

A germline-specific class of small RNAs binds mammalian Piwi proteins

Angélique Girard^{1,2}, Ravi Sachidanandam¹, Gregory J. Hannon¹ & Michelle A. Carmell¹

Small RNAs associate with Argonaute proteins and serve as sequence-specific guides to regulate messenger RNA stability, protein synthesis, chromatin organization and genome structure^{1–3}. In animals, Argonaute proteins segregate into two subfamilies⁴. The Argonaute subfamily acts in RNA interference and in microRNA-mediated gene regulation using 21–22-nucleotide RNAs as guides. The Piwi subfamily is involved in germline-specific events such as germline stem cell maintenance and meiosis. However, neither the biochemical function of Piwi proteins nor the nature of their small RNA guides is known. Here we show that MIWI, a murine Piwi protein, binds a previously uncharacterized class of ~29–30-nucleotide RNAs that are highly abundant in testes. We have therefore named these Piwi-interacting RNAs (piRNAs). piRNAs show distinctive localization patterns in the genome, being predominantly grouped into 20–90-kilobase clusters, wherein long stretches of small RNAs are derived from only one strand. Similar piRNAs are also found in human and rat, with major clusters occurring in syntenic locations. Although their function must still be resolved, the abundance of piRNAs in germline cells and the male sterility of *Miwi* mutants suggest a role in gametogenesis.

In multiple organisms, mutations in *Piwi*-family genes cause defects in germline development⁵. In flies, *piwi* mutations have both cell autonomous and non-autonomous effects on female germline stem cells^{6,7}. In addition, both Piwi and Aubergine are important for repression of repetitive elements^{8–15}. In mammals, expression of the three Piwi subfamily members, MIWI, MILI and MIWI2, is mainly germline-restricted. Moreover, mice bearing targeted mutations in *Miwi* (ref. 16), *Mili* (ref. 17) and *Miwi2* (M.A.C., A.G. and G.J.H., unpublished data) are male-sterile with distinct defects in spermatogenesis. Therefore, we searched for potential small RNA guides for mammalian Piwi proteins in the male germ line. As a first step, we surveyed small RNA populations in testes by pCp labelling (not shown). In addition to a 21–22-nucleotide species, a species of 30 nucleotides was strongly labelled. In parallel, we prepared total RNA from several mouse tissues and visualized small RNAs by staining with SYBR gold. Although this technique could barely detect microRNAs (miRNAs), a prominently staining 30-nucleotide RNA species was visible specifically in testis RNA (Fig. 1a and Supplementary Fig. S1). These small RNAs were cloned using established procedures¹⁸, and three sequenced RNAs, piR-1, piR-2 and piR-3, were used in northern blot analyses to confirm tissue specificity (Fig. 1b–d).

As these RNAs were abundant in testes, they represented possible small RNA partners for Piwi proteins. We therefore examined MIWI immunoprecipitates for their presence. Northern blotting revealed that the ~30-nucleotide piR-1 RNA species specifically co-purified with MIWI but not with Ago2. Similar results have been found with other randomly selected piRNAs (not shown). Interestingly, a

~22-nucleotide version of piR-1 was observed under some conditions, but the generality of such variants is unclear (not shown). Probing the same immunoprecipitates for let-7c showed a strong and specific signal in Ago2 complexes, but minimal signal in MIWI immunoprecipitates. Although MIWI may also function together with miRNAs, our results indicate that its major binding partner, and perhaps that of other Piwi family members, is a previously undescribed class of ~30-nucleotide small RNAs, which we have termed Piwi-interacting RNAs (piRNAs).

We obtained a more detailed picture of this new small RNA class using a highly parallel pyrosequencing methodology¹⁹. Some 87,463 reads were obtained, 52,934 of which are classified as candidate piRNAs (see Supplementary Information). Of these, 26% were cloned only once, indicating that the piRNA population in testes is complex but that sufficient sequencing depth was achieved to obtain a representative sample. One notable feature was an overwhelming bias for uracil at the first position (Supplementary Fig. S2B). Overall, 94.2% of sequences began with U. Remaining positions did not show any dramatic biases. Although the underlying cause of this unusual pattern is unknown, a less extreme 5' nucleotide bias has been noted in other classes of small RNAs, including repeat-associated siRNAs (rasiRNAs), miRNAs and effective short-interfering RNAs (siRNAs)²⁰.

piRNA distribution among chromosomes did not correlate either with gene density or repeat density. When normalized to chromosome length, some chromosomes were piRNA-rich whereas others were piRNA-poor (Supplementary Fig. S2C). Most piRNAs (83.7%)

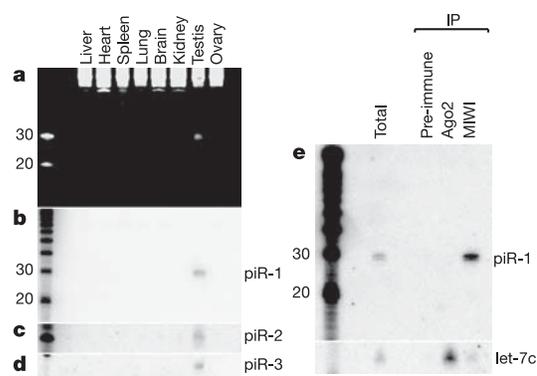


Figure 1 | Novel testis-specific small RNAs associate with MIWI. **a**, SYBR gold staining of 20 μ g of total RNA from various mouse tissues. **b–d**, Northern blot analyses using three cloned sequences: piR-1 (**b**), piR-2 (**c**) and piR-3 (**d**). **e**, Immunoprecipitates from testis extracts with antibodies directed against Argonaute2 and MIWI were analysed by northern blotting, using piR-1 and let-7c as probes. Size (nucleotides) is shown along the left of each panel.

¹Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. ²Ecole Nationale Supérieure des Mines de Paris, 60 boulevard Saint-Michel, 75272 Paris, France.

mapped uniquely in the genome (Supplementary Fig. S2D). Seventeen per cent mapped to repeats, including short interspersed elements (SINEs), long interspersed elements (LINEs), long terminal repeat (LTR) retrotransposons, and a variety of other repeat classes (Supplementary Table S1). Very few mapped to major and minor satellites that comprise the mouse centromere and none mapped to telomeric repeats. Although repeat-associated piRNAs form a significant portion of the total, this fraction is much less than would be expected based on a random distribution. Thus, mammalian piRNAs are not likely to be functional equivalents of *Drosophila* rasiRNAs, which are ~23–26 nucleotides in length and mostly derived from various classes of transposons. However, resolution of this issue will await a fuller understanding of both classes of RNAs and the identification of an Argonaute binding partner for rasiRNAs.

An examination of the 52,329 piRNAs that map to between one and five discrete loci revealed a non-uniform distribution in the genome. This analysis showed clustering of approximately 50,331 (96.2%) piRNAs into only a relatively small number of loci (Supplementary Table S2). Overall, most chromosomes contained identifiable clusters (Fig. 2a and Supplementary Tables S2 and S3). At the depth of our present study, we found that the major clusters range in size from 10 to 83 kilobases and contain between 10 and 4,500 small RNAs. They generally occur in gene- and repeat-poor regions,

although rare expressed sequence tags (ESTs) from these loci were often derived from testis libraries.

As one example, chromosome 17 contains 3,885 clustered piRNAs as compared to 111 piRNAs that map as non-clustered orphans (Fig. 2b). A detailed view of one large cluster revealed an asymmetric distribution of piRNAs along each genomic strand (Fig. 2c). The cluster began proximal to the centromere with a 43,400-base-pair stretch containing 1,371 piRNAs along the upper strand and then switched abruptly to a 49,300-base-pair stretch containing 1,290 piRNAs along the lower strand. Additional bi-directional clusters were found, although the two arms were not always of equal length. Other clusters were found only on one strand (unidirectional, see Supplementary Fig. S3 and Supplementary Table S2). Indeed, virtually all piRNA clusters showed a profound strand asymmetry. Density maps of piRNAs along the remaining mouse chromosomes can be found in Supplementary Figs S3–S11.

Mature piRNAs were single-stranded, as verified by northern blotting for each strand of a selected piRNA (Fig. 2d). When piRNAs were detected from protein-coding genes, they almost invariably arose only from the sense strand of the mRNA. The piRNA-generating loci themselves were not obviously enriched for secondary structure. Although some clusters did arise from a complex series of direct or inverted repeats, these were in the minority. Within any given cluster, there appeared to be gaps in piRNA production. This was confirmed by probing northern blots with immediately adjacent sequences, one of which matched a piRNA recovered in our sequence analysis (Fig. 2e). Such gaps could reflect the fact that different piRNAs show substantial variation in abundance, with low-abundance piRNAs being absent from our sequencing and not detected by northern blotting. Considered together, our observations seem to be at odds with conventional mechanisms for producing small RNAs through Dicer- and Drosha-mediated cleavage of double-stranded RNA precursors. However, there are examples, such as target-dependent accumulation of siRNAs in *Caenorhabditis elegans*, where apparently conventional small RNA generation mechanisms can give rise to a severely asymmetric population (Grishok, A. and Mello, C. C., personal communication²¹).

By SYBR gold staining, we also detected potential piRNA populations in testes of other mammals, including human, rat and bull (not shown). However, only a small number of mouse piRNAs could be mapped to the genomes of other species. We therefore sequenced candidate piRNA populations from human and rat testes, obtaining 52,099 and 47,024 sequences, respectively. The human and rat piRNAs have the same defining characteristics as mouse piRNAs, being roughly 29–30 nucleotides in length, having a strong bias for U in the first nucleotide, and clustering in their respective genomes. Despite a lack of primary sequence conservation, bi-directional clusters of piRNAs were found on human chromosome 6 and rat chromosome 20 at loci that correspond to the mouse chromosome 17 cluster shown in Fig. 2c (Fig. 3a). Indeed, the vast majority of large mouse piRNA clusters existed at syntenic loci in human and rat, whereas smaller clusters were much less likely to be found at corresponding chromosomal positions in other species (Fig. 3b and Supplementary Tables S3 and S4).

The mouse testis contains several distinct cell types including germ cells, Sertoli cells and Leydig cells. Northern blotting indicated that piRNAs are present in germ cells, as piRNAs were absent from a *c-kit* mutant (W/W^v) that lacks germ cells and from a Sertoli cell line (TM4) (Fig. 4a). These RNAs were also lacking from the epididymis, suggesting an absence from maturing and mature sperm. The first wave of spermatogenesis initiates shortly after birth and occurs fairly synchronously throughout the organ. Therefore, distinct cell types appear at defined times postnatally. Thus, the timing of the appearance of piRNAs provides clues to the stages of spermatogenesis in which they might act (Fig. 4b). piRNAs first became visible at 14 days post partum, which corresponds to the time when cells enter the pachytene stage of meiosis. Abundance increased steadily to reach

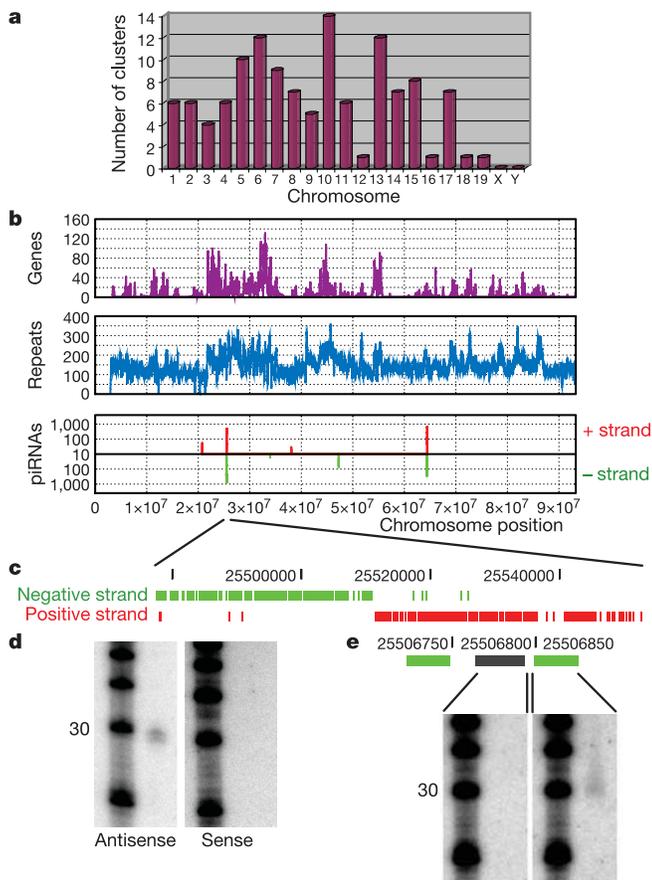


Figure 2 | Properties of the piRNA clusters. **a**, Absolute number of piRNA clusters per chromosome. **b**, Density analysis of exons (top), repeats (middle) and piRNAs (bottom, with positive strand in red and negative strand in green) along chromosome 17. Peaks correspond to clusters of 10 or more piRNAs. **c**, Strand bias of piRNAs. **d**, Strand specificity of piRNAs was analysed by northern blotting, using piR-1 sense and antisense probes. **e**, Non-uniform distribution of piRNAs across clusters shown by northern blotting using the frequently cloned piR-3 (green), and an uncloned, immediately adjacent region (black). Size (nucleotides) is shown along the left of panels **d** and **e**.

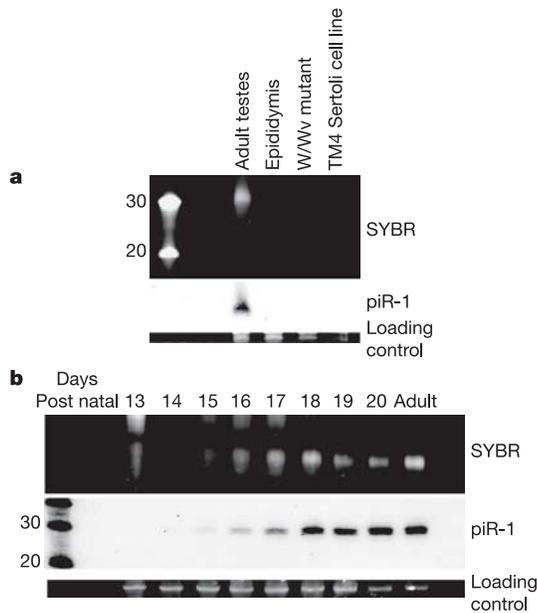


Figure 4 | piRNAs are germ cell specific and developmentally regulated. **a**, Germ cell specificity of the piRNAs. We examined the presence of piRNAs in somatic cells and sperm by SYBR gold staining and northern blot analysis using piR-1. **b**, Developmentally timed appearance of piRNAs. Testes were harvested during the first wave of spermatogenesis and assayed for the presence of piRNAs by SYBR gold staining and northern blotting using piR-1. Size (nucleotides) is shown along the left of each panel.

epigenetic fluidity, and mutations that disrupt the ability of cells to carry out epigenetic programming are known to have meiotic phenotypes^{22–24}. Finally, spermatogenesis involves dynamic regulation of both global protein synthesis and translation of individual mRNAs, and analogies to miRNA-mediated repression suggest possible roles for piRNAs in these processes. Comparative genomics has revealed conservation of the piRNA-producing capacity of a locus, rather than conservation of the piRNAs themselves. This may indicate that the key to the function of piRNAs may not be in their precise sequence. Instead, their function may stem from the fact that they are produced in large numbers from unique loci in the genome. Genetic studies of these complex RNA populations, combined with biochemical characterization of the complexes that contain them, will ultimately be required to fit this new addition to the small RNA world into our evolving picture of the biological roles of the Argonaute gene family.

METHODS

Total RNA was prepared using Trizol (mouse, rat) or purchased from Ambion (human). Small RNAs were separated by 15% denaturing polyacrylamide gel electrophoresis and stained with SYBR gold. Immunoprecipitations were carried out using a rabbit polyclonal antibody to MIWI after lysis of testis in RIPA or Triton buffer. Small RNA northern blots were carried out essentially as described previously²⁵. Probe sequences can be found in Supplementary Methods. Small RNA cloning was done as in ref. 18, and large-scale sequencing done as recommended by 454 Life Sciences. Detailed protocols and bioinformatics methods are available as Supplementary Information.

Received 18 February; accepted 18 May 2006.

Published online 4 June 2006.

- Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
- Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).

- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
- Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727 (1998).
- Lin, H. & Spradling, A. C. A novel group of *pumilio* mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* **124**, 2463–2476 (1997).
- Cox, D. N., Chao, A. & Lin, H. *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**, 503–514 (2000).
- Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9**, 315–327 (2002).
- Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303**, 669–672 (2004).
- Aravin, A. A. *et al.* Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. *Mol. Cell. Biol.* **24**, 6742–6750 (2004).
- Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11**, 1017–1027 (2001).
- Kalmykova, A. I., Klenov, M. S. & Gvozdev, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res.* **33**, 2052–2059 (2005).
- Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A. & Gvozdev, V. Telomere elongation is under the control of the RNAi-based mechanism in the *Drosophila* germline. *Genes Dev.* **20**, 345–354 (2006).
- Vagin, V. V. *et al.* The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*. *RNA Biol.* **1**, 51–55 (2004).
- Sarot, E., Payen-Groschene, G., Bucheton, A. & Pelisson, A. Evidence for a *piwi*-dependent RNA silencing of the *gypsy* endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. *Genetics* **166**, 1313–1321 (2004).
- Deng, W. & Lin, H. *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* **2**, 819–830 (2002).
- Kuramochi-Miyagawa, S. *et al.* *Mili*, a mammalian member of *piwi* family gene, is essential for spermatogenesis. *Development* **131**, 839–849 (2004).
- Ambros, V. & Lee, R. C. Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. *Methods Mol. Biol.* **265**, 131–158 (2004).
- Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
- Aravin, A. A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5**, 337–350 (2003).
- Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).
- Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99 (2004).
- Webster, K. E. *et al.* Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc. Natl Acad. Sci. USA* **102**, 4068–4073 (2005).
- Peters, A. H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
- Caudy, A. A., Myers, M., Hannon, G. J. & Hammond, S. M. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.* **16**, 2491–2496 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Sheth for help with the informatics analysis of piRNA populations, and W. R. McCombie for pilot sequencing. C. Cordon-Cardo and M. Golding provided tissue samples. We thank C. Mello, L. Murchison, S. Cheloufi and O. Tam for helpful discussions. This work was supported in part by grants from the NIH (G.J.H.). A.G. is a Florence Gould Fellow of the Watson School of Biological Sciences. G.J.H. is an Investigator of the Howard Hughes Medical Institute.

Author Information The GenBank accession numbers for piR-1, piR-2 and piR-3 are DQ539889, DQ539890 and DQ539891, respectively. Mouse piRNA accession numbers range from DQ539889 to DQ569912; human piRNA accession numbers range from DQ569913 to DQ601958; and rat piRNA accession numbers range from DQ601959 to DQ628526. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.edu).