Sciencexpress Report

Characterization of the piRNA Complex from Rat Testes

Nelson C. Lau,[1]* Anita G. Seto,[1]* Jinkuk Kim,[2,3] Satomi Kuramochi-Miyagawa,[4] Toru Nakano,[4] David P. Bartel,[3,5] Robert E. Kingston[1]†

[1]Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA. [2]Harvard-MIT Division of Health Sciences and Technology, E18-435, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [3]Howard Hughes Medical Institute and Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. [4]Department of Molecular Cell Biology, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamada-oka, Suita-shi, Osaka 565-0871, Japan. [5]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail:kingston@molbio.mgh.harvard.edu

Small noncoding RNAs regulate processes essential for cell growth and development, including messenger RNA degradation, translational repression, and transcriptional gene silencing (TGS). During a search for candidate mammalian factors for TGS, we purified a complex that contains small RNAs and Riwi, the rat homolog to human Piwi. The RNAs, frequently 29 to 30 nucleotides in length, are called Piwi-interacting RNAs (piRNAs), 94% of which map to 100 defined (<100 kb) genomic regions. Within these regions, the piRNAs generally distribute across only one genomic strand, or distribute on two strands but in a divergent, non-overlapping manner. Preparations of piRNA complex (piRC) contain rRecQ1, which is homologous to *qde-3* from *Neurospora*, a gene implicated in silencing pathways. Piwi has been genetically linked to TGS in flies, and slicer activity cofractionates with the purified complex. These results are consistent with a gene silencing role for piRC in mammals.

Gene-silencing pathways guided by small RNAs, essential for maintaining proper cell growth and differentiation, operate at either the transcriptional or posttranscriptional level (*1*). Posttranscriptional gene silencing acts through mRNA destabilization or inhibition of mRNA translation (*1*), whereas TGS represses gene expression by altering chromatin conformation (*2*). Each pathway utilizes a core complex containing small RNA associated with a member of the Argonaute (Ago) protein family; however, the different mechanistic needs of each pathway require differences in complex composition. Although RNA-mediated TGS has been studied in fission yeast and other eukaryotes (*2–5*), the mechanism of this process in mammals remains elusive.

To identify candidate complexes for TGS in mammals, we exploited the previous observations that TGS might utilize small RNAs longer than the 21 to 23 nucleotide microRNAs (miRNAs). In *Arabidopsis*, *Tetrahymena*, *Drosophila* and zebrafish, RNAs that are 24 nt and longer have been associated with TGS and/or genomic repeats, which are often silenced (*6–11*). In *Drosophila*, these repeat-associated small interfering RNAs (rasiRNAs) are enriched in testis (*6*, *12*). Therefore, we prepared extract from rat testes and fractionated it on an ion-exchange Q column, monitoring the small RNAs. A peak of small RNAs longer than a 22 nt marker eluted in mild salt conditions, which suggested the presence of a novel ribonucleoprotein complex (Fig. 1A).

To characterize the small RNAs, we sequenced cDNA libraries made from flowthrough and eluate fractions, obtaining 61,581 reads from the eluate that matched perfectly to the *Rattus norvegicus* genome (*13*). In contrast to the flowthrough RNAs, which were mostly miRNAs (69%), the eluate RNAs derived primarily from regions of the genome not previously thought to be expressed (Fig. 1A). Some eluate reads matched ESTs (11%), but only a small fraction matched annotated mRNAs (< 1.1%). Some also matched repeats (20%), but when considering that ~40% of the genome is annotated as repeats (*13*), the eluate reads were depleted in repeat sequences and thus as a class did not represent rasiRNAs.

The eluate RNAs were mostly 25 to 31 nt in length (Fig. 1A), and Northern analysis indicated a testis-specific expression pattern (Fig. 1B). Most eluate RNAs began with a 5' uridine (~84%), but no other sequence features or motifs were detected. A dominant subpopulation at 29 to 30 nt was observed (Fig. 1C), however these 29 to 30mers could not be distinguished from most of the remaining eluate reads by other criteria, including 5' nucleotide, genomic locus, or annotation. Thus, all the eluate RNAs that did not match annotated noncoding RNAs (miRNA, tRNA, rRNA, and

snRNA) were considered together as representing a single newly identified class of small RNAs.

To understand potential functions for these RNAs, we purified the associated proteins. By monitoring the RNAs, we developed a five-step scheme to purify the native complex to near homogeneity (Fig. 2, A and B). Mass spectrometry of the purified complex identified the rat homologs to Piwi (Riwi) and the human RecQ1 protein (Fig. 2 and fig. S1). Western blotting confirmed the copurification of Riwi and rRecQ1 with the small RNAs (Fig. 2C, but see also Fig. 4 below). We designate RNAs found with rat Piwi to be Piwi-interacting RNAs (piRNAs) and the complex to be the piRNA complex (piRC).

To gain insight into the origins of piRNAs, we examined the genomic loci from which they presumably derived. About two-thirds of the piRNA sequences each perfectly matched a single locus, and in some cases that unique locus was matched by multiple reads (up to 149). For the remaining one-third of the reads, which each mapped to multiple loci (up to 25,044 loci), we normalized the number of reads by the number of genomic hits and assigned this normalized hit count equally to all the loci; thus a piRNA read with four perfect genomic hits contributed a quarter of a count to each of its four loci. Counts were integrated into bins and plotted across each chromosome. The majority of counts (94%) fell into 100 genomic clusters that each contained at least 20 uniquely mapping reads (table S1). As exemplified by four clusters on chromosome 20 (Fig. 3A) and illustrated for all 100 (fig. S2), the clusters distributed across the genome, however some chromosomes were underrepresented in piRNA hits and clusters (fig. S2B). These clusters spanned 1 to 101 kb (table S1) and in aggregate comprised less than 0.1% of the rat genome.

Known silencing RNAs (siRNAs and miRNAs) derive from double-stranded RNA precursors or foldback structures (*1*). In contrast, piRNAs of most clusters mapped exclusively to either the plus or minus genomic strand, in irregular, sometimes overlapping patterns, with no evidence of extensive foldback structures or double-stranded origins (Fig. 3B). Sixteen clusters, such as cluster 1 (Fig. 3, B and C), contained regions of minus- and plus-strand hits that were juxtaposed with each other but separated by a gap of ~100 to 800 bp, an orientation suggesting divergent, bi-directional transcription, starting within the gap that separated the two distributions (table S1). Only two clusters had hits suggesting convergent or overlapping transcription (cluster 31 and 38, table S1). Northern analysis confirmed that piRNAs derived predominantly from one of the two genomic strands (Fig. 3D). RT-PCR results suggested that longer transcripts of the same polarity, perhaps piRNA precursors, also derived from these regions (fig. S3).

Analogous production of piRNAs from at least 94 clusters occurred in mouse, as indicated by the analysis of 68,794 piRNA reads generated in the same manner as those of rat (Fig. 3B, fig. S4, and tables S3 and S4). Most of the mouse clusters were homologous to rat clusters, with strikingly similar strand specificity and abundance profiles (Fig. 3B and tables S1 and S3). Nonetheless, their sequence conservation was low. Probes against rat clusters 4 and 6 hybridized only weakly to mouse piRNAs, as expected by the numerous point substitutions in the orthologous mouse piRNAs (Fig. 3D and fig. S5). Overall, the single-nucleotide substitution rate of the piRNA clusters was within the 15 to 20% expected for neutral residues (*13*). Nevertheless, residues represented by more reads had lower substitution and insertion/deletion rates, indicating detectable evolutionary pressure to conserve the sequence of the abundant piRNAs (Fig. 3E). We conclude that the production of piRNAs is highly conserved, but the sequence identities of the piRNAs are only weakly conserved. The weak conservation favors models in which piRNAs target the loci/transcripts that correspond to the same loci from which they derive.

We characterized two potential biochemical functions of piRC suggested by activities previously attributed to RecQ and Ago family members. Human RecQ1 is an ATP-dependent DNA helicase (*14*). Both ATPase and DNA unwinding activities followed the rRecQ1 protein of piRC (Fig. 4, A and B, and fig. S6). Riwi contains the catalytic residues that other Ago proteins use for RNA-guided cleavage of target RNAs (*15*) (fig. S7A). Using a substrate complementary to a piRNA, cleavage activity was detected, peaking with fractions containing Riwi and piRNAs (Fig 4D). However, it was not robust, perhaps because of the small representation of the cognate piRNA in the diverse population of piRNAs (< 0.2%).

Our purification of piRC uncovered a novel class of small RNAs and identified as copurifying factors Riwi and rRecQ1, two proteins with intriguing functions genetically determined in other species. Piwi represents a subclade of the Ago family of proteins (*16*), and was first discovered to regulate germ stem-cell maintenance in *Drosophila* (*17*). Subsequently, mammalian Piwi members were found to regulate germ cell maturation (*18*, *19*). Drosophila *piwi* mutants are also defective in small RNA-dependent transgene and retrotransposon silencing (*20*, *21*) and lose the inability to localize heterochromatic proteins, including the repressive Polycomb-group proteins (*22*, *23*). *Tetrahymena* Piwi (*TIWI*) is needed for siRNA-mediated DNA elimination (*24*).

In *Neurospora*, a screen for mutants in quelling (gene silencing during vegetative growth) identified both QDE-2, an Ago-family protein, and QDE-3, a RecQ1 homolog (*25*, *26*). When comparing RecQ homology in mammals and other organisms, *Neurospora* QDE-3 resided in the same clade as

rRecQ1 (Fig. 2D and fig. S7B). rRecQ1 did not always precisely cofractionate with Riwi and the piRNAs during our final purification step (Fig. 4A). The lack of tight association of rRecQ1 might have reflected conditions specific to this step, or might indicate that rRecQ1 is generally less tightly associated with piRNAs than is Riwi. Perhaps rRecQ1 is not critical for piRC function. However, the genetic links between the QDE-2 and QDE-3 silencing factors suggest that the biochemical association between Riwi and rRecQ1 has biological importance and furthermore implies a gene-silencing function for piRC. Addressing the functional role(s) of piRC, and understanding the genesis of the piRNA family, will be important questions in unraveling the role of piRC in the regulation of the genome.

References and Notes

1. P. D. Zamore, B. Haley, *Science* **309**, 1519 (2005).
2. M. A. Matzke, J. A. Birchler, *Nat. Rev. Genet.* **6**, 24 (2005).
3. I. M. Hall *et al.*, *Science* **297**, 2232 (2002).
4. A. Verdel *et al.*, *Science* **303**, 672 (2004).
5. T. A. Volpe *et al.*, *Science* **297**, 1833 (2002).
6. A. A. Aravin *et al.*, *Dev. Cell* **5**, 337 (2003).
7. P. Y. Chen *et al.*, *Genes Dev.* **19**, 1288 (2005).
8. A. Hamilton, O. Voinnet, L. Chappell, D. Baulcombe, *EMBO J.* **21**, 4671 (2002).
9. S. R. Lee, K. Collins, *Genes Dev.* **20**, 28 (2006).
10. K. Mochizuki, M. A. Gorovsky, *Genes Dev.* **18**, 2068 (2004).
11. D. Zilberman, X. Cao, S. E. Jacobsen, *Science* **299**, 716 (2003).
12. A. A. Aravin *et al.*, *Mol. Cell. Biol.* **24**, 6742 (2004).
13. R. A. Gibbs *et al.*, *Nature* **428**, 493 (2004).
14. S. Cui *et al.*, *J. Biol. Chem.* 278, **1424** (2003).
15. F. V. Rivas *et al.*, *Nat. Struct. Mol. Biol.* **12**, 340 (2005).
16. M. A. Carmell, Z. Xuan, M. Q. Zhang, G. J. Hannon, *Genes Dev.* **16**, 2733 (2002).
17. D. N. Cox *et al.*, *Genes Dev.* **12**, 3715 (1998).
18. W. Deng, H. Lin, *Dev. Cell* **2**, 819 (2002).
19. S. Kuramochi-Miyagawa *et al.*, *Development* **131**, 839 (2004).
20. A. I. Kalmykova, M. S. Klenov, V. A. Gvozdev, *Nucleic Acids Res.* **33**, 2052 (2005).
21. M. Pal-Bhadra, U. Bhadra, J. A. Birchler, *Mol. Cell* **9**, 315 (2002).
22. C. Grimaud *et al.*, *Cell* **124**, 957 (2006).
23. M. Pal-Bhadra *et al.*, *Science* **303**, 669 (2004).
24. K. Mochizuki, N. A. Fine, T. Fujisawa, M. A. Gorovsky, *Cell* **110**, 689 (2002).
25. C. Catalanotto, G. Azzalin, G. Macino, C. Cogoni, *Genes Dev.* **16**, 790 (2002).
26. C. Cogoni, G. Macino, *Science* **286**, 2342 (1999).
27. Supporting material is available on *Science* Online.

28. We thank N. Francis for initial work on this project; W. Johnston for technical assistance; L. Davidow, J. Morris, L. Lim, and J. Ruby for bioinformatics assistance; Z. Zhang and J. Goldman for performing preliminary assays; D. Schwarz for advice on slicer assays; C. Woo, E. Troemel, J. Song, and S. Aigner for comments on the manuscript. Supported by Helen Hay Whitney (N. L.) and Damon Runyon (A. S.) postdoctoral fellowships, a Korea Foundation for Advanced Studies predoctoral fellowship (J. K.), and grants from NIH (D. B. and R. K.). D. B. is a Howard Hughes Medical Institute investigator.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1130164/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Tables S1 to S6

18 May 2006; accepted 7 June 2006
Published Online 15 June 2006; 10.1126/science.1130164
Include this information when citing this paper.

Fig. 1. Testes contain a longer class of small RNAs. (**A**) Top section: Rat testes extract was fractionated on a Q column (0.1 to 1 M potassium acetate gradient). RNA from fractions was end-labeled and resolved on a gel. Bottom section: Small RNAs from column fractions were gel-purified (dashed boxes), converted to cDNAs, and sequenced. (**B**) Rat tissue Northern blot hybridized with body-labeled RNA probes corresponding to small RNA sequences. The blot was stripped before reprobing for the indicated chromosomal regions or let-7 miRNA (loading control). Migration of RNA markers is indicated (right). (**C**) Size distribution of small RNAs from flowthrough (white bars), and eluate (black bars). *Y*-axis scales are different for flowthrough RNAs (left) and eluate RNAs (right).

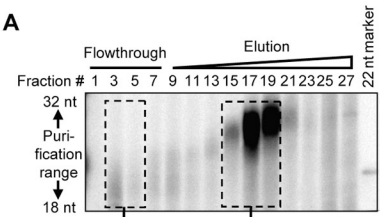
Fig. 2. Purification of a native small RNA-containing complex revealed Riwi and rRecQ1. (**A**) Schematic of piRC fractionation steps from rat testes extract. Numbers represent potassium acetate (KOAc) concentration (mM). (**B**) Proteins from the peak fraction of piRC (fract# 28, Superdex-200 column) were resolved on a gel and silver stained. Bands were excised and identified by mass spectrometry to be Riwi and rRecQ1 (See fig. S1) (**C**) Small RNAs, Riwi, and rRecQ copurify after five steps of chromatographic separation. Final Superdex-200 column fractions were assayed for the presence of small RNAs, Riwi, or rRecQ1. Top section: small RNAs were end-labeled and resolved on a gel. Middle and bottom sections: Western blots probed with anti-Miwi (mouse Piwi) and anti-hRecQ1 antibodies. Elution profile of protein size markers from Superdex-200 column indicated above.

(**D**) Phylogenetic comparison of RecQ DNA helicase family members revealed *Neurospora* QDE-3 to be a close homologue to rRecQ1.

Fig. 3. Genomic characteristics of rat piRNAs. (**A**) Chromosomal view indicating the number of piRNA reads mapping to clusters on Chr. 20 (see (27)). (**B**) Medium-resolution view of clusters 1, 15, and 26. Horizontal bars above and below the histograms indicate regions orthologous to mouse regions that also produce piRNAs (indicated if the bin matches > 2 uniquely mapping mouse piRNAs; table S1, S3). An asterisk denotes a group of piRNAs from cluster 15 that perfectly mapped to more than one locus in the genome. (**C**) High-resolution view centered on the gap region that separates minus- and plus-strand piRNA hits in cluster 1. Horizontal bars represent individual piRNAs. (**D**) Northern analysis with probes to the indicated clusters, testing strand-specific piRNA expression (left and middle) and cross-hybridization to mouse and human testes RNAs (right). Migration of RNA markers is indicated (far right). Blots were stripped and reprobed to let-7 miRNA (loading control). (**E**) piRNA conservation analysis. Orthologous rat and mouse clusters were identified and rat residues were binned based on the number of matching rat piRNA reads. The estimated substitution rate per residue (left) and estimated insertion/deletion rate (right), comparing rat to mouse, was calculated for each bin. Error bars indicate 95% confidence intervals of the estimates (*27*).

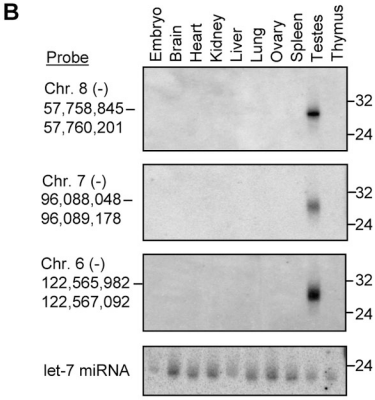
Fig. 4. piRC fractions contained ATP-dependent DNA helicase and slicer activities. (**A**) Visualization of proteins (top, silver stain) and end-labeled small RNAs (bottom) in fractions from final Superdex-200 column (independent purification from that shown in Fig. 2C). Size standards indicated on left. (**B**) Fractions containing rRecQ1 exhibit ATPase activity. Fractions were incubated with radiolabeled ATP. Free phosphate generated by ATPase activity was separated from unhydrolyzed ATP on thin layer chromatography. (**C**) Fractions containing rRecQ1 exhibited DNA unwinding activity. Fractions were incubated with a DNA substrate containing a 17 bp duplex. Reactions were resolved on a native gel. Appearance of the faster migrating radiolabeled 17 nt oligo indicated DNA duplex unwinding. (**D**) Fractions containing Riwi and piRNAs exhibit slicer activity. Cap-labeled substrates, complementary to piRNA or miRNA (negative control) sequences, and cleavage products were resolved on a gel.

A

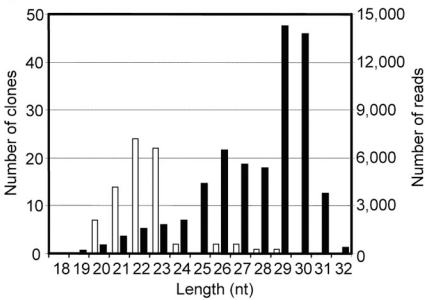


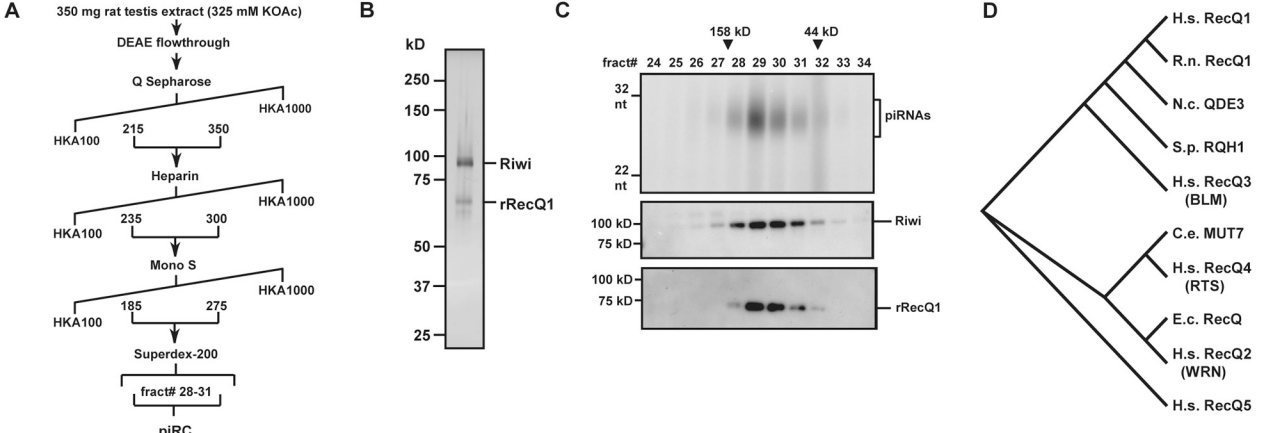
	Flow-through	Eluate
total clones/reads	75	61,581
No annotation	18%	65.8%
miRNAs	69%	1.8%
mRNA/EST	6.7%	12.1%
repeat associated	5.3%	19.9%
tRNA/rRNA/snRNA	1.3%	0.6%

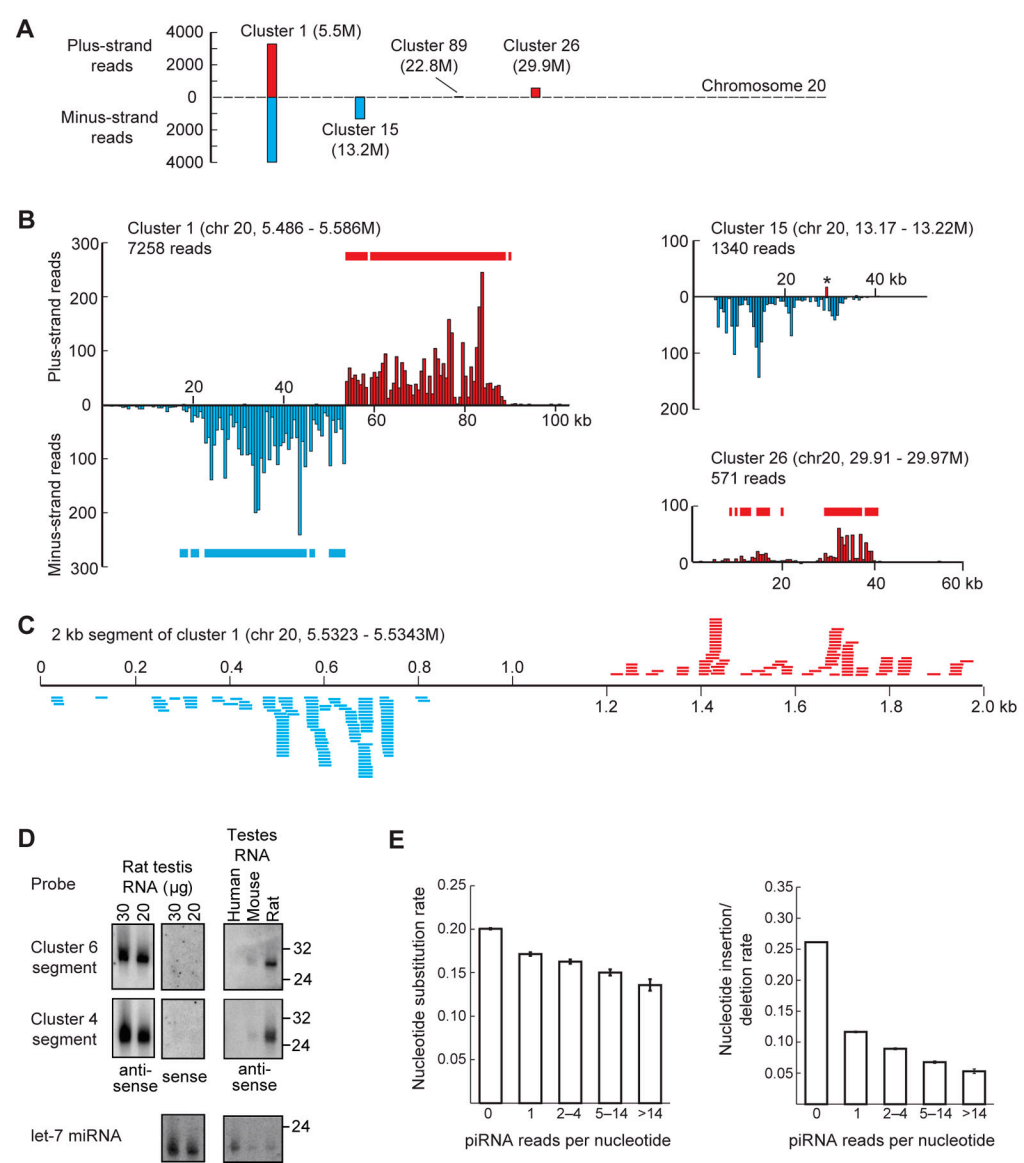
B

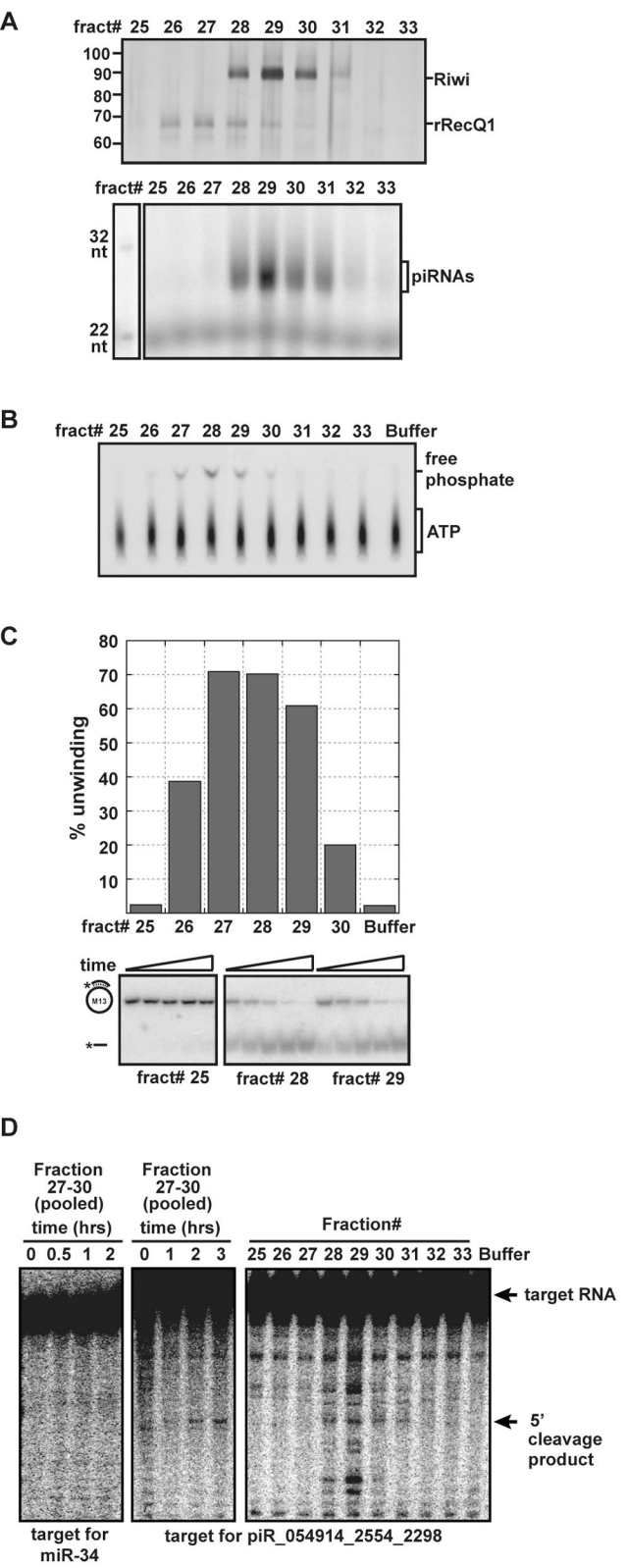


C




A
350 mg rat testis extract (325 mM KOAc)
DEAE flowthrough
Q Sepharose
HKA100
HKA1000
215
350
Heparin
235
300
Mono S
185
275
Superdex-200
fract# 28-31
piRC
B
kD
250
150
100
75
50
37
25
Riwi
rRecQ1
C
158 kD
44 kD
fract# 24 25 26 27 28 29 30 31 32 33 34
32 nt
22 nt
piRNAs
100 kD
75 kD
Riwi
rRecQ1
D
H.s. RecQ1
R.n. RecQ1
N.c. QDE3
S.p. RQH1
H.s. RecQ3 (BLM)
C.e. MUT7
H.s. RecQ4 (RTS)
E.c. RecQ
H.s. RecQ2 (WRN)
H.s. RecQ5


A
Plus-strand reads
Minus-strand reads
4000
2000
0
2000
4000
Cluster 1 (5.5M)
Cluster 89 (22.8M)
Cluster 26 (29.9M)
Cluster 15 (13.2M)
Chromosome 20
B
Cluster 1 (chr 20, 5.486 - 5.586M)
7258 reads
Plus-strand reads
Minus-strand reads
300
200
100
0
100
200
300
20
40
60
80
100 kb
Cluster 15 (chr 20, 13.17 - 13.22M)
1340 reads
100
0
100
200
20
*
40 kb
Cluster 26 (chr20, 29.91 - 29.97M)
571 reads
100
0
20
40
60 kb
C
2 kb segment of cluster 1 (chr 20, 5.5323 - 5.5343M)
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0 kb
D
Probe
Rat testis RNA (μg)
30
20
30
20
Testes RNA
Human
Mouse
Rat
Cluster 6 segment
Cluster 4 segment
32
24
32
24
anti-sense
sense
anti-sense
let-7 miRNA
24
E
Nucleotide substitution rate
0.25
0.20
0.15
0.10
0.05
0
1
2–4
5–14
>14
piRNA reads per nucleotide
Nucleotide insertion/deletion rate
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
1
2–4
5–14
>14
piRNA reads per nucleotide


A
fract# 25 26 27 28 29 30 31 32 33
100
90
80
70
60
Riwi
rRecQ1
fract# 25 26 27 28 29 30 31 32 33
32 nt
22 nt
piRNAs
B
fract# 25 26 27 28 29 30 31 32 33 Buffer
free phosphate
ATP
C
% unwinding
80
70
60
50
40
30
20
10
fract# 25 26 27 28 29 30 Buffer
time
M13
fract# 25
fract# 28
fract# 29
D
Fraction 27-30 (pooled)
time (hrs)
0 0.5 1 2
Fraction 27-30 (pooled)
time (hrs)
0 1 2 3
Fraction#
25 26 27 28 29 30 31 32 33 Buffer
target RNA
5' cleavage product
target for miR-34
target for piR_054914_2554_2298