# Small RNA pathways in platypus

Elizabeth P. Murchison[1], Pouya Kheradpour[2], Ravi Sachidanandam[1], Carly Smith[1], Zhenyu Xuan[1], Manolis Kellis[2,3], Frank Grűtzner[4], Alexander Stark[2,3], and Gregory J. Hannon[1*]

[1]Watson School of Biological Sciences
Howard Hughes Medical Institute
Cold Spring Harbor Laboratory
1 Bungtown Road
Cold Spring Harbor, NY 11724, USA

[2]Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

[3]Broad Institute of MIT and Harvard
Cambridge, MA 02141, USA

[4]Discipline of Genetics
School of Molecular & Biomedical Science
The University of Adelaide
5005 SA, Australia

*To whom correspondence should be addressed

Keywords: RNAi; microRNA; piRNA; platypus; echidna; monotreme

## Abstract

**Small RNA pathways play evolutionarily conserved roles in gene regulation and in defense from pathogenic and parasitic nucleic acids. The character and expression patterns of small RNAs show conservation throughout animal lineages, but specific animal clades also show variations on these recurring themes, including species-specific small RNAs. The monotremes, with only platypus and four species of echidna as extant members, represent the basal branch of the mammalian lineage. Here, we examine the small RNA pathways of monotremes by deep sequencing of six platypus and echidna tissues. We find that highly conserved microRNA species display their signature tissue-specific expression patterns. In addition, we find small RNAs with unusual features that are unique to monotremes. Platypus and echidna testes contain a robust piRNA system which appears to be participating in ongoing transposon defense.**

[Supplementary Figures 1 and 2 and Supplementary Tables 1 – 7 can be found online at < >]

## Introduction

Long thought by European biologists to be a hoax, the platypus, along with four species of echidna, represent the only remaining members of the basal mammalian clade of prototheria. Belonging to the monotreme taxon, these animals diverged from remaining mammalian lineages some * million years ago. Morphologically, monotremes display a mixture of characteristics of mammals, birds and reptiles. As mammals, monotremes have fur, but share with birds and reptiles a cloaca, the single external opening for both the gastrointestinal and urogenital systems. Monotremes lay eggs that are nourished in part through a placental structure before deposition, but they provide milk to their hatched young through mammary glands. Monotreme embryos show meroblastic rather than holoblastic cleavage, which is more typical of birds, reptiles and fish than of mammals.

The platypus is a specialized semi-aquatic carnivore, which displays a number of unique features. Its leathery bill serves as an electrosensory organ, and it is unique among mammals for its ability to produce venom, which the males deliver through spikes on their hind limbs. Perhaps the most remarkable feature of the platypus is being revealed by its genomic sequence, which holds a mixture of both mammalian and reptilian features. This makes it a uniquely placed comparative tool for studying mammalian evolution (Warren et al. 2007).

RNAi pathways are deeply conserved and can be traced at least as far back as the divergence of prokaryotes and eukaryotes, approximately 2.7 billion years ago. Individual eukaryotic lineages have adapted these pathways in different ways, though most use RNAi generally as both a genome defense and a gene regulatory mechanism. The RNAi machinery uses small RNAs of between 18 and 33 nucleotides in length as specificity determinants to regulate RNA stability and gene expression, through both post-transcriptional and transcriptional modes.

Two endogenous RNAi pathways have been described in mammals, the microRNA (miRNA) pathway and the piRNA pathway. The microRNA pathway uses small RNAs liberated from genome-encoded hairpins to regulate genes post-transcriptionally (reviewed in Bartel 2004). miRNAs act in diverse biological networks, with many having roles in reinforcing cell-type and tissue identity (reviewed in Plasterk 2006). Some families of miRNA genes have deeply conserved expression patterns, extending throughout animal lineages, whereas others are more rapidly-evolving and lineage-specific (Plasterk 2006). Piwi-interacting RNAs (piRNAs) are a larger size class of small RNAs that interact with the Piwi clade of Argonaute-family proteins. piRNAs show restricted expression patterns, being prominent in gonad, with mammalian piRNAs thus far being restricted to the germline compartment (Aravin et al. 2007a). In post-pubescent mammalian testis, piRNAs arise from dense strand-biased clusters whose biogenesis and functions remain largely unknown, though roles in transposon defense are supported for piRNAs that are expressed in germ cell precursors earlier in development (Aravin et al. 2007a).

Using the platypus genome sequence, we have performed a comparative analysis of small RNA pathways in platypus, echidna and eutherian mammals and have found deep conservation of many small RNA species as well as a surprisingly extensive set of monotreme-

specific and platypus-specific small RNAs. Strikingly, our analysis implicates small RNA functions in monotreme reproduction, as we find large clusters of germline-expressed fast-evolving miRNA loci in platypus and echidna, and evidence for ongoing transposon defense by piRNAs in the adult platypus testis.

## Results

### *Comparative analysis of RNAi components in platypus*

The completion of the platypus genome sequence offers an unprecedented opportunity to examine the conservation of small RNA pathways in prototherian mammals (Warren et al. 2007). Indeed, homologs of key RNAi components *Dicer*, *Drosha* and *Argonaute* (*Ago*) family members 1 through 4 could be readily identified in platypus (Fig. 1a). Moreover, synteny of the *Ago* gene family was conserved between platypus and eutherian mammals, with platypus *Agos 1, 3* and *4* closely linked on Ultracontig 472, and Ago2 on chromosome 4. Platypus also had obvious orthologs for Piwi family members *PiwiL1, PiwiL2* and *PiwiL4*, but appeared to be missing *PiwiL3* (Fig. 1b). As *PiwiL3* is only found in a subset of eutherian mammals, and could not be found in a marsupial genome (*Monodelphis domestica*), it is likely that this gene emerged after the divergence of the eutherian and prototherian lineages. The conservation pattern and longer branch length of *PiwiL2* suggest that it might be the ancestral Piwi protein in vertebrates, and that subsequent *Piwis* may have arisen by gene duplication events. This theory is consistent with observations that *PiwiL2* binds a smaller size range of piRNAs, more akin to the piRNAs found in invertebrates, and that it is the only Piwi protein in mammals that is expressed in both the embryonic and adult gonad (Aravin et al. 2006; Kuramochi-Miyagawa et al. 2004; Kuramochi-Miyagawa et al. 2001). Interestingly, the DDH motif, necessary for the endoribonuclease activity of Argonaute enzymes, was conserved in platypus *Ago2* and all three Piwi proteins as well as in *Ago3*, even though Ago3 appears to be non-catalytic (Liu et al. 2004; Meister et al. 2004). The conservation of this motif in the *Ago* clade throughout mammalian evolution is mysterious, as the biological significance of Argonaute-mediated RNA cleavage in mammals is unclear (Davis et al. 2005; O'Carroll et al. 2007; Yekta et al. 2004; Yu et al. 2005). Overall, our findings suggest that the platypus maintains a fully functional set of RNAi enzymes whose evolution reconstruct the phylogeny of the vertebrate lineage.

### *Identification of conserved and novel miRNAs in monotremes*

To determine the extent of conservation of miRNA pathways in monotremes, we cloned and deep sequenced the 18-24nt small RNA content of six tissues (brain, kidney, heart, lung, liver and testis) from platypus *(Ornithorhynchus anatinus)* and short-beaked echidna (*Tachyglossus aculeatus)*. 2,982,604 unique small RNAs were identified from these twelve libraries that mapped to the platypus genome: 1,409,428 small RNAs from platypus and 1,532,352 small RNAs from echidna. We devised a pipeline combining a computational heuristic with manual curation of miRNAs (Methods). This approach yielded 384 miRNA candidates and successfully identified all miRNAs annotated in the platypus genome by Ensembl as well as 23 additional well-conserved miRNAs that had not been annotated and 228 potential novel miRNAs (Supplementary Tables 1-4). Interestingly, we found several

miRNAs in platypus which were apparently shared with chicken but not with mammals(Warren et al. 2007).

miRNA expression patterns have been extensively characterized in several vertebrates (Ason et al. 2006; Chen et al. 2005; Darnell et al. 2006; Kloosterman et al. 2006a; Kloosterman et al. 2006b; Landgraf et al. 2007; Takada et al. 2006; Wienholds et al. 2005; Xu et al. 2006). Using miRNA cloning frequency data, we determined miRNA signatures for platypus and echidna brain, heart, lung, kidney, liver and testis (Fig. 2 and Supplementary Fig. 1). Although expression patterns of miRNAs tend to be conserved across vertebrate species, variation has been observed in some cases(Ason et al. 2006). We found strong conservation of expression domains and expression levels of orthologous miRNAs between both platypus and human and platypus and echidna (Fig. 2 and Supplementary Fig. 1 and 2). Interestingly, the tissue with the greatest differences in miRNA expression level between platypus and echidna was liver, perhaps reflecting the unique dietary adaptations of these two monotremes (Fig. 2b and Supplementary Fig. 1).

The platypus and echidna lineages diverged approximately 16.5 million years ago (Warren et al. 2007). Given this relatively recent evolutionary split, we expected to find the majority of potential monotreme-specific novel miRNAs in both platypus and echidna. However, of the 228 potential novel miRNA candidates identified through our screening approach, 131 were found only in the platypus libraries, 11 only in the echidna libraries (although their mature sequence was perfectly conserved in the platypus genome), and 86 were sequenced from both platypus and echidna (Fig. 3a and Supplementary Tables 1-3). Interestingly, we recovered the miRNA* sequence more frequently for miRNAs shared between monotremes. The miRNA* is the complementary strand of the miRNA duplex and a by-product of miRNA maturation. Although miRNA* strands are typically recovered during sequencing at only a fraction of the level of the mature miRNA, it provides additional confidence to a miRNA prediction. 50% of the novel miRNAs in both monotremes were cloned together with their miRNA* strands, compared with 40% in the platypus-only set and 0 in the echidna-only set (Supplementary Tables 1-3). The monotreme shared set also contained more miRNAs with high levels of expression (> 10,000 normalized cloning reads) (Fig. 3b). These results suggest that although we predict monotreme-specific miRNAs with greater confidence, there are also a number of platypus specific miRNAs that have either arisen subsequent to the divergence of platypus and echidna, or have been lost or are not expressed in the corresponding tissues examined from echidna.

We investigated tissue expression patterns of candidate novel miRNAs by comparing total normalized cloning frequencies across tissues. Surprisingly, the vast majority of novel miRNAs in both the monotreme and platypus-specific sets was found in testis (Fig. 3c). Indeed, closer inspection revealed that 68 of the novel miRNAs (30%), which constituted the majority of highly expressed species, arose from seven testis-expressed miRNA clusters on chromosome X1 and contigs 1754, 7160, 7359, 8388, 11344 and 22847. Revisiting the original scaffolding data allowed us to tentatively place contigs 1754, 7160, 7359 and 22847 on the long arm of chromosome X1, neighboring the miRNA cluster that can be definitively mapped to X1 (John Wallis, FG and EPM, personal communication). It is unclear whether the monotreme meiotic chain of ten sex chromosomes undergoes meiotic sex chromosome inactivation and

sex body formation as is observed in eutherian mammals and marsupials(Handel 2004; Namekawa et al. 2007). Abundant testis expression of at least one X-linked miRNA cluster in monotremes suggests either that either this cluster is expressed in pre-meiotic or somatic cells of the testis or that unpaired meiotic chromosomes are not completely silenced in monotremes.

The definitively mapped chromosome X1 cluster is ~40kb long and encodes at least 17 mature miRNAs (Fig. 3d and Supplementary Tables 2 and 3). Interestingly, although most miRNAs in this cluster are predominantly found in platypus testis, some are found equally in echidna testis, and one is almost entirely restricted to echidna (Fig. 3d). This pattern is also reflected in the testis clusters assigned to contigs, suggesting that these miRNA clusters are fast-evolving, even to the extent that the same sequence can be processed into a mature miRNA in one species but not another. Animal miRNAs generally regulate their targets through base-pairing interactions involving the "seed" region (positions 2-8 of the mature miRNA). Some of the miRNAs in testis clusters share seeds, and many have related seeds, suggesting that they may regulate a common set of targets.

*piRNAs in platypus*

piRNAs are a class of germline-expressed small RNAs that interact with Piwi proteins (Aravin et al. 2007a). Many piRNAs function as the specificity determinants of an RNA-based innate immune system that protects the germline from transposable elements and other nucleic acid pathogens (Aravin et al. 2007a). In eutherian mammals, piRNAs are ~26-30nt long, tend to start with a uracil and are abundantly expressed from discrete genomic intervals in testis(Aravin et al. 2007a). Radioactive end labeling of total RNA from platypus, echidna and mouse testes revealed an abundant class of ~30nt piRNAs in monotreme testis (Fig. 4a).

To investigate the relationships between the Piwi/piRNA pathways of monotremes and eutherians, we cloned and deeply sequenced RNAs sized between 25 and 33 nucleotides from platypus testis. 72% of uniquely mapping RNAs in this size range could be assigned to one of 50 large intervals in the platypus genome (Supplementary Table 5). These appeared similar to piRNA clusters observed in eutherian mammals, being an average of ~30kb long and exhibiting a strong strand bias for small RNA production. Based upon these characteristics and their 5' U bias, we conclude that these species are likely platypus piRNAs. As is the case in eutherian mammals, no piRNA clusters were observed on X chromosomes in platypus; however, as many of the platypus piRNA clusters are located on unmapped contigs we cannot rule out the possibility that platypus has piRNA clusters on sex chromosomes (Supplementary Table 5).

Platypus piRNAs were annotated based on genome mapping (Supplementary Table 6). In mouse, piRNAs expressed in pre-pubertal testis are repeat-rich and have a function in transposon control, but piRNAs expressed in adult testis are repeat poor and have no known function (Aravin et al. 2007a; Aravin et al. 2007b). Interestingly the platypus piRNA library contained a much larger proportion of repeat-associated species than an equivalent piRNA library from adult mouse testis (Girard et al. 2006) (Fig. 4b). Furthermore, the repeat composition of adult platypus and mouse repeat-associated piRNAs was quite distinct.

Whereas in platypus the majority of repeat-associated piRNAs were from LINE and SINE elements, these repeats only made up 16% of the repeat-derived adult mouse piRNAs (Fig. 4b). As LINE and SINE elements, specifically L2 and Mon1, are thought to be active in platypus (Warren et al. 2007), we speculated that perhaps a difference in transposon activity may underlie the differential contribution of these repeat elements to the piRNA populations of adult testis in monotreme and eutherian mammals

Enrichment for A residues at position 10 of piRNAs is a signature of their participation in ongoing transposon defense (Aravin et al. 2007a; Brennecke et al. 2007). In other systems, piRNA populations with a 10A bias are secondary piRNAs and largely arise from transcripts from active transposons (Brennecke et al. 2007). If LINE and SINE elements are active in platypus a higher proportion of 10A piRNAs might be expected from these elements than from the bulk piRNA population. In accord with this prediction, the percentage of 10A was significantly greater in L2 and Mon1 piRNAs (Fig. 4c, $p < 10^{-125}$). We also reasoned that the more times a repeat-derived piRNA matched the genome the greater the likelihood that it is derived from an active transposon, and thus greater the likelihood of 10A. To test whether we could see this effect in the platypus piRNA population, we calculated the 10A bias for piRNAs that mapped to the genome at least ten times. Strikingly, 10A was at 49% in this population, and increased to 53% and 56% in L2 and Mon1 piRNAs respectively (Fig. 4c, $p < 10^{-4}$). Moreover, piRNAs that mapped once to the genome but lay outside of the 50 piRNA clusters (Supplementary Table 5) and thus presumably represented individual active transposons recognizable through their divergent sequences, had a higher proportion of 10A, particularly if they were derived from L2 or Mon1 elements (Fig. 4c). Overall, this analysis provides evidence that piRNA pathways in platypus are active in transposon defense against L2 and Mon1 elements, and this may underlie the high proportion of LINE and SINE piRNAs in platypus.

One characteristic feature of piRNAs is that they are generally larger than miRNAs and siRNAs. However, a surprisingly large proportion of the platypus testis 18-24nt library was composed of small RNAs with piRNA-specific properties. Indeed, 68% of small RNAs in this library that mapped to the genome once were derived from one of the 50 piRNA clusters (Fig. 4d). Furthermore, 79% of small RNAs in this library that mapped to the genome at least once had a U at position 1 (Fig. 4d). Further studies will be required to determine if this small piRNA class has unique biogenesis or Piwi binding characteristics as compared to the abundant (Fig 4A), larger piRNA species.

## Discussion

Overall, these studies indicate that that the RNAi machinery and its functions in genome defense and gene regulation are conserved in monotremes. Cloning and deep sequencing of miRNAs and piRNAs in monotremes has revealed small RNAs and regulatory systems similar to those found in other mammals, as well as monotreme-specific components.

In general we found conservation of both sequence and expression patterns of miRNAs in monotremes and eutherian mammals. We also examined conservation of miRNAs between platypus and non-mammalian vertebrates. Although miRNAs have not been extensively characterized in the chicken genome, we nevertheless identified some miRNAs that are

shared between the platypus and chicken genomes but that are not shared with mammals. As these miRNAs appear to have emerged and subsequently disappeared in the mammalian lineage, they are presumably involved in biological functions that have either become redundant or obsolete during mammalian evolution.

We have identified 228 novel miRNA candidates in platypus and echidna, including a large cluster of testis-expressed miRNAs on chromosome X1. There are several examples of such expanded fast-evolving miRNA clusters in other species (e.g. the *miR-467* cluster in rodents). It is tempting to speculate that although miRNAs and their expression patterns are often deeply conserved across lineages, their flexibility may allow them to facilitate or even catalyze evolutionary change. Analysis of miRNA genes in additional vertebrate groups will illuminate the significance of such fast-evolving miRNA clusters in specialized biological niches.

Although platypus piRNAs are organized in large strand-biased clusters similar to those found in other mammals, the piRNA system appears to be actively engaged in combating transposon activity in mature platypus testis. This is in contrast to piRNAs in adult testis of other mammals, which are depleted of repeat sequences and have no known function (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Watanabe et al. 2006). The piRNAs of platypus adult testis have features of both fish and eutherian piRNA systems (Houwing et al. 2007), and suggest that mammalian piRNA clusters that are expressed late in germ cell development may have progressively lost repeat content as they acquired novel, and as yet unknown functions (Aravin et al. 2007b; Carmell et al. 2007).

The small RNA complement of platypus has highlighted not only unique and conserved roles for miRNAs and piRNAs in monotremes but also provides insight into the evolution of small RNA pathways in mammals.

## Methods

*Tree building*

The phylogenetic trees of Argonaute and Piwi families in multiple species were constructed using the Phylip package (version 3.67) (Felsenstein 1997). The full-length coding region of each gene was used to calculate the distance, and a neighbor-joining method used to obtain the topology of the tree.

*Small RNA cloning, sequencing and annotation*

RNA was isolated by a standard Trizol method from snap frozen tissue of adult male platypus and echidna (*Tachyglossus aculeatus)* heart, liver, lung, kidney, brain and testis collected from animals captured at the upper Barnard river, New South Wales, Australia during breeding season (AEC permit no. S-49-2006 to F.G.). MiRNA (18-24nt) and piRNA (25-33nt) fractions of total RNA were used to clone libraries as described (Brennecke et al. 2007). Small RNA libraries were sequenced on the Illumina 1G sequencer. In the case of miRNA libraries, 3' linkers were clipped using a dynamic alignment algorithm which required perfect matches at positions 25-30, but allowed mismatches towards the end of the read. The piRNA library was uniformly clipped to 26nt in length. Platypus and echidna small RNAs with perfect matches to the platypus genome were annotated using Ensembl platypus genome annotations.

*miRNA prediction*

Clipped small RNAs of 18nt or longer from platypus and echidna 18-24nt libraries were mapped to the platypus genome requiring perfect sequence identity. The potential for each position of the genome to be a 5' end of a mature miRNA sequence was evaluated by counting the total number of sequence counts starting at that position. Two windows of 150 base pairs were folded around the potential mature miRNA, one starting 30 bases upstream (for a mature in the left arm) and one starting 100 bases upstream (for a mature in the right arm). Folding was done by sampling 500 structures using RNAsubopt from the Vienna RNA Package (Wuchty et al. 1999) and taking the one that had the greatest number of paired arm bases. The longest pair of substrings whose bases only matched to each other and whose starts and ends were paired were defined as the two arms, and each hairpin was trimmed to the ends of the arms. Only hairpins that had at least 20 paired bases in the arms were considered. It was required that at least 18 of 20 bases following the potential mature 5' end overlapped with the intended arm and that none of the bases overlapped with the opposite arm. We considered only hairpins for which at least 50% of the reads within 10 bases of the hairpin, and 40% of reads within the hairpin belonged to the putative 5'end of the mature miRNA, and at least 40% of reads within the entire hairpin aligned exactly to the 5'end of the putative mature miRNA. In order to avoid repeat associated sequences, no read within the hairpin could have more than 5 total matching positions in the genome, and 99% of the reads must not have more than 3 matching positions. It was required that there be no more than 6 positions that had more than 1% of the hairpin and no more than 4 positions that had more than 10% of the hairpin reads. The 2350 hairpins that fit these criteria were inspected manually to remonve piRNAs, repeats, degradation products and other types of non-coding RNAs, leaving 384 candidate miRNAs. The most frequently cloned small RNA in the candidate miRNA hairpin was designated as the mature sequence, and the most frequently cloned sequence from the other arm was

designated as the miRNA* sequence. In cases where cloning frequencies were equal from both arms, sequences from both arms were designated as mature. Platypus and echidna orthologs of known miRNAs were identified by Blastn with the Rfam miRNA database, requiring at least 15 nucleotide matches. Platypus and echidna candidate miRNAs were arbitrarily named Platypus-1 to Platypus-426 (not inclusive), pending approval of Rfam nomenclature.

*Heatmap construction*

Human orthologs for candidate mature miRNAs in platypus brain, heart, liver and testis (for comparison with human) or platypus brain, heart, liver, testis, kidney and lung (for comparison with mouse) were found by blastn using the human or mouse Rfam database, and requiring at least 15 matched bases. miRNA read counts in platypus and those in corresponding tissues in a human and mouse miRNA expression atlas (Landgraf et al. 2007) were first normalized to the tissue with the least number of total miRNA reads. MicroRNAs that did not have at least 10 normalized reads in both platypus and human/mouse were discarded. Each miRNA in each tissue was assigned an expression score, a ratio of the number of reads in that tissue versus all tissues in that organism. The miRNAs were then clustered by their tissue expression ratios in platypus using the repeated bisection option in the Cluto clustering package (Zhao and Karypis 2005). The distance between the confidence intervals of the expression ratios for each platypus/human or platypus/mouse pair of miRNAs was computed after Wilson correction (Brown et al. 2001) and was displayed as a difference score on the heatmap. The following human and mouse libraries from the miRNA expression atlas (Landgraf et al. 2007) were used for comparison with platypus: brain, hsa_Frontal-cortex-adult (human), average mmu_Brain-WT and mmu_Cortex (mouse); heart, hsa_Heart (human) and mmu_Heart (mouse); liver, hsa_Liver (human) and mmu_Liver (mouse); testis, hsa_Testis (human) and mmu_Testis (mouse); lung, mmu_Lung (mouse); kidney, mmu_Kidney (mouse). For platypus/echidna expression profiles, the same procedure was used, except read counts for candidate mature miRNAs in platypus for brain, kidney, lung, heart, liver and testis were compared to those for identical miRNAs found in echidna.

**Acknowledgements**

We thank members of the Hannon Lab for helpful discussions. We thank Russell Jones (Newcastle University) for help with the tissue collection. Facilities were provided by Macquarie Generation and Glenrock Station, New South Wales. Approval to collect animals was granted by the New South Wales National Parks and Wildlife Services, New South Wales Fisheries and the Animal Ethics Committee, University of Adelaide. FG is an Australian Research Council Research Fellow. EPM is a Sir Keith Murdoch Fellow of the American Australian Association. AS was supported in part by the Schering AG/Ernst Schering Foundation and in part by the Human Frontier Science Program Organization (HFSPO). PK was supported in part by a National Science Foundation Graduate Research Fellowship. This work was funded in part by grants from the NIH (GJH) and by a kind gift from Kathryn W. Davis (GJH).

## References


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: 203-207.

Aravin, A.A., G.J. Hannon, and J. Brennecke. 2007a. The Piwi/piRNA pathway: adaptive defense for the transposon arms race. Science In press.

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 316: 744-747.

Ason, B., D.K. Darnell, B. Wittbrodt, E. Berezikov, W.P. Kloosterman, J. Wittbrodt, P.B. Antin, and R.H. Plasterk. 2006. Differences in vertebrate microRNA expression. Proc Natl Acad Sci U S A 103: 14385-14389.

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

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: 1089-1103.

Brown, L.D., T.T. Cai, and A. DasGupta. 2001. Interval Estimation for a Binomial Proportion. Statistical Science 16: 101-133.

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: 503-514.

Chen, P.Y., H. Manninga, K. Slanchev, M. Chien, J.J. Russo, J. Ju, R. Sheridan, B. John, D.S. Marks, D. Gaidatzis, C. Sander, M. Zavolan, and T. Tuschl. 2005. The developmental miRNA profiles of zebrafish as determined by small RNA cloning. Genes Dev 19: 1288-1293.

Darnell, D.K., S. Kaur, S. Stanislaw, J.H. Konieczka, T.A. Yatskievych, and P.B. Antin. 2006. MicroRNA expression during chick embryo development. Dev Dyn 235: 3156-3165.

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

Felsenstein, J. 1997. An alternating least squares approach to inferring phylogenies from pairwise distances. Syst Biol 46: 101-111.

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.

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

Handel, M.A. 2004. The XY body: a specialized meiotic chromatin domain. Exp Cell Res 296: 57-63.

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

Kloosterman, W.P., F.A. Steiner, E. Berezikov, E. de Bruijn, J. van de Belt, M. Verheul, E. Cuppen, and R.H. Plasterk. 2006a. Cloning and expression of new microRNAs from zebrafish. Nucleic Acids Res 34: 2558-2569.

Kloosterman, W.P., E. Wienholds, E. de Bruijn, S. Kauppinen, and R.H. Plasterk. 2006b. In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. Nat Methods 3: 27-29.

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: 839-849.

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: 121-133.

Landgraf, P., M. Rusu, R. Sheridan, A. Sewer, N. Iovino, A. Aravin, S. Pfeffer, A. Rice, A.O. Kamphorst, M. Landthaler, C. Lin, N.D. Socci, L. Hermida, V. Fulci, S. Chiaretti, R. Foa, J. Schliwka, U. Fuchs, A. Novosel, R.U. Muller, B. Schermer, U. Bissels, J. Inman, Q. Phan, M. Chien, D.B. Weir, R. Choksi, G. De Vita, D. Frezzetti, H.I. Trompeter, V. Hornung, G. Teng, G. Hartmann, M. Palkovits, R. Di Lauro, P. Wernet, G. Macino, C.E. Rogler, J.W. Nagle, J. Ju, F.N. Papavasiliou, T. Benzing, P. Lichter, W. Tam, M.J. Brownstein, A. Bosio, A. Borkhardt, J.J. Russo, C. Sander, M. Zavolan, and T. Tuschl. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129: 1401-1414.

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: 363-367.

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: 1437-1441.

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

Namekawa, S.H., J.L. VandeBerg, J.R. McCarrey, and J.T. Lee. 2007. Sex chromosome silencing in the marsupial male germ line. Proc Natl Acad Sci U S A 104: 9730-9735.

O'Carroll, D., I. Mecklenbrauker, P.P. Das, A. Santana, U. Koenig, A.J. Enright, E.A. Miska, and A. Tarakhovsky. 2007. A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. Genes Dev 21: 1999-2004.

Plasterk, R.H. 2006. Micro RNAs in animal development. Cell 124: 877-881.

Takada, S., E. Berezikov, Y. Yamashita, M. Lagos-Quintana, W.P. Kloosterman, M. Enomoto, H. Hatanaka, S. Fujiwara, H. Watanabe, M. Soda, Y.L. Choi, R.H. Plasterk, E. Cuppen, and H. Mano. 2006. Mouse microRNA profiles determined with a new and sensitive cloning method. Nucleic Acids Res 34: e115.

Warren, W.C., L.W. Hillier, J.A. Marshall Graves, E. Birney, C.P. Ponting, F. Grutzner, K. Belov, W. Miller, L. Clarke, A.T. Chinwalla, S.P. Yang, A. Heger, D. Locke, P. Miethke, P.D. Waters, F. Veyrunes, L. Fulton, B. Fulton, T. Graves, J. Wallis, X.S. Puente, C. Lopez-Otin, G.R. Ordonez, E.E. Eichler, L. Chen, Z. Cheng, J.E. Deakin, A. Alsop, K. Thompson, P. Kirby, A.T. Pappenfuss, M.J. Wakefield, T. Olender, D. Lancet, G.A.

Huttley, A.F.A. Smit, A. Pask, P. Temple-Smith, M.A. Batzer, J.A. Walker, M.K. Konkel, R.S. Harris, C.M. Whittington, E.S.W. Wong, N. Gemmell, E. Buschiazzo, I. Vargas Jentzsch, A. Merkel, J. Schmitz, A. Zemann, G. Churakov, J. Ole Kriegs, J. Brosius, E.P. Murchison, R. Sachidanandam, C. Smith, A. Stark, P. Kheradpour, G.J. Hannon, E. Tsend-Ayush, E. McMillan, R. Attenborough, W. Rens, M. Ferguson-Smith, C.M. Lefèvre, J.A. Sharp, K.R. Nicholas, D.A. Ray, M. Kube, R. Reinhard, T.H. Pringle, J. Taylor, R.C. Jones, B. Nixon, J. Dacheux, H. Niwa, Y. Sekita, X. Huang, P. Flicek, C. Webber, R. Hardison, J. Nelson, K. Hallsworth-Peppin, M. Renfree, W.U.G.S. Center, E.R. Mardis, and R.K. Wilson. 2007. Genome analysis of the platypus reveals unique signatures of evolution. Nature Submitted.

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: 1732-1743.

Wienholds, E., W.P. Kloosterman, E. Miska, E. Alvarez-Saavedra, E. Berezikov, E. de Bruijn, H.R. Horvitz, S. Kauppinen, and R.H. Plasterk. 2005. MicroRNA expression in zebrafish embryonic development. Science 309: 310-311.

Wuchty, S., W. Fontana, I.L. Hofacker, and P. Schuster. 1999. Complete suboptimal folding of RNA and the stability of secondary structures. Biopolymers 49: 145-165.

Xu, H., X. Wang, Z. Du, and N. Li. 2006. Identification of microRNAs from different tissues of chicken embryo and adult chicken. FEBS Lett 580: 3610-3616.

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

Yu, Z., T. Raabe, and N.B. Hecht. 2005. MicroRNA Mirn122a reduces expression of the posttranscriptionally regulated germ cell transition protein 2 (Tnp2) messenger RNA (mRNA) by mRNA cleavage. Biol Reprod 73: 427-433.

Zhao, Y. and G. Karypis. 2005. Data clustering in life sciences. Mol Biotechnol 31: 55-80.

## Figure legends

**Figure 1: RNA interference genes in platypus.** Phylogenetic trees of **(A)** *Argonaute* and **(B)** *Piwi* families in multiple vertebrate species. The platypus *PiwiL4* (*oanPiwiL4*) sequence was incomplete, and a partial coding sequence was used to place this gene on the tree (dotted line) (Supplementary Table 7). Platypus genes are highlighted in red. Scale: 0.1 = 0.1 base substitutions expected. *oan, Ornithorhynchus anatinus; cfa, Canis familiaris; hsa, Homo sapiens; mmu, Mus musculus; rno, Rattus norvegicus; mdo, Monodelphis domesctica; gga, Gallus gallus; xtr, Xenopus tropicalis; dre, Danio rerio.*

**Figure 2: microRNA expression patterns in monotremes.** Comparison of **(A)** platypus and human and **(B)** platypus and echidna miRNA tissue expression profiles. The expression score represented by blue squares in columns two and three is a ratio of the number of normalized reads in that tissue versus all tissues in that organism and the difference score represented by red squares in column one is the difference between the confidence intervals of the expression ratios for each platypus/human or platypus/echidna pair of miRNAs. For the platypus/echidna comparison a subset of miRNAs with the most distinct tissue-specific expression patterns are shown here clustered by tissue expression; the complete dataset can be found in Supplementary Fig. 1.

**Figure 3: Novel microRNAs in monotremes.** 228 novel miRNAs were predicted in platypus and echidna. **(A)** Pie graph of monotreme species distribution of novel miRNAs. Of the 228 novel miRNAs, 86 were sequenced in both platypus and echidna small RNA libraries (blue); 131 were only in platypus (green); and 11 were cloned only in echidna (red) although their sequences were present in the platypus genome. **(B)** Frequency distribution of novel miRNAs. miRNA cloning frequency was normalized so that all tissues had an equal number of miRNA reads. Novel miRNAs were binned based on their normalized cloning frequencies. For the monotreme class, frequency represents the average of platypus and echidna frequencies in this class. Monotreme (blue), platypus (green) and echidna (red). **(C)** Tissue distribution of novel miRNAs. Total miRNA normalized cloning frequency is plotted for each tissue. Monotreme cloning frequency is an average of platypus and echidna cloning frequencies for each miRNA in this class. Monotreme (blue), platypus (green) and echidna (red). **(D)** MiRNAs in novel testis-expressed chromosome X1 cluster. Only miRNAs that map to the genome three times or less are shown. The unnormalized cloning frequency divided by the number of genome matches for each miRNA is represented by a bar. Platypus miRNAs are plotted in green, echidna in red; platypus and echidna bars are independent and not additive.

**Figure 4: piRNAs in platypus. (A)** Total RNA labeling from mouse, platypus and echidna testis reveals a distinct ~30nt piRNA species in monotremes. RNA was dephosphorylated, end labeled with $^{32}$P and run on a denaturing 15% polyacrylamide gel. Sizes of markers (left) are in nucleotides. **(B)** Comparison of platypus and mouse piRNA libraries. A platypus piRNA library was generated by cloning small RNAs in a 25-33nt size window from total platypus testis RNA. Sequences from this library and from an equivalent mouse piRNA library(Girard et al. 2006) were classed as repeat (red), unannotated (blue) or other (beige). The repeat piRNAs were broken down into classes of repeat (bars). **(C)** 10A profiles for platypus piRNAs. An A at piRNA position 10 is a signature of an active piRNA amplification system. “Clustered” piRNAs are

defined as those that map to the genome once and fall into one of 50 piRNA clusters (Supplementary Table 5), while “non-clustered” piRNAs are those that map to the genome once but do not fall into a cluster. 10A percentages for all piRNAs are in black, those that are derived from LINE2 elements in blue, and those from Mon1 SINE elements in white. **(D)** Characteristics of “small” piRNAs. “Percent 1U” includes all small RNAs that match the genome; “percent in clusters” includes only small RNAs that map only once to the genome.

**Supplementary Figures and Tables**

**Supplementary Figure 1.** Expression heatmap for all miRNAs conserved in platypus and echidna brain, kidney, testis, lung, heart and liver with a normalized cloning multiplicity of at least 10 in each tissue.

**Supplementary Figure 2.** Expression heatmap for miRNAs conserved in platypus and mouse brain, testis, liver and heart, kidney and lung.

**Supplementary Tables 1 – 4.** 384 predicted platypus and echidna miRNAs were divided into those that matched Rfam miRNAs (Supplementary Table 1) and those that did not ("novel miRNAs"). Putative novel miRNAs were further divided into those that were shared between platypus and echidna (Supplementary Table 2) and those that were cloned only in platypus (Supplementary Table 3) or only in echidna (Supplementary Table 4). Putative miRNAs were named Platypus-1 through Platypus-426 (non-inclusive) pending approval of Rfam nomenclature. Chromosomal or contig coordinates are listed. Cloning multiplicities for each putative miRNA are shown for each tissue, as well as the total cloning multiplicities in platypus and echidna. Cloning multiplicities have been normalized such that the total number of miRNA reads per tissue in this set is identical. Cloning multiplicities refer only to the shown annotated mature miRNA and do not include multiplicities of miRNA variants with variable 3' ends or of miRNA* strands. **Supp. Table 1. Conserved miRNAs in platypus.** Platypus miRNAs related to known Rfam miRNAs in other species. **Supp. Table 2. Novel miRNAs in monotremes.** MiRNAs cloned in both platypus and echidna but absent from Rfam. MiRNAs from testis clusters are highlighted in grey. **Supp. Table 3. Novel miRNAs in platypus.** MiRNAs cloned in platypus but not echidna and absent from Rfam. MiRNAs from testis clusters are highlighted in grey. **Supp. Table 4. Novel miRNAs in echidna.** MiRNAs cloned in echidna but not platypus that nevertheless match the platypus genome.

Figure 1

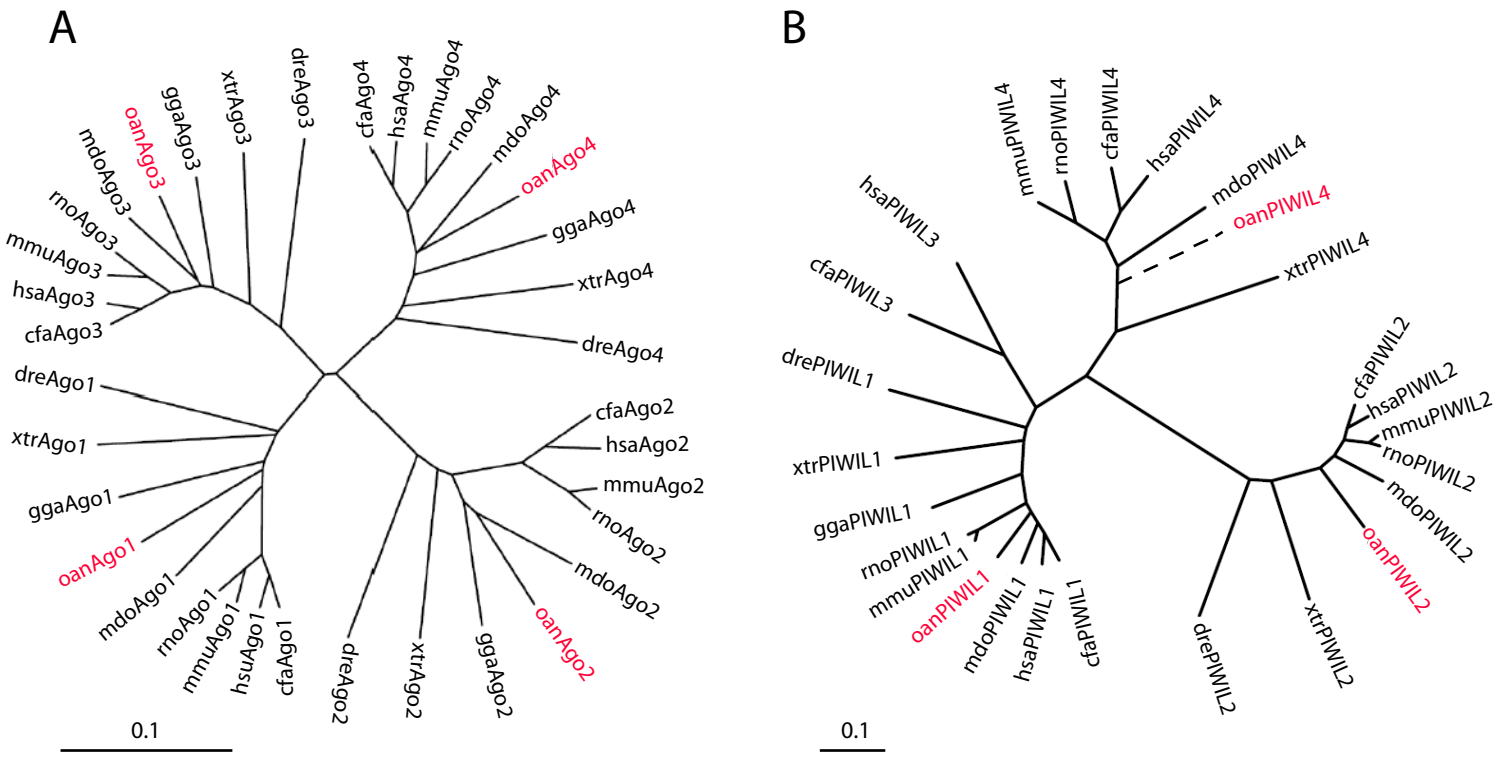

A
dreAgo3
cfaAgo4
hsaAgo4
mmuAgo4
rnoAgo4
mdoAgo4
oanAgo4
ggaAgo4
xtrAgo4
dreAgo4
cfaAgo2
hsaAgo2
mmuAgo2
rnoAgo2
mdoAgo2
oanAgo2
ggaAgo2
xtrAgo2
dreAgo2
cfaAgo1
hsuAgo1
mmuAgo1
rnoAgo1
mdoAgo1
oanAgo1
ggaAgo1
xtrAgo1
dreAgo1
cfaAgo3
hsaAgo3
mmuAgo3
rnoAgo3
mdoAgo3
oanAgo3
ggaAgo3
xtrAgo3
0.1
B
mmuPIWIL4
rnoPIWIL4
cfaPIWIL4
hsaPIWIL4
mdoPIWIL4
oanPIWIL4
xtrPIWIL4
cfaPIWIL2
hsaPIWIL2
mmuPIWIL2
rnoPIWIL2
mdoPIWIL2
oanPIWIL2
xtrPIWIL2
drePIWIL2
cfaPIWIL1
hsaPIWIL1
mdoPIWIL1
oanPIWIL1
mmuPIWIL1
rnoPIWIL1
ggaPIWIL1
xtrPIWIL1
drePIWIL1
cfaPIWIL3
hsaPIWIL3
0.1

Figure 2

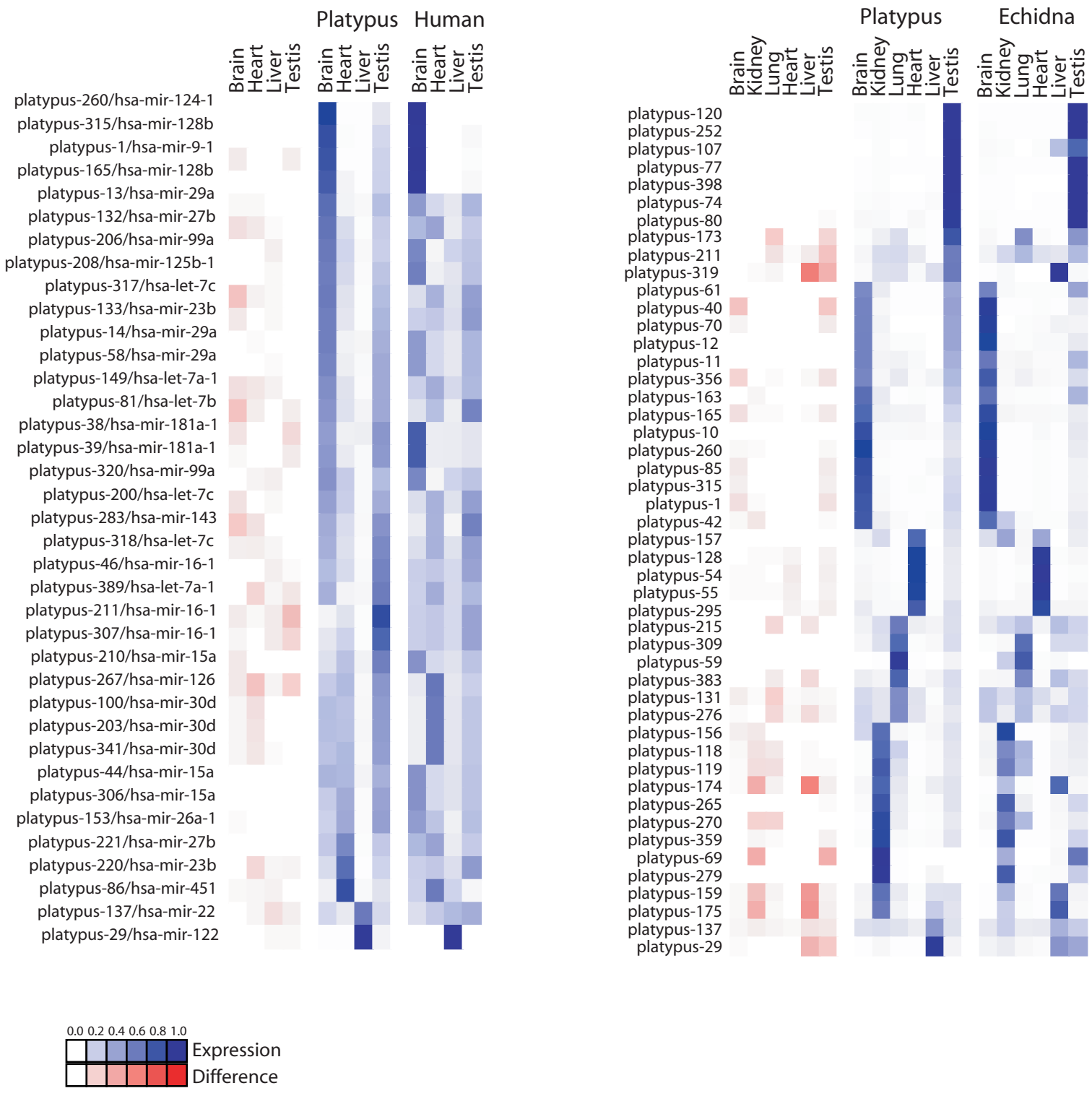

A
Platypus
Human
Brain
Heart
Liver
Testis
platypus-260/hsa-mir-124-1
platypus-315/hsa-mir-128b
platypus-1/hsa-mir-9-1
platypus-165/hsa-mir-128b
platypus-13/hsa-mir-29a
platypus-132/hsa-mir-27b
platypus-206/hsa-mir-99a
platypus-208/hsa-mir-125b-1
platypus-317/hsa-let-7c
platypus-133/hsa-mir-23b
platypus-14/hsa-mir-29a
platypus-58/hsa-mir-29a
platypus-149/hsa-let-7a-1
platypus-81/hsa-let-7b
platypus-38/hsa-mir-181a-1
platypus-39/hsa-mir-181a-1
platypus-320/hsa-mir-99a
platypus-200/hsa-let-7c
platypus-283/hsa-mir-143
platypus-318/hsa-let-7c
platypus-46/hsa-mir-16-1
platypus-389/hsa-let-7a-1
platypus-211/hsa-mir-16-1
platypus-307/hsa-mir-16-1
platypus-210/hsa-mir-15a
platypus-267/hsa-mir-126
platypus-100/hsa-mir-30d
platypus-203/hsa-mir-30d
platypus-341/hsa-mir-30d
platypus-44/hsa-mir-15a
platypus-306/hsa-mir-15a
platypus-153/hsa-mir-26a-1
platypus-221/hsa-mir-27b
platypus-220/hsa-mir-23b
platypus-86/hsa-mir-451
platypus-137/hsa-mir-22
platypus-29/hsa-mir-122
0.0 0.2 0.4 0.6 0.8 1.0
Expression
Difference
B
Platypus
Echidna
Brain
Kidney
Lung
Heart
Liver
Testis
platypus-120
platypus-252
platypus-107
platypus-77
platypus-398
platypus-74
platypus-80
platypus-173
platypus-211
platypus-319
platypus-61
platypus-40
platypus-70
platypus-12
platypus-11
platypus-356
platypus-163
platypus-165
platypus-10
platypus-260
platypus-85
platypus-315
platypus-1
platypus-42
platypus-157
platypus-128
platypus-54
platypus-55
platypus-295
platypus-215
platypus-309
platypus-59
platypus-383
platypus-131
platypus-276
platypus-156
platypus-118
platypus-119
platypus-174
platypus-265
platypus-270
platypus-359
platypus-69
platypus-279
platypus-159
platypus-175
platypus-137
platypus-29

Figure 3

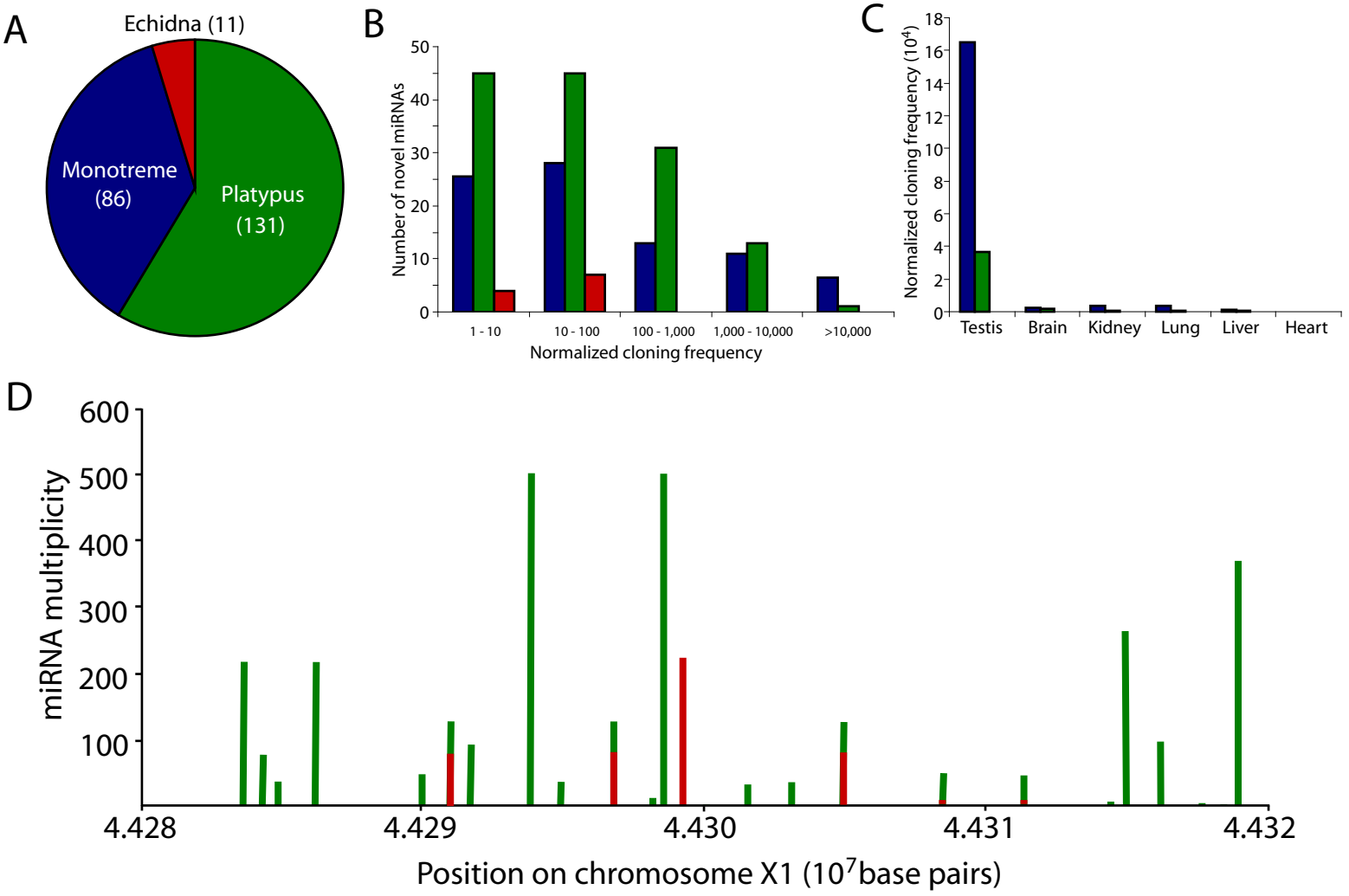
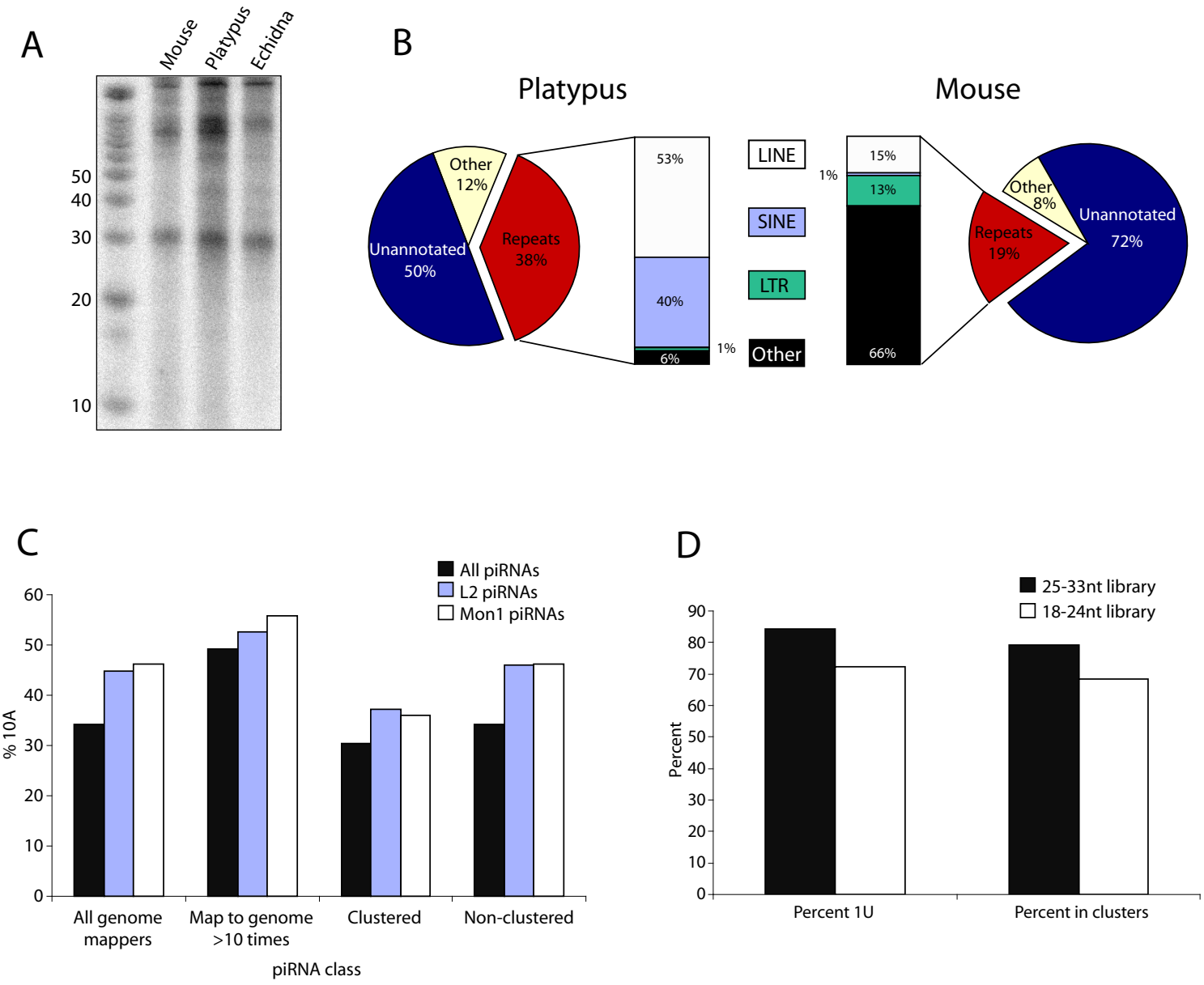

A
Echidna (11)
Monotreme (86)
Platypus (131)
B
Number of novel miRNAs
50
40
30
20
10
0
1 - 10
10 - 100
100 - 1,000
1,000 - 10,000
>10,000
Normalized cloning frequency
C
Normalized cloning frequency ($10^4$)
18
16
14
12
10
8
6
4
2
0
Testis
Brain
Kidney
Lung
Liver
Heart
D
miRNA multiplicity
600
500
400
300
200
100
4.428
4.429
4.430
4.431
4.432
Position on chromosome X1 ($10^7$base pairs)

Figure 4


A
Mouse
Platypus
Echidna
50
40
30
20
10
B
Platypus
Other 12%
Unannotated 50%
Repeats 38%
53%
40%
6%
1%
LINE
SINE
LTR
Other
Mouse
1%
15%
13%
66%
Other 8%
Repeats 19%
Unannotated 72%
C
All piRNAs
L2 piRNAs
Mon1 piRNAs
% 10A
60
50
40
30
20
10
0
All genome mappers
Map to genome >10 times
Clustered
Non-clustered
piRNA class
D
25-33nt library
18-24nt library
Percent
90
80
70
60
50
40
30
20
10
0
Percent 1U
Percent in clusters