,372	**(30.0%)**	426,844	**(25.5%)**	18,753,351	**(35.3%)**
Other ncRNAs	132,654	**(5.7%)**	3,149,552	**(9.2%)**	83,458	**(5.7%)**	2,983,596	**(5.6%)**	94,270	**(5.6%)**	2,781,619	**(5.2%)**
mRNA	384,313	**(16.5%)**	1,685,828	**(4.9%)**	232,309	**(15.8%)**	2,000,540	**(3.7%)**	274,651	**(16.4%)**	2,284,614	**(4.3%)**
Unannotated	1,503,665	**(64.7%)**	7,551,131	**(22.1%)**	412,766	**(28.2%)**	5,377,349	**(10.0%)**	466,901	**(27.9%)**	4,330,899	**(8.2%)**
TOTAL	2,325,500	**(100%)**	34,111,418	**(100%)**	1,466,161	**(100%)**	53,616,426	**(100%)**	1,671,944	**(100%)**	53,064,988	**(100%)**

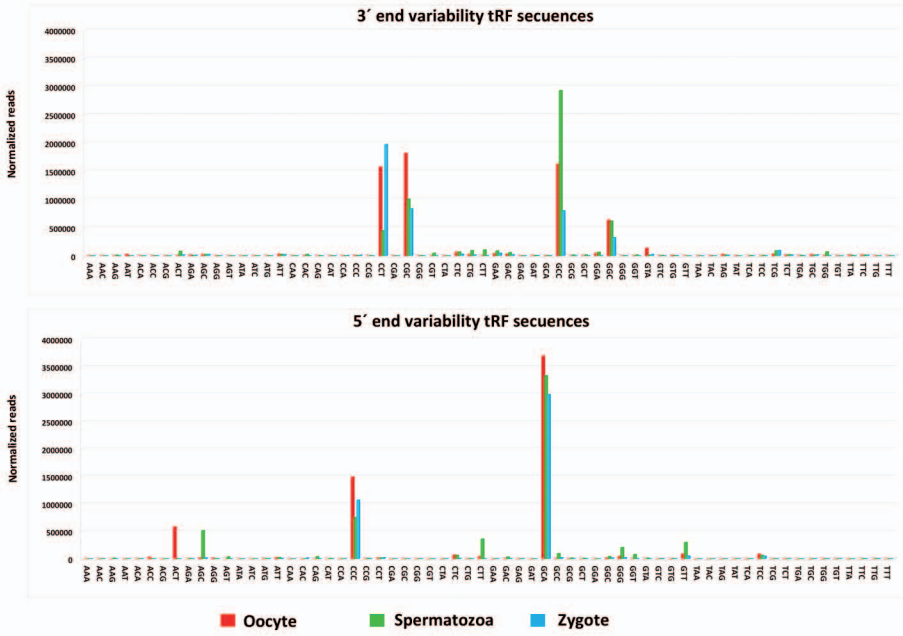

A
rasRNA association
TE
Non-TE
Oocyte
28%
72%
% Oocyte tags
1%
2%
4%
16%
6%
36%
35%
Oocyte Reads
Number of reads
2.5E+07
2.0E+07
1.5E+07
1.0E+07
5.0E+06
0.0E+00
LINE
LTR
SINE
tRNA
rRNA
Satellite
Unknown
Other
B
rasRNA association
TE
Non-TE
Spermatozoa
90%
10%
Spermatozoa
% Sperm. tags
6%
20%
23%
0.4%
13%
23%
8%
7%
Spermatozoa Reads
Number of reads
2.5E+07
2.0E+07
1.5E+07
1.0E+07
5.0E+06
0.0E+00
LINE
LTR
SINE
tRNA
rRNA
Satellite
Unknown
Other
C
rasRNA association
TE
Non-TE
Zygote
15%
85%
Zygote
% Zygote tags
1%
2%
4%
14%
7%
37%
35%
Zygote Reads
Number of reads
2.5E+07
2.0E+07
1.5E+07
1.0E+07
5.0E+06
0.0E+00
LINE
LTR
SINE
tRNA
rRNA
Satellite
Unknown
Other
Non-TE
TE
LINE
LTR
SINE
tRNA
rRNA
Satellite
Unknown
Other

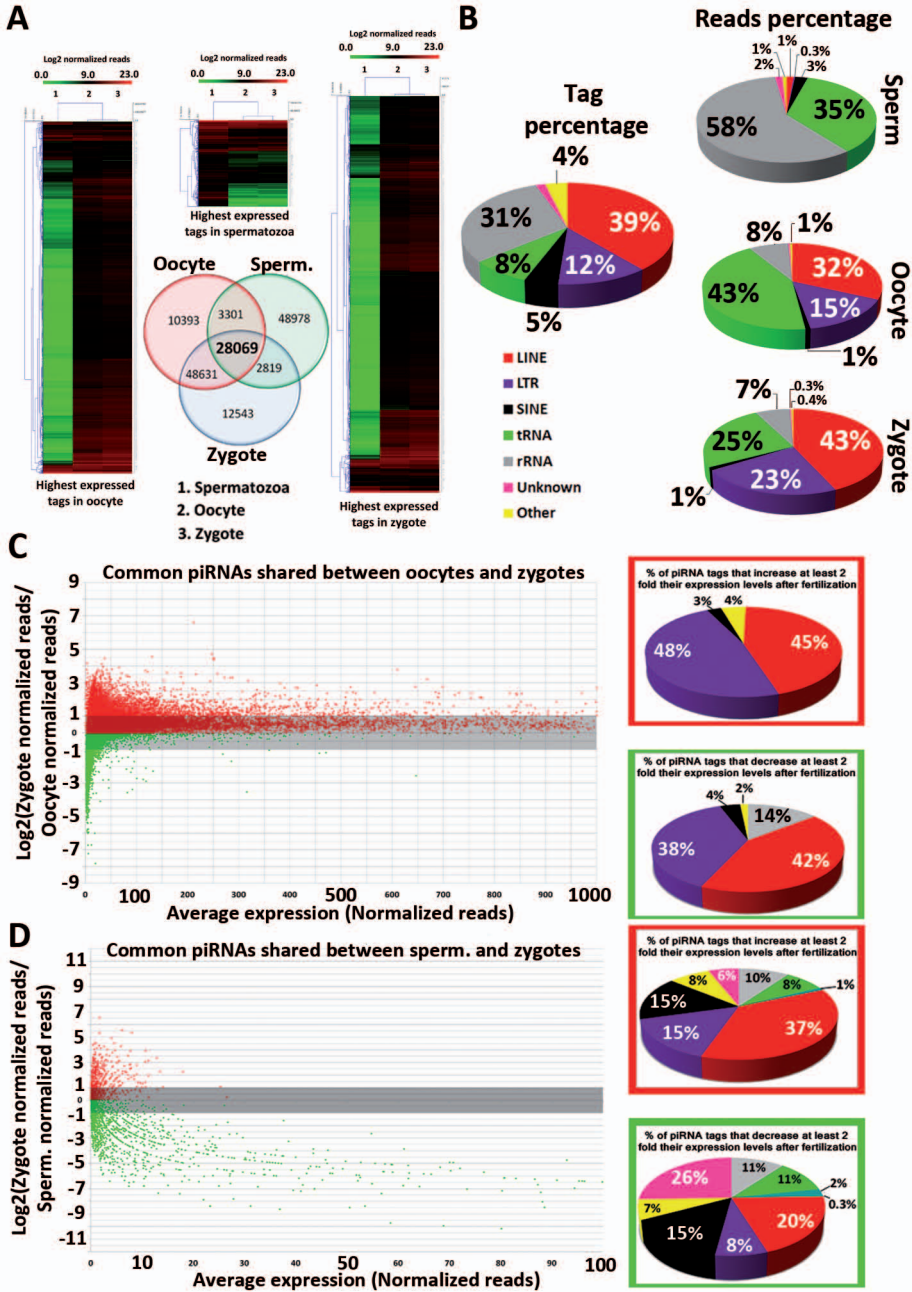

A
Oocyte piRNA tags
42%
33%
4%
3%
14%
2%
2%
Oocyte piRNA reads
Number of reads
1.2E+07
1.0E+07
8.0E+06
6.0E+06
4.0E+06
2.0E+06
0.0E+00
LINE
LTR
SINE
tRNA
rRNA
Satellite
Unknown
Other
B
Spermatozoa piRNA tags
18%
10%
5%
4%
0.3%
45%
13%
5%
Spermatozoa piRNA reads
C
Zygote piRNA tags
43%
34%
4%
3%
12%
2%
2%
Zygote piRNA reads
D
Oocyte endo-siRNA tags
33%
36%
7%
4%
17%
3%
Oocyte endo-siRNA reads
E
Spermatozoa endo-siRNA tags
28%
15%
10%
10%
31%
0.4%
6%
Spermatozoa endo-siRNA reads
7.0E+05
6.0E+05
5.0E+05
4.0E+05
3.0E+05
2.0E+05
1.0E+05
F
Zygote endo-siRNA tags
34%
36%
8%
4%
15%
3%
Zygote endo-siRNA reads
1.6E+07
1.4E+07
1.2E+07
1.0E+07
8.0E+06
6.0E+06
4.0E+06
2.0E+06
LINE
LTR
SINE
tRNA
rRNA
Satellite
Unknown
Other

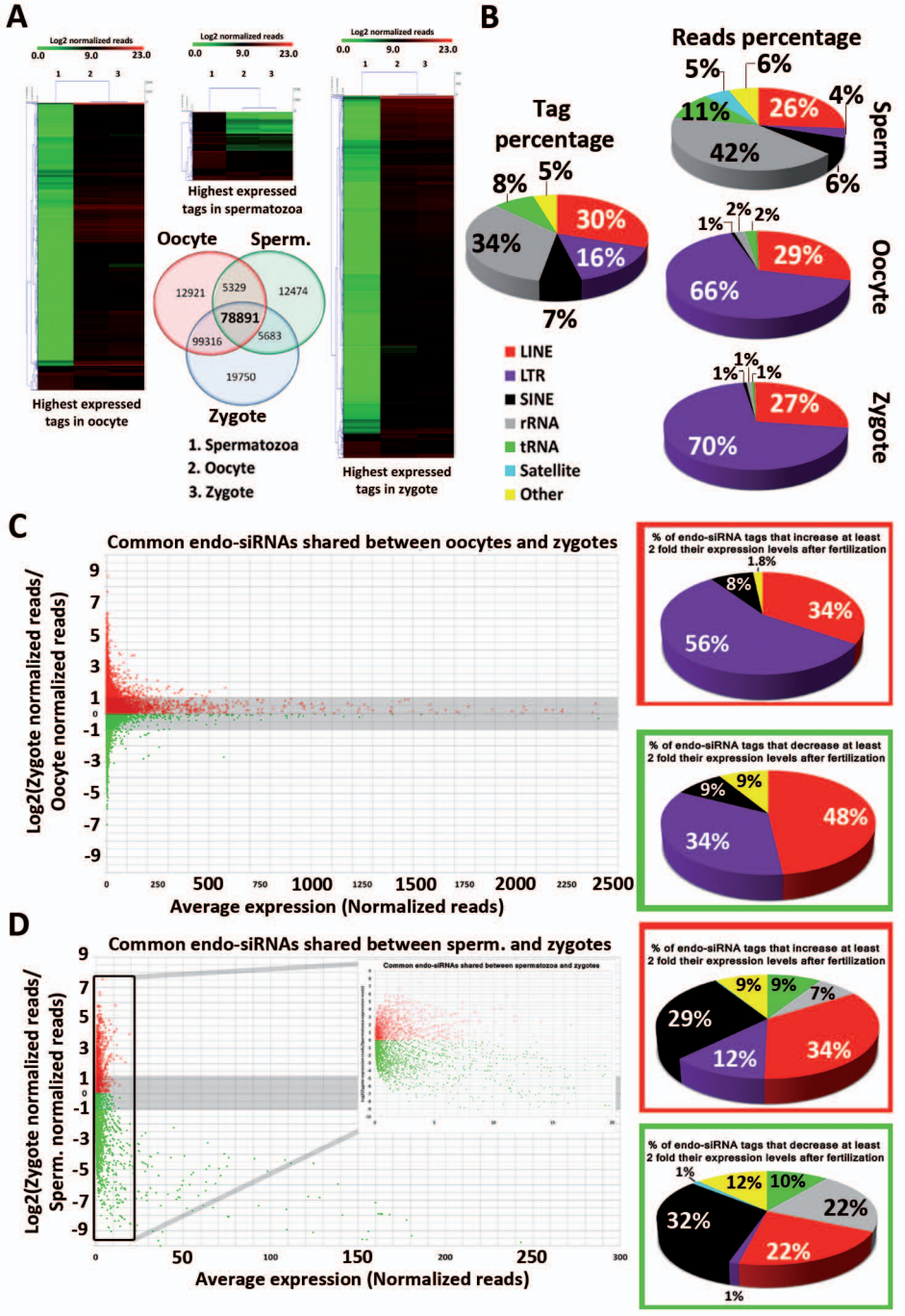
A


Log2 Normalized reads
0.0
12.0
22.0
Spermatozoa
Oocyte
Zygote
tRNA-Glu-AAG
tRNA-Ala-AGC
tRNA-Leu-CTA
tRNA-Val-GTG
tRNA-Ala-TGC
tRNA-Thr-CGT
tRNA-Trp-CCA
tRNA-Phe-GAA
tRNA-Met
tRNA-Ile-AAT
tRNA-Ala-CGC
tRNA-Arg-ACG
tRNA-Thr-TGT
tRNA-Tyr-GTA
tRNA-Asn-GTT
tRNA-Arg-CCT
tRNA-Arg-TCG
tRNA-Gly-TCC
tRNA-Met-i
tRNA-Val-CAC
tRNA-Gly-CCC
tRNA-Glu-CTC
tRNA-Cys-TGC
tRNA-Lys-AAG
tRNA-His-CAC
tRNA-Ala-AGC
tRNA-Asp-GAT
tRNA-Cys-GCA
tRNA-Pro-TGG
tRNA-Val-AAC
tRNA-Ser-GCT
tRNA-Lys-CTT
tRNA-His-CTG
tRNA-Asp-GTC
tRNA-Ser-TGA
tRNA-Arg-CCG
tRNA-SeC(e)-TCA
tRNA-Gln-CAG
tRNA-Lys-TTT
tRNA-Lys-AAA
tRNA-Glu-GAG
tRNA-Gly-GCC
tRNA-Gly-GGC

B


3´ end variability tRF secuences
5´ end variability tRF secuences
Normalized reads
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
AAA AAC AAG AAT ACA ACC ACG ACT AGA AGC AGG AGT ATA ATC ATG ATT CAA CAC CAG CAT CCA CCC CCG CCT CGA CGC CGG CGT CTA CTC CTG CTT GAA GAC GAG GAT GCA GCC GCG GCT GGA GGC GGG GGT GTA GTC GTG GTT TAA TAC TAG TAT TCA TCC TCG TCT TGA TGC TGG TGT TTA TTC TTG TTT
Oocyte
Spermatozoa
Zygote


A
Log2 normalized reads
0.0 9.0 23.0
1 2 3
Highest expressed tags in oocyte
Highest expressed tags in spermatozoa
Oocyte
Sperm.
Zygote
10393
3301
48978
28069
48631
2819
12543
1. Spermatozoa
2. Oocyte
3. Zygote
Highest expressed tags in zygote
B
Tag percentage
4%
31%
39%
8%
12%
5%
LINE
LTR
SINE
tRNA
rRNA
Unknown
Other
Reads percentage
1%
1%
2%
0.3%
3%
35%
58%
Sperm
8%
1%
32%
43%
15%
1%
Oocyte
7%
0.3%
0.4%
25%
43%
23%
1%
Zygote
C
Common piRNAs shared between oocytes and zygotes
Log2(Zygote normalized reads/ Oocyte normalized reads)
Average expression (Normalized reads)
% of piRNA tags that increase at least 2 fold their expression levels after fertilization
3%
4%
48%
45%
% of piRNA tags that decrease at least 2 fold their expression levels after fertilization
4%
2%
14%
38%
42%
D
Common piRNAs shared between sperm. and zygotes
Log2(Zygote normalized reads/ Sperm. normalized reads)
Average expression (Normalized reads)
% of piRNA tags that increase at least 2 fold their expression levels after fertilization
8%
6%
10%
8%
1%
15%
15%
37%
% of piRNA tags that decrease at least 2 fold their expression levels after fertilization
26%
11%
11%
2%
0.3%
7%
15%
8%
20%


A
Log2 normalized reads
0.0 9.0 23.0
1 2 3
Highest expressed tags in oocyte
Highest expressed tags in spermatozoa
Oocyte
Sperm.
12921
5329
12474
78891
99316
5683
19750
Zygote
1. Spermatozoa
2. Oocyte
3. Zygote
Highest expressed tags in zygote
B
Tag percentage
8%
5%
30%
34%
16%
7%
LINE
LTR
SINE
rRNA
tRNA
Satellite
Other
Reads percentage
5%
6%
11%
26%
4%
42%
6%
Sperm
1%
2%
2%
29%
66%
Oocyte
1%
1%
1%
27%
70%
Zygote
C
Common endo-siRNAs shared between oocytes and zygotes
Log2(Zygote normalized reads/ Oocyte normalized reads)
Average expression (Normalized reads)
% of endo-siRNA tags that increase at least 2 fold their expression levels after fertilization
1.8%
8%
34%
56%
% of endo-siRNA tags that decrease at least 2 fold their expression levels after fertilization
9%
9%
48%
34%
D
Common endo-siRNAs shared between sperm. and zygotes
Log2(Zygote normalized reads/ Sperm. normalized reads)
Average expression (Normalized reads)
% of endo-siRNA tags that increase at least 2 fold their expression levels after fertilization
9%
9%
7%
29%
12%
34%
% of endo-siRNA tags that decrease at least 2 fold their expression levels after fertilization
1%
12%
10%
22%
32%
22%
1%

SUPPLEMENTAL MATERIAL

FIGURE S1. Read length distribution. The figure shows the distribution of read length in each sample.

FIGURE S2. Specific rasRNAs detected in gametes and zygotes. The figure shows the specific piRNAs and endo-siRNAs detected in spermatozoa, oocytes and zygotes.

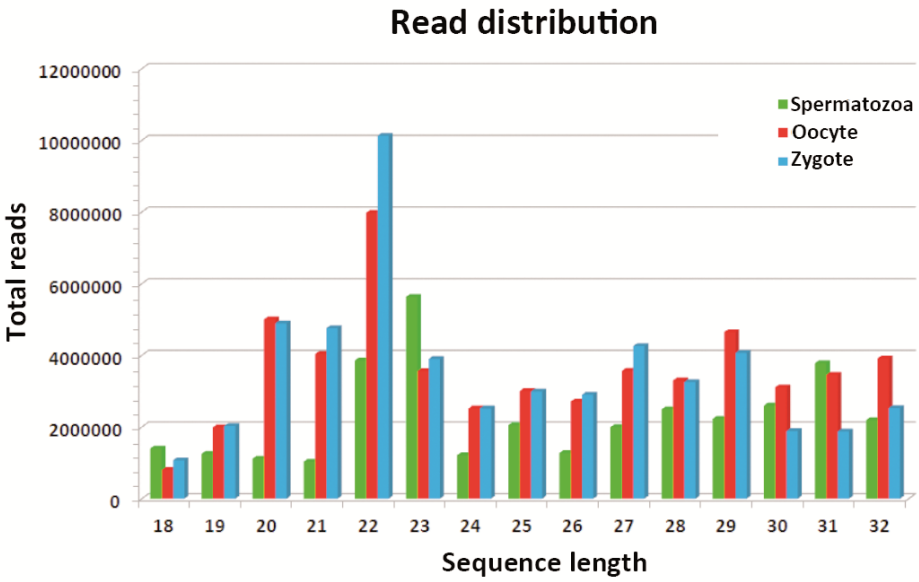
FIGURE S3. Sequence level analysis of piRNAs. The figure shows the distribution of piRNA classes in each cell type and the ping-pong rates for each sample. A) The bar diagram shows for each cell type the number of piRNA sequences classified as primary, secondary, 5U10A and other. B) Four examples are depicted of the equivalent types of piRNAs detected. C) Inter- and intra-sample rates of ping-pong pairs detected.

FIGURE S4. **Analysis of piRNAs classified as other.** The figure shows the read and tag distribution of piRNAs sequences that lack the ping-pong signatures. (A) piRNAs classified as other in spermatozoa. (B) piRNAs classified as other in oocytes. (C) piRNAs classified as other in zygotes.

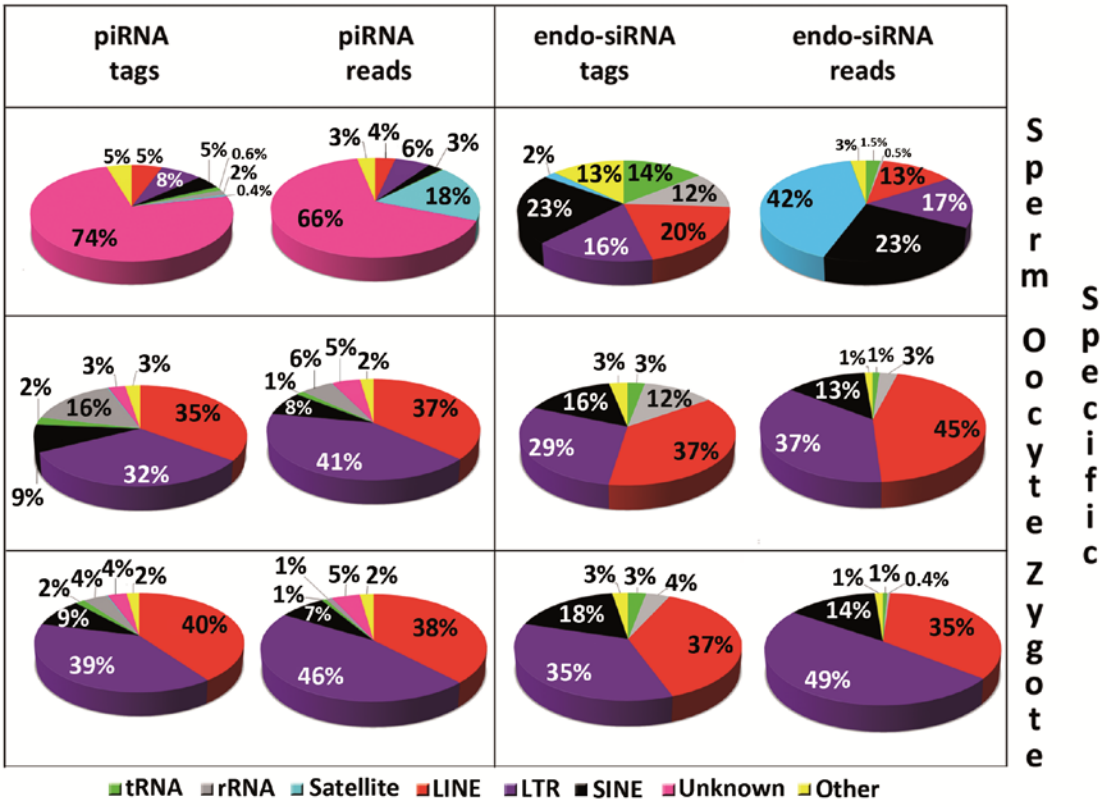
FIGURE S5. **Pipeline.** The figure shows the pipeline used to small RNA classification.

FIGURE S6. **Distribution of mRNAs potentially targeted by piRNAs.** Venn diagram shows the number of shared and specific protein-coding genes that are potential targets of piRNAs detected.

Garcia-Lopez_Figure S1

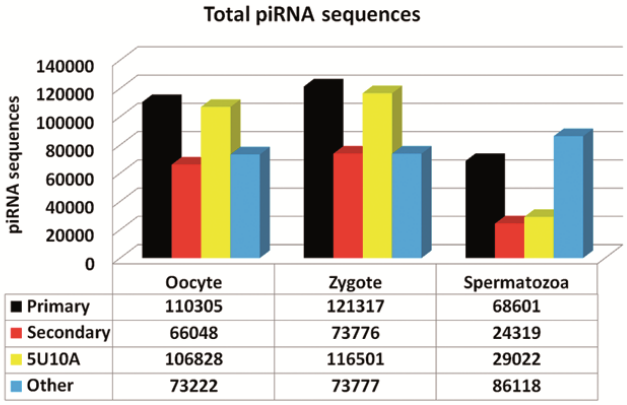


Garcia-Lopez_Figure S2



Garcia-Lopez_Figure S3

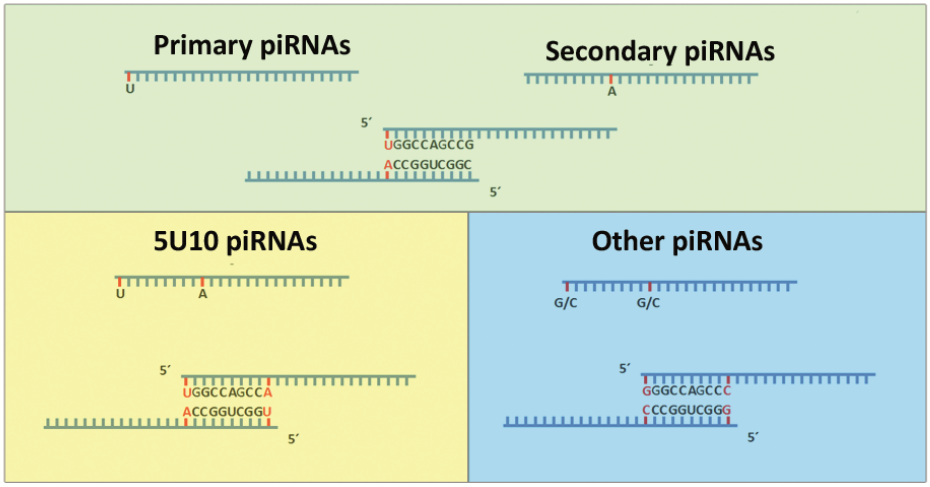
A



	Oocyte	Zygote	Spermatozoa
Primary	110305	121317	68601
Secondary	66048	73776	24319
5U10A	106828	116501	29022
Other	73222	73777	86118

B

Types of piRNA sequences and ping-pong pairing

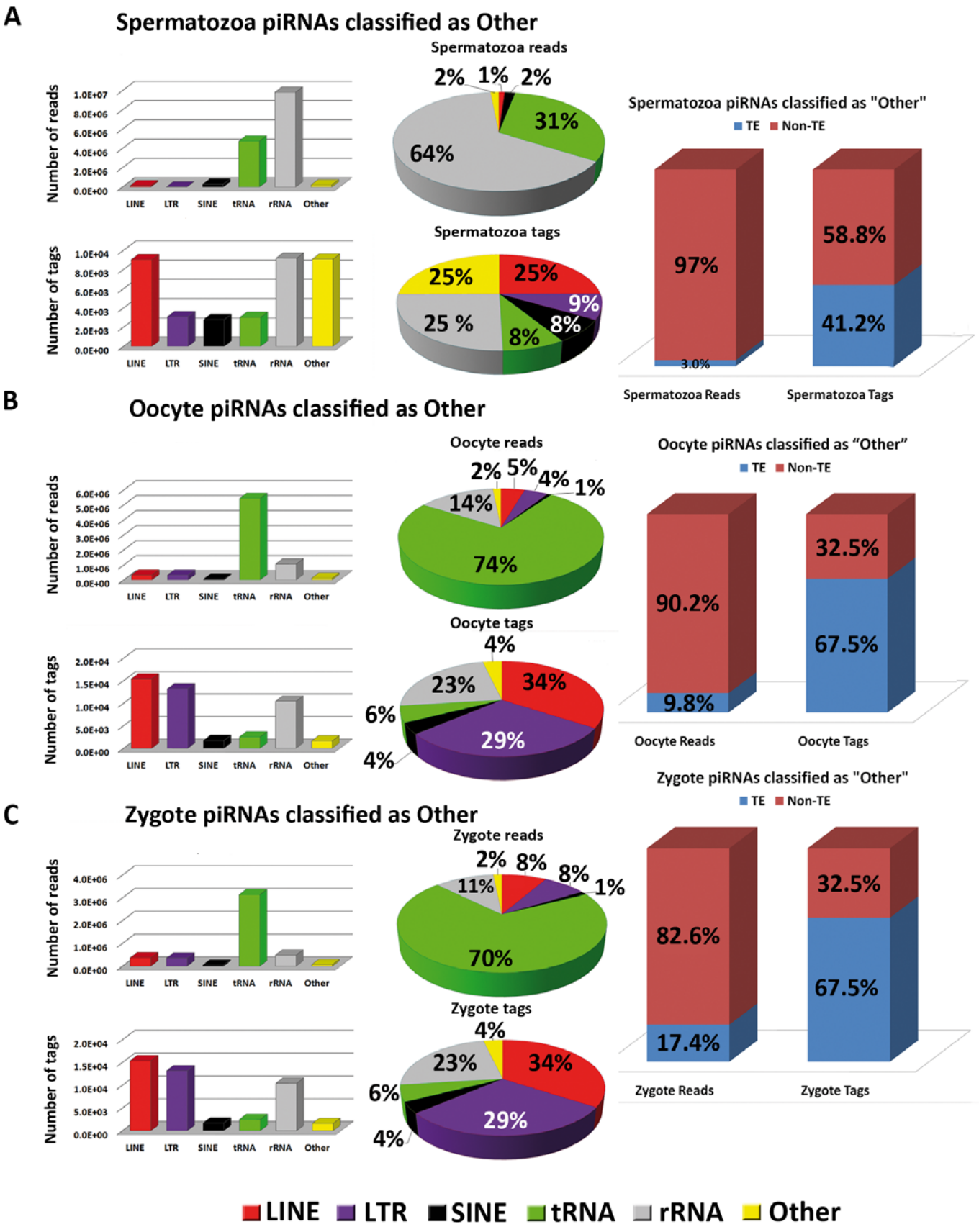


C

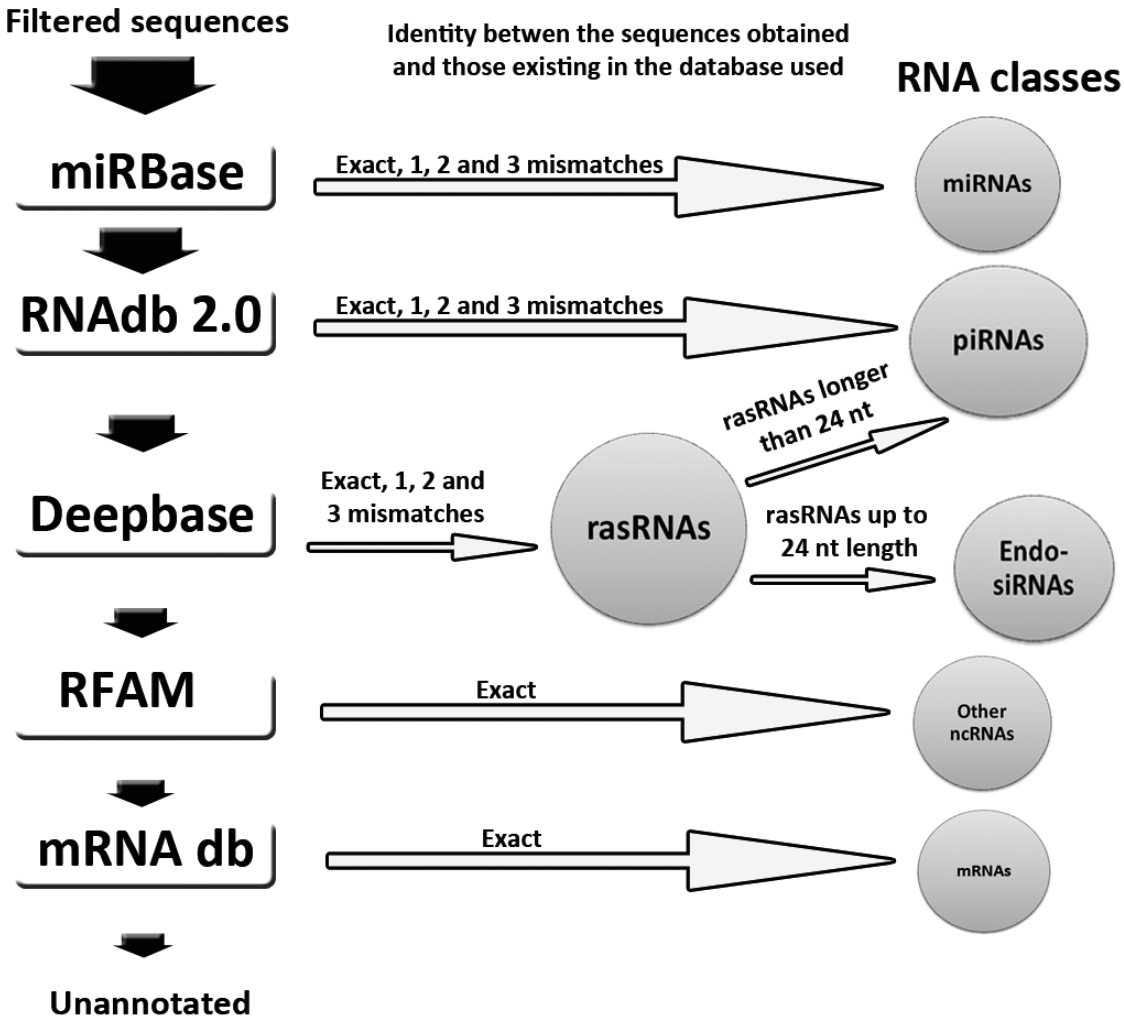
percentage of sequences that participate in ping-pong pairs

	Oocyte	Zygote	Spermatozoa
Percentage of primary piRNA sequences that have secondary piRNA partner	48.1%	49.4%	11.4%
Percentage of secondary piRNA sequences that have primary piRNA partner	65%	67%	33.8%
Percentage of 5U10A sequences that have ping-pong partner	74.4%	74.4%	29.1%
Percentage of other sequences that have ping-pong partner	23%	25.5%	28.5%

Garcia-Lopez_Figure S4



Garcia-Lopez_Figure S5



Garcia-Lopez_Figure_S6