www.sciencemag.org/cgi/content/full/1130164/DC1

Supporting Online Material for

Characterization of the piRNA Complex from Rat Testes

Nelson C. Lau, Anita G. Seto, Jinkuk Kim, Satomi Kuramochi-Miyagawa,
Toru Nakano, David P. Bartel, Robert E. Kingston*

*To whom correspondence should be addressed: kingston@molbio.mgh.harvard.edu

Published 15 June 2006 on *Science* Express
DOI: 10.1126/science.1130164

This PDF file includes:

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/1130164/DC1)

Tables S1 to S4 as zipped files

Supporting Online Material for

Characterization of the piRNA complex (piRC) from rat testes

Nelson Lau, Anita Seto, Jinkuk Kim, Satomi Kuramochi-Miyagawa, Toru Nakano, David Bartel, and Robert Kingston*

*To whom correspondence should be addressed. Ph: 617-726-5990; Fax: 617-643-2119; Email: kingston@molbio.mgh.harvard.edu

This PDF file includes:
Materials and Methods
References
Online text
Figures S1-S7
Tables S5-S6

Supplemental Online Materials

Materials and Methods:

Preparation of rat testes extract

Frozen trimmed rat testicles from adult Sprague-Dawley rats were purchased from Pel-Freez and were quick-thawed in washes of ice-cold 1X PBS. All subsequent steps were carried out on ice or at 4^{o} C. Thawed testicles were washed twice in Buffer A, and then minced in a Waring blender with 3 ml / 1 g tissue of Buffer A with six bursts of 6 seconds each. The homogenate was filtered through 3 layers of cheesecloth, and then centrifuged at 10,000 g for 10 minutes. The supernatant was set aside, and the pellet was resuspended in 0.5 volume / g tissue in Buffer B. The suspension was then dounced with 6 strokes of a type B pestle, and 0.5 volume / g tissue of Buffer C was added dropwise. The suspension was incubated with gentle rocking for 1 hr, dounced periodically to homogenize extract, and then centrifuged at 30,000 g for 1 hr. The supernatant was then dialyzed in Buffer D, and then centrifuged again at 16,000 g for 30 min. The final supernatant represented the rat testes extract and was flash frozen.

Sixty grams of testicles yielded ~50 ml of extract with a concentration ranging from 7-10 mg/ml of protein. Buffer D: 20 mM HEPES KOH, pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 0.2 mM PMSF, and 1.0 mM DTT. Buffer A: same as Buffer D but with 0.5 μg/ml leupeptin and 0.5 μg/ml aprotinin. Buffer B: same as Buffer A but with 20 mM KCl. Buffer C: same as Buffer A but with 1.2 M KCl.

Analysis of Small RNAs

Chromatography fractions were deproteinized by phenol/chloroform extraction, pH 6.6, and RNAs were precipitated with ethanol and glycogen carrier. RNAs were end-labeled with ^{32}P-cordycepin triphosphate and yeast poly-A polymerase (USB) according to the manufacturer's instructions. Small RNAs for cDNA libraries were extracted from column fractions with Tri-Reagent (Sigma). Northern blots were performed and cDNA libraries were constructed essentially as previously described (*1*). To prepare Q column eluate cDNA libraries for high-throughput pyrosequencing, the original cDNA libraries were diluted 200 fold into a PCR that was thermocycled for 12-16 cycles and that contained the forward primer: (GCCTCCCTCGCGCCATCAGTAT CGTAGGCACC TGAGA) and the reverse primer: (AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA /iSp18/GCCTTGCCAGCCCGCTCAGTATTGATGGTGCCTACAG). The asymmetric PCR products were then resolved on a 90% formamide, 8% polyacrylamide gel, and the longer DNA strands are gel purified and submitted to the 454 Life Sciences' Genome Sequencer 20 system (*2*).

Protein Chromatography and Analysis

Purification of piRC

All steps were performed on ice or at 4°C. Approximately 350 mg of protein from rat testes extract was brought to 325mM KOAc by the addition of HKA1000. 25 ml DEAE FF Sepharose resin (GE Healthcare) was equilibrated in HKA100. Rat testes extract was added to resin and bound with stirring on ice under vacuum for 2-3 hours. Slurry was poured into a column (20x3 cm) and supernatant flowed through by gravity.

Buffer HKA100 was added to keep column wet. 10 ml flowthrough fractions were collected until protein was no longer detectable by Bradford protein assay (Bio-Rad). To pooled DEAE flowthrough fractions, one volume of HKA0 buffer was slowly added while swirling. Flowthrough material was loaded onto HiLoad 16/10 Q Sepharose HP (GE Healthcare), equilibrated in HKA100. The bound material was eluted with a salt gradient from 0.1–1 M KOAc. Fractions were assayed for piRNAs by 3' end-labeling with polyA polymerase and ^{32}P-cordycepin triphosphate. The peak of RNA was pooled (eluted at approximately 215-350 mM KOAc). Pooled fractions were split in half and each was loaded onto a HiPrep 16/10 Heparin FF column (GE Healthcare). The bound material was eluted with a salt gradient from 0.1–1 M KOAc. The piRNAs eluted in a peak of approximately 235–300 mM KOAc. Peak fractions from both heparin columns were pooled and diluted with one volume of HKA0. This material was then loaded onto a Mono S 5/50 GL column (GE Healthcare), equilibrated in 100 mM KOAc. The bound material was eluted with a salt gradient from 0.1–1 M KOAc. The piRNAs eluted in 2 fractions of approximately 185–275 mM salt. The lower salt fraction was loaded onto a Superdex-200 10/300 GL column equilibrated and eluted in HKA200. 500 ul fractions were collected from Superdex-200 column immediately upon sample injection. HKA buffer: 30 mM HEPEs pH 7.5, 2 mM MgOAc, 10% glycerol, and 0–1 M KOAc.

Protein analysis

Five-20 μl of Superdex-200 column fractions were resolved on a 4-12% Bis-Tris PAGE gel run in MOPS buffer (Invitrogen). Gels were either stained with SilverQuest silver staining kit (Invitrogen) or transferred to PVDF membrane (Bio-Rad) for Western blotting. Blots were probed with anti-Miwi antibody (*3*) or anti-human RecQ1 antibody

(Bethyl BL2071) followed by anti-rabbit IgG (GE Healthcare), then visualized by ECL Plus (GE Healthcare). LC/MS/MS of bands excised from silver-stained gels was performed by the Taplin Biological Mass Spectrometry Facility (Harvard Medical School).

Activity assays

ATPase assays

γ-[^{32}P]-ATP was incubated with 4 mM $MgCl_2$, 4 μg plasmid DNA, and 3 μl of Superdex-200 fraction in a 5 μl reaction for 30 min. at 30 C. Reactions were quenched with 12.5 μl of stop solution (3% SDS, 100 mM EDTA, 50 mM Tris [pH 7.7]). Inorganic phosphate and ATP were separated on PEI-cellulose TLC plates using 0.5M LiCl and 1 M formic acid, and visualized on a PhosphorImager (GE Healthcare).

Helicase assays

[^{32}P]-5' end labeled 17-mer (5'-GTTTTCCCAGTCACGAC-3') was annealed to M13mp18 ssDNA as described (*4*). The dsDNA substrate was purified away from the excess 17-mer by successive purifications over four MicroSpin S-400 HR columns (GE Healthcare). Duplexed DNA substrate was incubated with 0.67 μl of Superdex-200 column fractions in a 10 μl reaction at 37°C as described (*4*). Reactions were stopped at indicated timepoints with 0.3% SDS, 10 mM EDTA, 5% glycerol and then resolved on a 12% nondenaturing polyacyramide gel. Gels were dried, visualized, and quantitated on a PhosphorImager (GE Healthcare).

Slicer assays

RNA cleavage substrates were transcribed from DNA oligos to form the following RNAs: 5'-GGAACCGAGCUC-[antisense sequence to piRNA]-AGCUAGCAACC-3'. Cleavage substrates were cap-labeled and utilized in slicer reactions essentially as described (*5*), except that 10 μl of purified fractions were mixed 1:1 with 10 μl 2X reaction components containing 8 mM $MgCl_2$, and incubations were performed at 35°C.

Computational analysis of piRNAs

Processing and annotation of large-scale sequencing reads

High-throughput sequencing of the eluate cDNA library from rat and mouse testes extract yielded 99,753 and 105,793 raw sequencing reads, respectively. After filtering out reads that did not match 5' and 3' linker sequences, reads that contained an ambiguous base ('N'), reads with lengths outside of the gel-purification size range (18-32 nt), or reads matching size marker RNAs used in library construction, 85,489 rat and 95,423 mouse reads remained. Because some sequences represented more than one read, these corresponded to 61,293 unique small RNA sequences in rat, and 65,681 in mouse.

For the sequences of each species, WU-BLAST (parameters: nogaps, E=0.01, W=[length of a sequencing read], hspsepSmax=0, hspmax=60000, B=60000) (*6*) was used to find matches to: (a) the mammalian (human/mouse/rat/dog) miRNA hairpin sequences registered at mirBase (*7*), (b) the cluster of rat or mouse 18S, 5.8S, and 28S rRNA sequences (accession number: V01270 for rat, J01871, X00686 and X00525 for mouse), and (c) the *Rattus norvegicus* genome (build rn3) or *Mus musculus* genome

(build mm7) (*8*). As a result, 40,698 unique sequences representing 61,581 reads of rat and 43,332 sequences representing 68,794 reads of mouse were confirmed to derive from the rat and mouse genome respectively. For each unique sequence, the number of genomic hits was counted (table S2, S4).

We annotated the unique sequences with determinable genomic coordinates according to a hierarchical manner that classified RNAs into specific groups. First, sequences that perfectly matched to microRNA hairpin sequences were classified as 'miRNA'. Second, the remaining unique sequences that perfectly matched rRNA cluster sequences were classified as 'rRNA'. Third, the remaining unique sequences not classified as miRNA nor rRNA were classified as 'rmsk' if they mapped to at least one locus from the RepeatMasker (*9*) annotation tracks that were downloaded from the UCSC genome browser (*10*).

To analyze the repeat-associated sequences (classified as 'rmsk') in greater detail, we examined the classifications of this sequence group amongst the 15 subclasses in RepeatMasker annotations (5S rRNA, tRNA, snRNA, scRNA, srpRNA, LTR, LINE, SINE, Satellite, Low_complexity, Simple_repeat, DNA, RNA, Other and Unknown), which include not only repeat elements, but also non-coding RNA species. Many of these sequences hit multiple loci, of which some loci carry a RepeatMasker annotation, while other loci may lack any annotation. For example, a sequence can hit one locus that has no annotation, and it can hit a second locus that is annotated as both a tRNA and a LTR. To minimize bias and improve consistency in classification, we would assign classifications in such an example by first evenly distributing an identity score (100%) to each of the two loci, giving a "score" of 50% to each locus. For the locus that was

annotated as a tRNA and LTR, the identity score was further divided, so that the tRNA and LTR classifications each would receive a 25% identity score. So, if this sequence was read 10 times, a score of 5 (= 10*25/50) was assigned to "tRNA", and another score of 5 was assigned to "LTR". Annotations on both sense and antisense strands were considered.

Finally, the remaining sequences that were not yet classified were examined for coordinate overlap with the coordinates of mRNAs and ESTs. The genomic coordinates of mRNAs and ESTs were downloaded from UCSC genome browser, and sequences yet containing at least one locus annotated by mRNA or EST were classified as 'mRNA' and 'EST', respectively. However, rat mRNA sequences whose accession number started with 'DQ' were disregarded because these represented a large group of annotation in which the mRNAs did not derive from the annotated loci (their annotated exons had multiple mismatches to the cDNA sequences and their annotated introns did not have canonical splicing sites).

Detecting motifs

The 40,698 unique rat sequences that did not represent fragments of rRNA or other annotated non-coding RNAs were divided into subset, based on a common length or annotation. Each subset was examined for the presence of sequence motifs. Every subset strongly exhibited a 5' U at the first nucleotide, but no other significant enrichment of motifs was detected from any subsets. The 5-nt long sequences immediately upstream of the 5' end of the mapped loci were collected and examined for the presence of sequence motifs for piRNA processing. This search did not result in any significant enrichment of motifs regardless of the density of reads.

Calculating genomic proportions of each annotation class

The aggregate proportions of the genome that are comprised of RepeatMasker annotations, mRNAs, and ESTs were calculated in a hierarchical manner. Genomic regions annotated by RepeatMasker were determined first, and proportions of regions covered by different RepeatMasker subclasses were further analyzed in a manner analogous to the repeat subclass analysis of the small RNA sequences. After determining the repetitive proportion of the genome, the proportions of the genome comprised of mRNAs and ESTs were determined in succession. Calculations were based on number of bases annotated divided by number of bases in the genome.

Detecting piRNA clusters

Clusters were identified by scanning a 20 kb window off-set by 1 kb across chromosomes and detecting genomic regions where at least 20 or more normalized hit counts were mapped. When qualified genomic regions were first encountered, the right hand side boundary of the window was extended progressively further in 1 kb intervals until the total normalized hit counts within 20 kb dropped below 20.

The identities and boundaries of some putative and bonafide clusters were confounded by piRNAs that mapped to several genomic loci, making it difficult to assess the defined origin of these particular sequences. Thus, we disregarded genomic regions in boundaries or putative small cluster where the total number of unique reads was less than 20. Then, both the left and right hand side boundaries were, independently, fine-tuned by trimming them progressively by 1 kb at a time toward the center of the cluster until the normalized hit counts within 1 kb became at least 1.

piRNA clusters were defined into four types (divergent, plus-strand, minus-strand, and mixed) by the following algorithm. Each clusters was scanned first on the plus-strand (from the left boundary to the right boundary) and sequentially on the minus-strand (from the right boundary to the left boundary) for 5 consecutive loci where reads were mapped uniquely. Searches that identified 5 consecutive loci only from one strand in a cluster logically classified the cluster as either a plus-strand or minus-strand type. If 5 consecutive loci were identified on both the plus- and minus-strand searches, the algorithm determined whether plus-strand loci were located downstream of minus-strand loci. Such a cluster would then be classified as a divergent type and the distance between the two plus- and minus-strand loci found by each search was calculated as a gap. In other cases, the cluster is classified as mixed type. This procedure identified exactly 100 clusters from rat and 94 from mouse and the clusters were ranked by the total normalized hit counts within each cluster. In total, the 100 rat clusters contained 56,738 normalized hit counts, which was 94% of total reads, and they covered 2,733 kb, which was less than 0.1% of the genome. Similarly, the 94 mouse clusters contained 62,466 normalized hit counts, which was 93% of total reads, and they covered 2,489 kb, which was less than 0.1 % of the genome.

Displaying the spatial distribution of sequencing reads

For each unique sequence, we normalized the number of reads by the number of genomic hits and assigned this normalized hit count equally to all the loci. For the whole chromosome view plot (Fig. 3A, fig. S2, S4), the hit counts were integrated into 1 Mb bins based on the start position of their loci and plotted across each chromosome. For the cluster view plot (table S1, S3), the cluster region was divided into 150 equal-sized bins,

the hit counts were integrated into the bins, and plotted across the cluster region. However, for the three cluster view plot in Fig. 3B, in order to compare the number of reads in a bin across three clusters, fixed-size (600 nt) bins are used regardless of size of clusters. For divergent type clusters, bins were defined so that a bin boundary can lie within the gap. For the 1 nt resolution plot, the hit counts were rounded up to the closest integer and this number was used to determine the number of horizontal bars, which were duplicated accordingly and plotted in a stacked representation on the corresponding locus (Fig. 3C, fig. S3, S5).

Conservation analysis of the piRNAs in rat and mouse

In order to determine the mouse chromosomal regions orthologous to rat chromosomal regions and to estimate the primary sequence conservation between the two regions, we downloaded rat centric rat-mouse pairwise alignments from UCSC genome browser. We then used a 30-nt window to scan along a rat piRNA cluster region of interest by 1nt offsets and counted the number of conserved residues. The conservation levels ([# conserved residues] / 30) were plotted across rat piRNA cluster regions (Fig. S5).

Estimation of substitution and insertion/deletion rate was performed as follows. For each residue within 100 rat clusters, we calculated the number of reads that uniquely mapped to that residue (and disregarded reads that mapped to more than one locus). The residues were divided into 5 groups based on the calculated number of reads: (1) 0 reads (2) 1 reads (3) 2-4 reads (4) 5-14 reads (5) more than 14 reads. For each group, the substitution rate was calculated as (total number of substituted residues) divided by (total number of residues aligned without gap). The insertion/deletion rate was calculated as

(total number of residues missing in mouse) divided by (total number of residues). The calculated substitution rate was adjusted with Jukes-Cantor multiple hit correction. Estimation of the 95% confidence interval was performed as described in (*11*).

To examine the conservation of piRNAs between rat and mouse, for all the mouse reads which uniquely mapped to the genome, we determined the rat loci that were orthologous to the mouse loci where mouse piRNA reads were mapped by parsing mouse centric mouse-rat pairwise alignments. The mouse reads were binned by their calculated rat loci and such bins that matched more than 2 mouse reads were depicted above and below the histograms in rat cluster view plots (Fig. 3B. table S1). Similarly, rat reads were mapped on mouse cluster view plots (table S3).

References:

S1. N. C. Lau, L. P. Lim, E. G. Weinstein, D. P. Bartel, *Science* **294**, 858 (2001).

S2. M. Margulies *et al.*, *Nature* **437**, 376 (2005).

S3. S. Kuramochi-Miyagawa *et al.*, *Development* **131**, 839 (2004).

S4. S. Cui *et al.*, *J Biol Chem* **278**, 1424 (2003).

S5. J. Martinez, A. Patkaniowska, H. Urlaub, R. Luhrmann, T. Tuschl, *Cell* **110**, 563 (2002).

S6. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J Mol Biol* **215**, 403 (1990).

S7. S. Griffiths-Jones, R. J. Grocock, S. van Dongen, A. Bateman, A. J. Enright, *Nucleic Acids Res* **34**, D140 (2006).

S8. R. A. Gibbs *et al.*, *Nature* **428**, 493 (2004).

S9. A. Smit, Hubley, R & Green, P. (1996-2004).

S10. A. S. Hinrichs *et al.*, *Nucleic Acids Res* **34**, D590 (2006).

S11. J. L. Devore, in *Probability and statistics for engineering and the sciences*. (Duxbury, 2000) pp. 291.

S12. GenBank accession numbers are DQ684678 to DQ727400 for our mouse piRNAs and DQ727401 to DQ767590 for our rat piRNAs .

Supporting text:

Sequence complexity of the rat small RNA library from the eluate of the crude purification of rat testes extract.

Regarding the complexity of the high-throughput pyrosequenced library, we believe we have not approached saturation of the library, and we expect the potential breadth of different piRNA sequences to be very wide due to the irregular, overlapping patterns of the piRNAs within the cluster regions. For reads that were mapped to rat genomic sequences (61,581 reads), singleton and doubleton reads comprised ~81% and ~6%, respectively.

Supplemental Table legends and tables:

Supplemental Table S1. One hundred piRNA clusters from the rat genome (build rn3). [Expand archive file TableS1_RatClusterRank.zip. Open HTML file: TableS1_RatClusterRank.html (note: this archive contains a directory of images linked to the HTML table. Expanded size: 8.4 MB)]

Supplemental Table S2. Unique rat small RNA sequences (61,294 in total) that derived from the eluate of a Q column fractionation of rat testes extract. Note that some sequences in this table do not have perfect matches to the rat genome.
[Expand archive file: TableS2_RatUniqSeq.zip. Open Excel file: TableS2_UniqSeq.xls (Expanded size: 9.4 MB)]

Supplemental Table S3. Ninety-four piRNA clusters from the mouse genome (build mm7).
[Expand archive file TableS3_MouseClusterRank.zip. Open HTML file: TableS3_MouseClusterRank.html (note: this archive contains a directory of images linked to the HTML table. Expanded size: 7.9 MB)]

Supplemental Table S4. Unique mouse small RNA sequences that were derived from the eluate of a Q column fractionation of mouse testes extract. The first 65,536 unique sequences are detailed in this file, representing all reads mapped to the mouse genome. The last 145 unique sequences that could not be added to this file due to the file size limitation of the Excel program are reads that do not match to the mouse genome.
[Expand archive file: TableS2_MouseUniqSeq.zip. Open Excel file: TableS2_UniqSeq.xls (Expanded size: 10.0 MB)]

Supplemental fig. S1. LC/MS/MS identification of Riwi and rRecQ1

A Identification of Riwi protein (accession# XP_344106). Sequences of polypeptides found by mass spectrometric analysis of gel-excised band from purification shown in Fig. 2C. Bottom section, schematic of peptide coverage across Riwi.

coverage: 43.0% (by amino acid count)

Amino acid number	Peptide
63-80	SQELQISAGFQELSLAER
85-95	RDFHDLGVNTR
106-114	TGSSGIIVK
122-144	LTSRPQWALYQYHIDYNPLMEAR
148-160	SALLFQHEDLIGR
190-198	VTEVFSQTR
234-242	IMNLQQIGR
243-257	NYYNPSDPIDIPNHR
288-313	SETVLDFMFNLYQQTEEHKFQEQVSK
314-323	ELIGLIVLTK
328-341	TYRVDDIDWDQNPK
360-379	KQYNQEITDLKQPVLVSQPK
410-423	NDFNVMKDLAVHTR
436-448	LIDYIHKDDNVQR
449-467	ELRDWGLSFDSNLLSFSGR
480-492	TFDYNPQFADWSK
496-514	GAPLISVKPLDNWLLIYTR
515-529	RNYEAANSLIQNLFK
541-555	AIMIEVDDRTEAYLR
624-630	MGGELWR
654-669	RSIAGFVASINEGMTR
680-688	GQELVDGLK
760-767	FFAQSGGR
768-797	LQNPLPGTVIDVEVTRPEWYDFFIVSQAVR
850-869	LAFLVGQSIHREPNLSLSNR

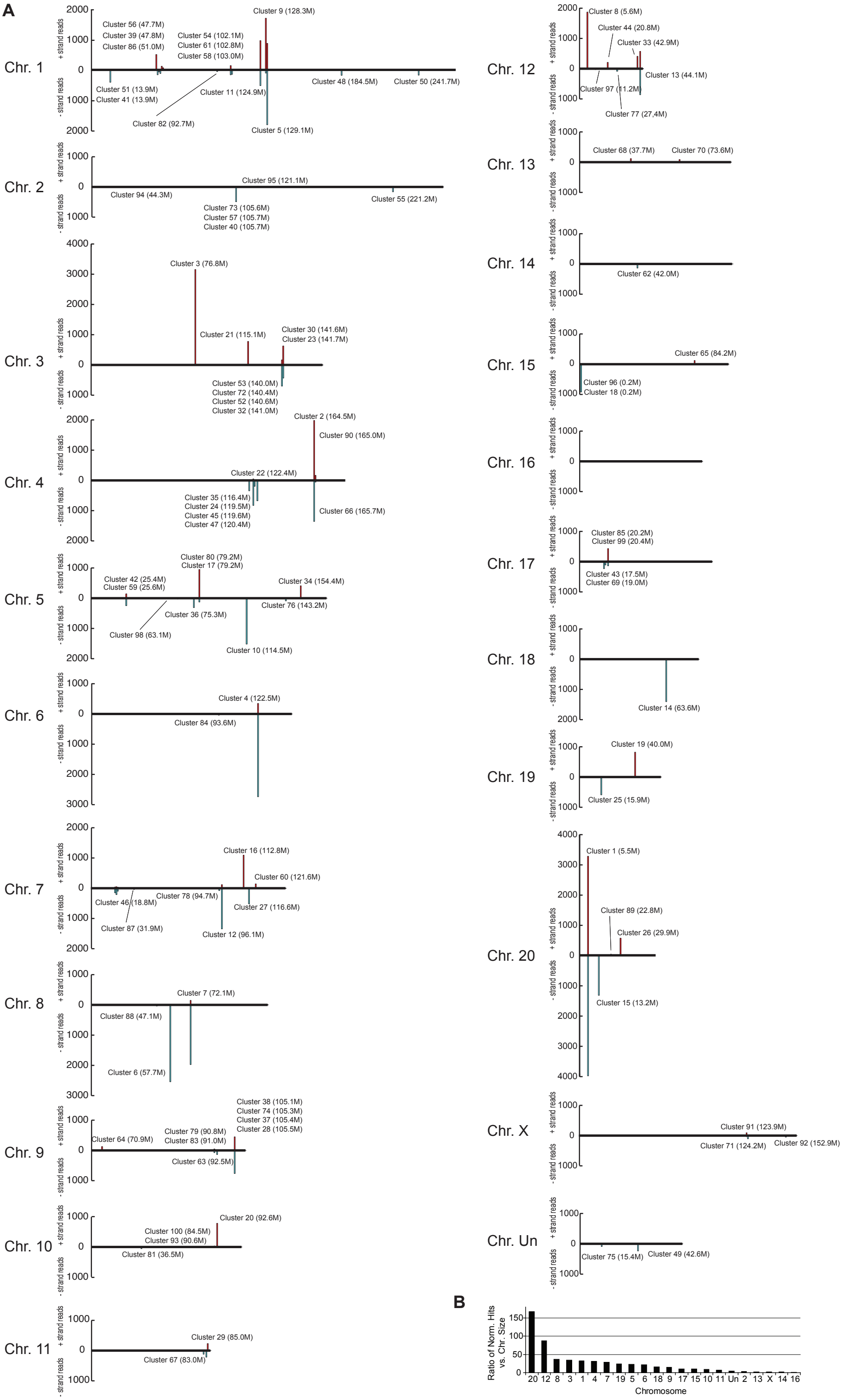
B Identification of rRecQ1 protein (accession# XP_575714). Sequences of polypeptides found by mass spectrometric analysis of gel-excised band from purification shown in Fig. 4A. Bottom section, schematic of peptide coverage across rRecQ1.

coverage: 33.4% (by amino acid count)

Amino acid number	Peptide
92-106	FRPLQLETVNATMAR
108-119	DIFLVMPTGGGK
186-193	LIYVTPEK
243-263	RQFPNISLIGLTATATNHVLK
292-311	QKPSSAEDFIENIANLINGR
326-336	DSEQVTISLQK
356-384	VHTQWSANELQVVVATVAFGMGIDKPDVR
394-404	SMENYYQESGR
426-439	ISSMVVMENVGQQK
515-522	LIDAWMGK
529-544	VAGVAVPALPREDLEK
545-556	IIVHALLQQYLK
557-571	EDYSFTAYATISYLK
576-591	ASLLSNEGHAVTMQVK

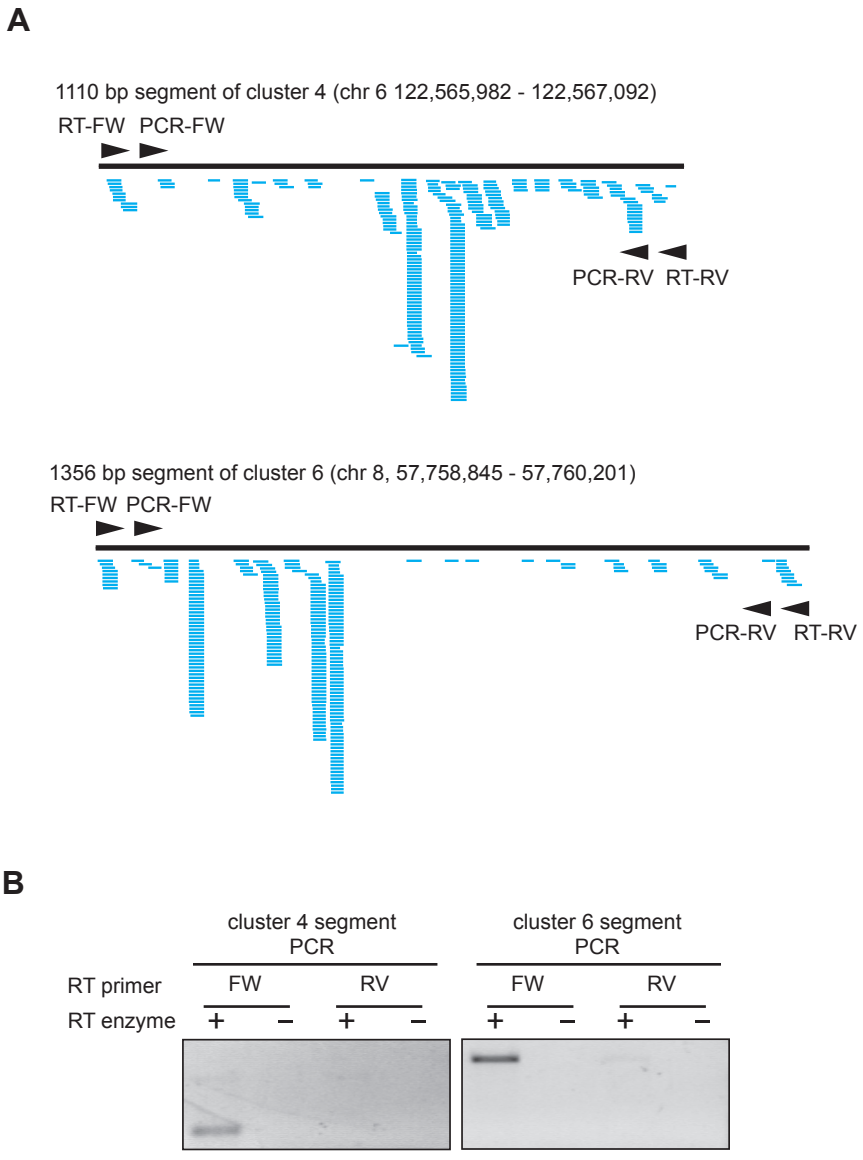
Supplemental fig. S2. Map of piRNA Clusters on the Rat Genome

(A) Graphical histograms represent the relative scaled lengths of the 20 autosomes, the X sex chromosome, and an artificial conglomerate of unassembled contigs called Chr. Unknown. Vertical bars in red and blue represent bins that tally plus and minus strand reads, respectively. One hundred rat piRNA clusters and their genomic location are labeled (see Table S1). (B) Ranking of chromosomes that contain the most piRNA hits relative to the size of the chromosome. The ratio is calculatd as the number of normalized hits on each chromosome multiplied by 10^-6 and dividing by the chromosome size.



Supplemental Fig. S3. piRNA cluster transcripts appear to derive from a single genomic strand.

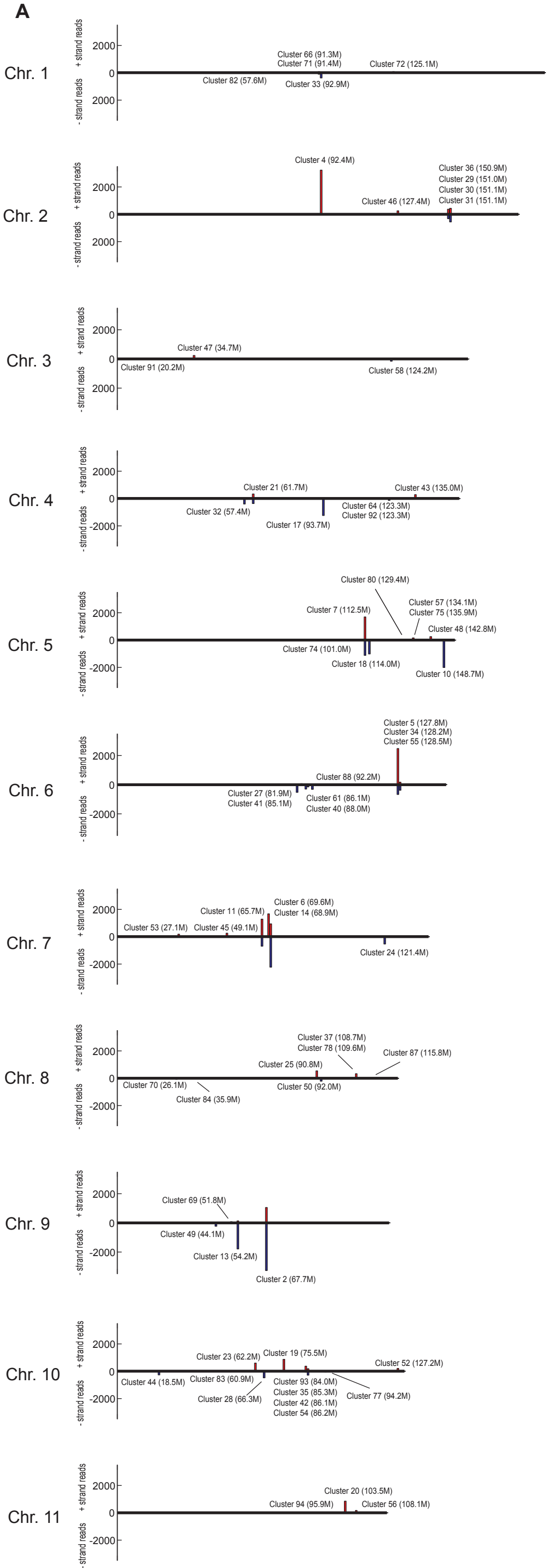
(A) High resolution schematic of segments of cluster 4 and 6 from rat. (B) PCR products generated from internal primers to rat cluster 4 and 6 are detected only for a single transcript that corresponds to the mapped piRNAs. All RT primers were validated for their ability to hybridize by the generation of effective constructs for Northern blot probes.



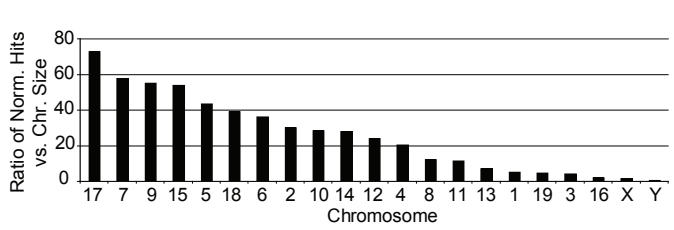
Supplemental fig. S4. Map of piRNA Clusters on the Mouse Genome

(A) Graphical histograms represent the relative scaled lengths of the 19 autosomes and the X and Y sex chromosome. Vertical bars in red and blue represent bins that tally plus and minus strand reads, respectively. Ninety-four piRNA clusters and their genomic location are labeled (see Table S3). B) Ranking of chromosomes that contain the most piRNA hits relative to the size of the chromosome. The ratio is calculatd as the number of normalized hits on each chromosome multiplied by 10^-6 and dividing by the chromosome size.

A

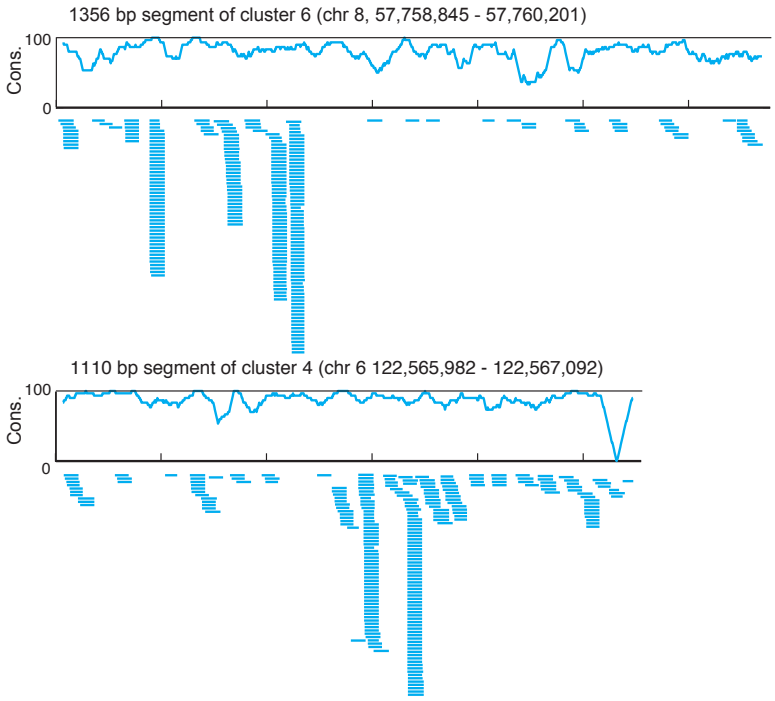


B



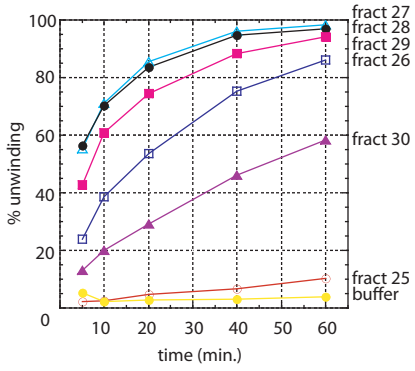
Supplemental fig. S5. The primary sequences of piRNAs are poorly conserved

Mapping and conservation analysis of cluster 6 and 4 segments. Line plots above the molecule maps of piRNAs examine the degree of primary sequence conservation between the rat and mouse syntenic regions. The jagged blue line marks the conservation percentage within 30-nt windows sliding 1-nt across. The patterns of individual piRNAs in the clusters do not strongly correlate with more highly conserved regions.



Supplemental fig. S6. Timecourse shows peak of helicase activity to reside in fractions 27-28 of Superdex-200 column

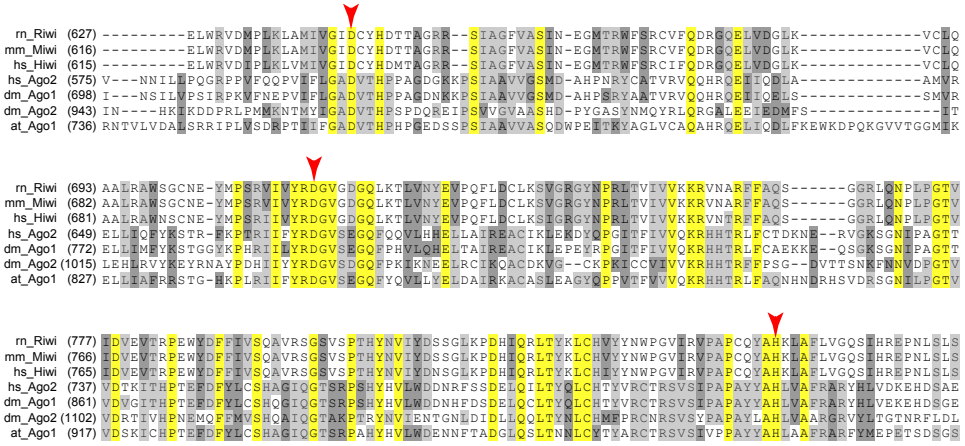
Fractions from Superdex-200 column were incubated with a DNA substrate containing a 17-bp duplex. Timecourse of unwinding of the duplex is shown.



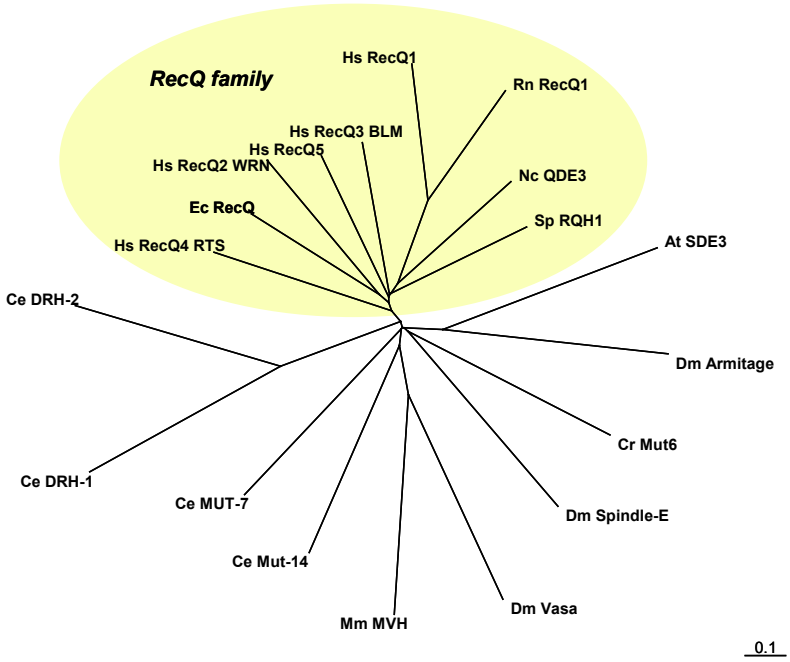
Supplemental fig. S7. Sequence comparisons of Riwi and rRecQ1 homologs

(A) Alignment of the Piwi domains from Riwi, Miwi, and Hiwi with other Argonaute-family members that have been demonstrated to exhibit Slicer activity. Arrowheads mark the invariant DDH catalytic triad essential for Slicer activity. (B) Phylogenetic tree comparing RecQ family members and other helicases with genetic links to RNAi. Accession numbers of protein sequences are: at_Ago1_NP_849784, at_SDE3_NP_172037, ce_DRH1_NP_501018, ce_DRH2_NP_501019, ce_Mut14_AAA96106, ce_MUT7_NP_499105, cr_Mut6_AAG33228, dm_Ago1_NP_725341, dm_Ago2_NP_648775, dm_Armitage_AAT12000, Dm_SpindleE_NP_476741, Dm_Vasa_NP_723899, ec_RecQ_NP_756603, hs_Ago2_Q8CJG0, hs_Hiwi_AAC97371, hs_RecQ1_AAH01052, hs_RecQ2_WRN_Q14191, hs_RecQ3_BLM_P54132, hs_RecQ4_RTS_O94761, hs_RecQ5_O94762, mm_Miwi_NP_067286, Mm_MVH_Q9NQI0, nc_QDE3_AAF31695, rn_RecQ1_AAH79026, rn_Riwi_XP_344106, sp_RQH1_CAA70577.

A



B



Supplemental Table S5. Annotation of the one hundred piRNA clusters from the rat genome

Rank	Genomic coordinate	Length	LTR	LINE	SINE	Satellite	Low complexity	Simple repeat	DNA	Unknown	mRNA	EST	No annotation	Total # normalized reads
1	chr20:5480000-5581000	101000	5.64%	1.74%	2.89%	0.00%	0.06%	1.36%	2.31%	0.00%	0.00%	14.43%	71.58%	7257.72
2	chr4:164469000-164546000	77000	6.03%	11.22%	8.01%	0.00%	0.26%	0.47%	0.21%	0.00%	0.00%	0.18%	73.63%	3315.75
3	chr3:76817000-76892000	75000	2.61%	3.34%	8.29%	0.00%	0.00%	0.10%	0.00%	0.00%	0.03%	1.55%	84.08%	3161.93
4	chr6:122540000-122624000	84000	7.93%	3.95%	7.41%	0.00%	0.03%	0.03%	0.58%	0.00%	2.95%	0.03%	77.07%	3082.76
5	chr1:129070000-129164000	94000	6.02%	4.38%	1.71%	0.00%	0.00%	0.93%	1.60%	0.00%	0.00%	0.00%	85.36%	2689.82
6	chr8:57725000-57777000	52000	9.56%	4.60%	17.93%	0.00%	0.19%	0.31%	0.00%	0.00%	0.29%	0.00%	67.11%	2564.28
7	chr8:72069000-72129000	60000	3.81%	7.62%	7.37%	0.00%	0.05%	1.13%	1.14%	0.05%	0.00%	6.87%	71.96%	2126.17
8	chr12:5597000-5633000	36000	7.20%	3.52%	3.64%	0.00%	0.00%	0.64%	0.00%	0.00%	0.00%	84.99%	0.00%	1873.2
9	chr1:128332000-128392000	60000	3.75%	5.11%	4.38%	0.00%	1.57%	0.36%	1.19%	0.00%	0.00%	0.00%	83.64%	1810.17
10	chr5:114521000-114575000	54000	4.38%	3.42%	0.72%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	91.48%	1519.49
11	chr1:124891000-124959000	68000	3.09%	8.76%	1.88%	0.00%	0.00%	0.40%	0.13%	0.00%	0.03%	0.47%	85.23%	1488.88
12	chr7:96057000-96116000	59000	4.24%	6.04%	13.73%	0.00%	0.07%	0.07%	3.90%	0.00%	0.00%	0.00%	71.95%	1460.01
13	chr12:44134000-44189000	55000	32.32%	1.50%	2.59%	0.00%	0.56%	0.07%	0.02%	0.00%	0.00%	0.42%	62.51%	1430.33
14	chr18:63622000-63658000	36000	5.23%	8.52%	1.78%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	84.47%	1404.66
15	chr20:13186000-13223000	37000	4.50%	2.28%	9.96%	0.00%	0.04%	0.52%	0.15%	0.52%	0.41%	12.61%	69.00%	1339.81
16	chr7:112772000-112797000	25000	8.53%	0.00%	2.51%	0.00%	0.00%	0.28%	0.00%	0.00%	0.00%	0.00%	88.68%	1089.84
17	chr5:79245000-79257000	12000	11.30%	0.00%	4.01%	0.00%	0.00%	1.29%	0.00%	0.00%	0.00%	0.00%	83.40%	1009
18	chr15:218000-262000	44000	7.86%	5.28%	2.02%	0.00%	0.56%	0.50%	0.00%	0.00%	0.00%	0.45%	83.33%	893.43
19	chr19:40002000-40017000	15000	0.00%	13.01%	24.80%	0.00%	0.00%	0.43%	9.85%	0.00%	0.00%	5.47%	46.44%	822.51
20	chr10:92629000-92654000	25000	19.27%	0.00%	3.82%	0.00%	0.13%	0.13%	0.00%	0.00%	0.00%	0.19%	76.47%	784.67
21	chr3:115118000-115142000	24000	0.69%	0.26%	1.89%	0.00%	0.00%	7.25%	2.44%	0.00%	0.00%	42.19%	45.28%	777.42
22	chr4:122403000-122440000	37000	15.10%	25.37%	5.36%	0.00%	0.00%	0.15%	0.45%	0.15%	0.00%	1.34%	52.08%	672.04
23	chr3:141656000-141670000	14000	0.00%	0.00%	1.11%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.48%	98.41%	628
24	chr4:119527000-119539000	12000	0.00%	0.00%	0.16%	0.00%	0.00%	16.95%	0.00%	0.00%	0.00%	3.79%	79.10%	619.5
25	chr19:15907000-15945000	38000	17.85%	3.37%	1.52%	0.00%	0.00%	0.00%	0.51%	0.00%	0.00%	0.00%	76.77%	594.02
26	chr20:29911000-29965000	54000	8.41%	0.70%	11.95%	0.00%	1.31%	0.70%	2.10%	0.00%	0.26%	0.00%	74.57%	571.28
27	chr7:116628000-116678000	50000	9.78%	2.32%	12.00%	0.00%	0.00%	0.00%	1.36%	0.00%	0.00%	8.91%	65.63%	516.5
28	chr9:105522000-105551000	29000	5.83%	2.66%	5.79%	0.00%	0.61%	0.20%	0.00%	0.00%	1.19%	2.17%	81.55%	488.04
29	chr11:84973000-85028000	55000	11.27%	0.87%	1.30%	0.00%	0.87%	0.22%	0.00%	0.65%	0.00%	0.00%	84.83%	461.5
30	chr3:141591000-141601000	10000	0.00%	0.74%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.69%	98.56%	433.22
31	chr17:20742000-20810000	68000	11.59%	2.33%	3.03%	0.00%	0.47%	0.54%	0.00%	0.00%	0.00%	5.48%	76.55%	428.5
32	chr3:140968000-140994000	26000	2.81%	5.36%	1.99%	0.00%	0.00%	0.23%	0.00%	0.00%	0.00%	0.00%	89.61%	427.21
33	chr12:42905000-42939000	34000	28.58%	2.13%	8.05%	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	0.24%	60.92%	422.12
34	chr5:154408000-154412000	4000	0.00%	5.12%	14.63%	0.00%	0.00%	0.00%	0.00%	0.00%	36.59%	1.46%	42.20%	410
35	chr4:116384000-116396000	12000	4.08%	0.58%	27.70%	0.00%	0.00%	1.46%	0.00%	0.00%	0.00%	21.28%	44.90%	343
36	chr5:75292000-75303000	11000	26.28%	2.24%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	41.35%	30.13%	312
37	chr9:105449000-105469000	20000	6.83%	3.97%	9.79%	0.00%	0.00%	0.00%	0.00%	0.00%	1.90%	3.09%	74.42%	305.69
38	chr9:105120000-105160000	40000	8.56%	4.64%	3.10%	0.00%	0.00%	0.38%	0.00%	0.76%	2.20%	9.47%	70.88%	263.35
39	chr1:47839000-47873000	34000	4.80%	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%	0.01%	95.11%	0.00%	257.19
40	chr2:105696000-105728000	32000	1.45%	15.93%	4.07%	0.00%	0.81%	0.00%	1.63%	0.81%	0.00%	0.00%	75.29%	245.71
41	chr1:13931000-13959000	28000	25.64%	32.91%	6.41%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.28%	33.76%	234
42	chr5:25405000-25430000	25000	8.27%	0.43%	0.00%	7.23%	0.00%	0.00%	0.00%	1.07%	0.00%	0.86%	82.13%	232.69
43	chr17:17515000-17556000	41000	9.68%	1.51%	1.53%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.68%	61.60%	228.9
44	chr12:20765000-20783000	18000	3.81%	0.00%	46.70%	0.00%	2.38%	0.00%	0.00%	0.00%	0.00%	18.56%	28.55%	210.14
45	chr4:119575000-119601000	26000	0.48%	0.00%	0.00%	0.00%	0.00%	4.81%	0.00%	0.00%	0.00%	29.33%	65.38%	208
46	chr7:18786000-18819000	33000	0.62%	12.44%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.22%	49.25%	37.48%	205.87
47	chr4:120368000-120385000	17000	25.25%	0.00%	2.57%	0.00%	0.00%	0.51%	3.08%	0.00%	0.00%	12.59%	56.00%	194.65
48	chr1:184495000-184506000	11000	0.00%	23.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	73.33%	180
49	chrUn:42595000-42641000	46000	5.83%	3.35%	0.01%	0.00%	0.11%	0.00%	0.00%	0.00%	0.00%	77.99%	12.70%	178.84
50	chr1:241697000-241724000	27000	13.48%	2.25%	6.74%	0.00%	0.00%	0.00%	10.67%	0.00%	0.00%	6.18%	60.67%	178
51	chr1:13901000-13927000	26000	8.48%	0.57%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	42.37%	48.58%	177.02
52	chr3:140578000-140591000	13000	8.58%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.39%	91.03%	172.83
53	chr3:140041000-140077000	36000	7.71%	1.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.80%	90.48%	166.71
54	chr1:102090000-102111000	21000	3.35%	0.00%	2.64%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%	1.22%	92.37%	164.37
55	chr2:221181000-221209000	28000	0.00%	1.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.00%	88.75%	160
56	chr1:47729000-47753000	24000	5.47%	0.00%	0.00%	0.00%	0.13%	0.00%	0.00%	0.00%	0.02%	88.57%	5.81%	155.64
57	chr2:105657000-105694000	37000	1.32%	3.64%	2.65%	0.00%	0.00%	2.65%	0.00%	2.65%	0.05%	0.00%	87.04%	151.08
58	chr1:103000000-103019000	19000	17.38%	0.00%	0.91%	0.00%	0.00%	0.00%	0.00%	0.00%	0.82%	0.00%	80.89%	146.7
59	chr5:25621000-25636000	15000	16.00%	0.00%	0.00%	15.16%	0.00%	0.00%	0.00%	3.13%	0.00%	1.74%	63.98%	143.98
60	chr7:121600000-121604000	4000	0.00%	0.00%	11.19%	0.00%	0.00%	1.40%	0.00%	0.00%	0.00%	30.77%	56.64%	143
61	chr1:102827000-102845000	18000	3.15%	0.00%	0.93%	0.00%	0.00%	0.00%	0.00%	0.00%	0.90%	1.40%	93.62%	142.78
62	chr14:41998000-42011000	13000	0.00%	0.72%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.28%	138
63	chr9:92473000-92479000	6000	0.00%	0.37%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	44.61%	55.02%	134.5
64	chr9:7094000-7096000	2000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.78%	74.22%	128
65	chr15:84157000-84160000	3000	0.00%	0.00%	1.63%	0.00%	0.00%	8.16%	0.00%	0.00%	0.00%	77.14%	13.06%	122.5
66	chr4:165701000-165712000	11000	0.00%	6.77%	18.88%	0.00%	5.13%	0.00%	0.00%	0.00%	0.27%	36.94%	32.01%	121.83
67	chr11:82952000-82970000	18000	76.13%	14.30%	0.83%	0.00%	0.00%	0.00%	0.00%	1.66%	0.42%	2.50%	4.16%	120.22
68	chr13:37654000-37658000	4000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.86%	0.00%	0.00%	13.79%	85.34%	116
69	chr17:18976000-19001000	25000	1.82%	10.91%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	60.91%	26.36%	110
70	chr13:73630000-73667000	37000	3.57%	2.02%	11.11%	0.00%	0.00%	0.50%	3.03%	0.00%	78.76%	0.00%	1.01%	99.04
71	chrX:124261000-124262000	1000	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	97.46
72	chr3:140378000-140397000	19000	1.57%	0.00%	1.57%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	96.86%	95.5
73	chr2:105598000-105623000	25000	0.00%	0.54%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.46%	93
74	chr9:105273000-105298000	25000	12.11%	6.93%	14.10%	0.00%	2.21%	0.00%	0.00%	0.00%	0.00%	4.74%	59.91%	90.42
75	chrUn:15394000-15406000	12000	13.48%	5.62%	0.00%	0.00%	0.00%	1.12%	3.37%	0.00%	0.00%	0.00%	76.40%	89
76	chr5:143156000-143164000	8000	0.58%	3.49%	2.91%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	93.02%	86
77	chr12:27425000-27443000	18000	0.00%	0.00%	9.88%	0.00%	1.16%	0.58%	0.00%	0.00%	0.00%	19.77%	68.60%	86
78	chr7:94716000-94745000	29000	29.49%	17.95%	7.69%	0.00%	0.00%	0.00%	1.28%	0.00%	0.00%	2.56%	41.02%	78
79	chr9:90835000-90839000	4000	22.37%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	76.32%	1.32%	76
80	chr5:79201000-79204000	3000	0.00%	0.00%	1.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	98.50%	66.5
81	chr10:36520000-36523000	3000	0.00%	0.00%	16.22%	0.00%	10.81%	0.00%	0.00%	0.00%	0.00%	0.90%	72.07%	55.5
82	chr1:92677000-92707000	30000	3.74%	3.74%	9.35%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	83.18%	53.5
83	chr9:90967000-90984000	17000	0.00%	0.00%	0.00%	0.00%	4.08%	4.08%	0.00%	0.00%	0.00%	79.59%	12.24%	49
84	chr6:93598000-93616000	18000	6.84%	4.56%	0.00%	0.00%	0.00%	0.00%	0.00%	6.84%	0.00%	1.90%	79.85%	43.83
85	chr17:20188000-20198000	10000	16.05%	4.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.47%	77.38%	40.49
86	chr1:51025000-51039000	14000	7.67%	0.12%	0.00%	0.00%	0.00%	0.00%	0.00%	0.15%	0.08%	91.98%	0.00%	40.41
87	chr7:31878000-31887000	9000	29.49%	0.00%	0.00%	0.00%	0.00%	0.00%	3.85%	0.00%	0.00%	35.90%	30.77%	39
88	chr8:47138000-47143000	5000	5.26%	0.00%	5.26%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	57.89%	31.58%	38
89	chr20:22761000-22765000	4000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	37
90	chr4:165034000-165054000	20000	8.72%	0.73%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.45%	89.10%	34.42
91	chrX:123931000-123946000	15000	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	33.75
92	chrX:152895000-152918000	23000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	33
93	chr10:90649000-90661000	12000	0.00%	0.00%	6.25%	0.00%	0.00%	0.00%	0.00%	0.00%	93.75%	0.00%	0.00%	32
94	chr2:44320000-44347000	27000	3.50%	0.00%	19.97%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	76.54%	30.05
95	chr2:121139000-121146000	7000	0.00%	0.00%	3.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	17.24%	79.31%	29
96	chr15:187000-217000	30000	10.34%	6.90%	6.90%	0.00%	0.00%	0.00%	3.45%	0.00%	0.00%	65.52%	6.90%	29
97	chr12:11152000-11160000	8000	3.56%	0.00%	46.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	46.28%	3.56%	28.09
98	chr5:63093000-63106000	13000	0.00%	11.54%	0.00%	0.00%	0.00%	3.85%	0.00%	0.00%	0.00%	65.38%	19.23%	26
99	chr17:20354000-20367000	13000	4.55%	9.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.55%	81.82%	22
100	chr10:84482000-84488000	6000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	20
All		2733000	7.16%	4.40%	6.03%	0.07%	0.18%	0.78%	0.90%	0.06%	0.68%	9.39%	70.36%	56737.94
genomic proportion			7.92%	20.82%	6.42%	0.21%	0.60%	2.16%	0.74%	0.30%	13.46%	9.74%	37.44%	99.82%

Supplemental Table S6. Annotation of the ninety-four piRNA clusters from the mouse genome

Rank	Genomic coordinate	Length	LTR	LINE	SINE	Satellite	Low complexity	Simple repeat	DNA	RNA	Unknown	Other	mRNA	EST	No annotation	All
1	chr17:25458000-25553000	95000	5.68%	1.15%	2.45%	0.00%	0.00%	0.71%	0.02%	0.00%	0.00%	0.00%	4.80%	21.17%	64.02%	5091.86
2	chr9:67715000-67803000	88000	6.57%	14.84%	3.35%	0.00%	0.16%	0.53%	1.91%	0.00%	0.00%	0.02%	0.57%	1.26%	70.79%	4329.55
3	chr18:67419000-67485000	66000	3.24%	7.97%	2.70%	0.00%	0.03%	0.76%	0.06%	0.00%	0.00%	0.00%	0.32%	0.61%	84.31%	3430.2
4	chr2:92374000-92462000	88000	3.44%	3.38%	7.58%	0.00%	0.00%	0.53%	0.00%	0.00%	0.00%	0.00%	1.89%	1.17%	82.01%	3237.57
5	chr6:127799000-127879000	80000	15.73%	1.65%	7.70%	0.00%	0.03%	0.63%	0.43%	0.00%	0.00%	0.00%	5.07%	68.75%	0.00%	3153.19
6	chr7:69622000-69701000	79000	5.26%	3.03%	0.67%	0.00%	0.10%	0.08%	1.05%	0.00%	0.00%	0.05%	63.91%	0.95%	24.92%	3150.45
7	chr5:112507000-112564000	57000	33.41%	4.09%	3.33%	0.00%	0.50%	0.20%	0.80%	0.00%	0.00%	0.00%	5.68%	5.55%	46.44%	2810.02
8	chr12:95876000-95950000	74000	9.61%	2.88%	0.99%	0.00%	0.26%	0.39%	2.13%	0.00%	0.00%	0.00%	11.76%	71.98%	0.00%	2674.54
9	chr15:59313000-59370000	57000	5.88%	8.28%	10.54%	0.00%	0.12%	0.44%	3.67%	0.00%	0.00%	0.00%	20.30%	1.01%	49.76%	2586.18
10	chr5:148721000-148766000	45000	5.18%	1.64%	8.42%	0.00%	0.10%	0.77%	0.00%	0.00%	0.00%	0.00%	21.77%	0.75%	61.38%	2012.21
11	chr7:65734000-65810000	76000	2.84%	2.91%	0.66%	0.00%	0.05%	0.25%	0.05%	0.00%	0.00%	0.00%	0.08%	7.26%	85.91%	1982.41
12	chr14:21743000-21796000	53000	2.94%	8.66%	2.42%	0.00%	3.04%	0.21%	0.00%	0.00%	0.00%	0.00%	0.00%	17.07%	65.65%	1939.11
13	chr9:54236000-54298000	62000	7.49%	4.29%	12.71%	0.00%	0.00%	1.72%	0.00%	0.00%	0.00%	0.00%	0.05%	0.00%	73.74%	1921.75
14	chr7:68946000-69002000	56000	5.51%	6.03%	2.91%	0.00%	1.10%	0.14%	5.31%	0.00%	0.00%	0.00%	2.77%	11.19%	65.04%	1733.5
15	chr15:74701000-74752000	51000	3.53%	0.97%	0.44%	0.00%	0.00%	0.09%	0.00%	0.00%	0.00%	0.00%	0.91%	6.58%	87.46%	1641.8
16	chr17:64393000-64458000	65000	5.74%	1.12%	4.68%	0.00%	0.00%	0.63%	0.16%	0.00%	0.00%	0.00%	87.66%	0.00%	0.00%	1261.86
17	chr4:93669000-93725000	56000	6.82%	9.88%	2.81%	0.00%	0.08%	0.80%	0.00%	0.00%	0.00%	0.00%	79.60%	0.00%	0.00%	1244.91
18	chr5:114033000-114076000	43000	3.73%	4.77%	0.26%	0.00%	0.10%	0.05%	0.05%	0.00%	0.00%	0.00%	2.35%	18.48%	70.21%	1006.28
19	chr10:75491000-75537000	46000	3.24%	0.50%	6.94%	0.00%	0.00%	2.08%	0.00%	0.00%	0.00%	0.00%	9.00%	7.96%	70.28%	866.49
20	chr11:103533000-103559000	26000	16.55%	0.41%	3.10%	0.00%	0.12%	0.12%	0.00%	0.00%	0.00%	0.00%	3.34%	17.45%	58.91%	847.94
21	chr4:61654000-61673000	19000	11.55%	0.00%	2.15%	0.00%	1.15%	0.22%	0.00%	0.00%	0.00%	0.00%	0.00%	84.94%	0.00%	697
22	chr15:78725000-78767000	42000	16.07%	7.07%	7.14%	0.00%	0.00%	0.83%	0.00%	0.00%	0.00%	0.00%	1.80%	2.56%	64.53%	664.83
23	chr10:62171000-62224000	53000	3.73%	2.21%	14.49%	0.00%	2.29%	1.27%	3.65%	0.00%	0.00%	0.00%	72.36%	0.00%	0.00%	589.4
24	chr7:121421000-121461000	40000	2.15%	3.22%	0.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	93.92%	0.00%	0.54%	559.01
25	chr8:90782000-90795000	13000	5.04%	3.17%	1.87%	0.00%	0.00%	0.37%	0.00%	0.00%	0.00%	0.00%	0.00%	5.22%	84.33%	536
26	chr14:40245000-40263000	18000	3.12%	0.41%	0.00%	0.00%	0.19%	0.00%	0.00%	0.00%	0.00%	0.00%	96.28%	0.00%	0.00%	528.17
27	chr6:81912000-81933000	21000	0.57%	2.09%	12.18%	0.00%	0.00%	0.00%	8.19%	0.00%	0.00%	0.00%	7.42%	0.25%	69.29%	525.35
28	chr10:66270000-66293000	23000	6.52%	0.40%	0.80%	0.00%	2.41%	1.00%	0.00%	0.00%	0.00%	0.00%	0.90%	87.96%	0.00%	497.96
29	chr2:150989000-151008000	19000	3.58%	5.30%	0.19%	0.00%	0.07%	0.00%	2.16%	0.00%	0.00%	0.00%	88.70%	0.00%	0.00%	447.16
30	chr2:151055000-151073000	18000	3.27%	5.47%	0.19%	0.00%	0.08%	0.00%	2.23%	0.00%	0.00%	0.00%	88.76%	0.00%	0.00%	433.16
31	chr2:151143000-151162000	19000	3.10%	5.35%	0.19%	0.00%	0.08%	0.00%	2.25%	0.00%	0.00%	0.00%	89.03%	0.00%	0.00%	430
32	chr4:57398000-57406000	8000	11.94%	2.24%	0.00%	0.00%	0.00%	0.25%	0.00%	0.00%	0.00%	0.00%	70.15%	0.00%	15.42%	402
33	chr1:92900000-92905000	5000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.30%	0.00%	0.00%	0.00%	0.00%	10.13%	88.57%	385
34	chr6:128190000-128224000	34000	0.00%	1.19%	3.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	66.45%	28.67%	0.00%	378.5
35	chr10:85297000-85334000	37000	5.08%	1.25%	0.73%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.59%	1.50%	87.86%	367.68
36	chr2:150882000-150899000	17000	2.32%	6.06%	0.14%	0.00%	0.27%	0.00%	2.45%	0.00%	0.00%	0.00%	88.75%	0.00%	0.00%	366.63
37	chr8:108718000-108731000	13000	0.00%	7.54%	19.47%	0.00%	0.00%	4.11%	17.49%	0.00%	0.00%	0.00%	0.00%	0.00%	51.40%	328.78
38	chr15:83436000-83450000	14000	0.00%	0.00%	18.91%	0.00%	0.00%	0.31%	0.00%	0.00%	0.00%	0.00%	7.60%	30.70%	42.48%	322.5
39	chr14:17284000-17311000	27000	32.38%	2.79%	0.00%	0.00%	0.00%	0.93%	3.41%	0.00%	0.00%	0.00%	60.18%	0.00%	0.31%	322.48
40	chr6:88042000-88069000	27000	1.19%	44.70%	3.87%	0.00%	0.00%	0.34%	1.18%	0.00%	0.00%	0.00%	48.73%	0.00%	0.00%	297.55
41	chr6:85140000-85156000	16000	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.99%	0.00%	0.00%	295.03
42	chr10:86113000-86163000	50000	1.25%	0.97%	1.28%	0.00%	0.00%	0.35%	0.00%	0.00%	0.00%	0.00%	4.30%	1.28%	90.58%	287.06
43	chr4:134996000-135009000	13000	0.00%	12.59%	12.94%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	74.48%	0.00%	0.00%	286
44	chr10:18519000-18554000	35000	4.45%	2.59%	2.22%	0.00%	5.56%	0.37%	0.37%	0.00%	0.00%	0.00%	84.44%	0.00%	0.00%	270.01
45	chr7:49122000-49150000	28000	0.00%	27.08%	0.00%	0.00%	0.00%	0.78%	0.00%	0.00%	0.00%	0.00%	0.00%	50.03%	22.11%	257.83
46	chr2:127432000-127445000	13000	0.00%	1.20%	7.60%	0.00%	0.00%	0.40%	0.80%	0.00%	0.00%	0.00%	87.60%	0.00%	2.40%	250
47	chr3:34676000-34687000	11000	0.00%	14.29%	0.42%	0.00%	0.00%	1.26%	0.00%	0.00%	0.00%	0.00%	0.00%	7.56%	76.47%	238
48	chr5:142790000-142801000	11000	5.92%	0.00%	29.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	64.90%	0.00%	0.00%	236.51
49	chr9:44078000-44086000	8000	2.22%	0.00%	10.22%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	87.56%	0.00%	0.00%	225
50	chr8:91999000-92013000	14000	13.24%	1.37%	0.46%	0.00%	0.46%	3.65%	0.00%	0.00%	0.00%	0.00%	80.82%	0.00%	0.00%	219
51	chr14:22009000-22038000	29000	1.94%	23.40%	0.97%	0.00%	0.00%	0.48%	0.00%	0.00%	0.00%	0.00%	5.56%	67.65%	0.00%	206.94
52	chr10:127189000-127202000	13000	0.00%	0.00%	1.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.00%	0.00%	0.00%	200
53	chr7:27076000-27084000	8000	14.65%	1.09%	26.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	58.08%	0.00%	0.00%	183.35
54	chr10:86171000-86183000	12000	0.33%	1.31%	2.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	5.08%	0.00%	91.22%	177.79
55	chr6:128462000-128474000	12000	4.62%	1.16%	3.47%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	90.75%	0.00%	0.00%	173
56	chr11:108102000-108107000	5000	4.91%	0.00%	24.07%	0.00%	0.00%	0.60%	0.00%	0.00%	0.00%	0.00%	70.41%	0.00%	0.00%	166.17
57	chr5:134131000-134139000	8000	0.00%	10.19%	12.10%	1.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	76.43%	0.00%	0.00%	157
58	chr3:124237000-124263000	26000	1.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	98.72%	0.00%	0.00%	156
59	chr19:36956000-36962000	6000	2.67%	0.00%	1.33%	0.00%	0.00%	0.00%	24.67%	0.00%	0.00%	0.00%	54.00%	0.00%	17.33%	150
60	chr13:52433000-52440000	7000	0.00%	3.36%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	96.64%	0.00%	0.00%	149
61	chr6:86062000-86079000	17000	5.41%	0.00%	4.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	39.57%	48.25%	2.71%	147.83
62	chr17:47253000-47256000	3000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	135
63	chr14:22077000-22113000	36000	6.93%	3.01%	0.75%	0.00%	0.00%	0.00%	3.01%	0.00%	0.00%	0.00%	75.73%	2.26%	8.29%	132.7
64	chr4:123278000-123287000	9000	0.00%	0.00%	2.65%	0.00%	0.00%	0.88%	0.00%	0.00%	0.00%	0.00%	1.33%	95.13%	0.00%	113
65	chr13:49938000-49944000	6000	3.55%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	96.45%	0.00%	112.83
66	chr1:91310000-91314000	4000	17.92%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	26.42%	5.66%	50.00%	106
67	chr13:51154000-51161000	7000	1.89%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	98.11%	0.00%	0.00%	106
68	chr13:49820000-49833000	13000	12.55%	0.38%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	87.07%	0.00%	87.67
69	chr9:51827000-51831000	4000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	83
70	chr8:26050000-26065000	15000	2.55%	0.00%	1.87%	0.00%	0.00%	0.62%	0.00%	0.00%	0.00%	0.00%	0.25%	51.09%	43.61%	80.25
71	chr1:91445000-91467000	22000	0.00%	0.00%	0.00%	0.00%	0.00%	4.00%	0.00%	0.00%	0.00%	0.00%	94.66%	0.00%	1.33%	75
72	chr1:125054000-125057000	3000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.85%	0.00%	0.00%	0.00%	0.00%	0.00%	93.15%	73
73	chr15:73176000-73188000	12000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	70.5
74	chr5:101021000-101040000	19000	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.98%	0.00%	0.00%	64.01
75	chr5:135888000-135908000	20000	0.06%	0.00%	56.87%	1.72%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	37.91%	1.72%	1.72%	58.03
76	chr15:80000000-80039000	39000	8.77%	0.00%	15.81%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	59.64%	15.79%	0.00%	57.01
77	chr10:94234000-94238000	4000	12.96%	0.00%	1.85%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	85.19%	0.00%	0.00%	54
78	chr8:109603000-109621000	18000	0.93%	0.00%	17.76%	0.00%	0.00%	4.67%	0.00%	0.00%	0.00%	0.00%	76.63%	0.00%	0.00%	53.5
79	chrX:62174000-62206000	32000	0.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.94%	52.03
80	chr5:129378000-129393000	15000	0.00%	0.00%	8.33%	0.00%	4.17%	0.00%	0.00%	0.00%	0.00%	0.00%	85.42%	0.00%	2.08%	48
81	chr14:40079000-40097000	18000	99.99%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	46.59
82	chr1:57602000-57611000	9000	13.04%	0.00%	2.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.87%	73.91%	46
83	chr10:60946000-60952000	6000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	46
84	chr8:35860000-35873000	13000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	40.24%	59.76%	0.00%	41.83
85	chr13:24271000-24273000	2000	13.16%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	86.84%	0.00%	0.00%	38
86	chr17:20779000-20785000	6000	13.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	86.82%	0.00%	36.67
87	chr8:115820000-115829000	9000	0.00%	0.00%	2.78%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	8.33%	69.44%	19.44%	36
88	chr6:92248000-92270000	22000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	35.5
89	chr13:49373000-49377000	4000	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	29.5
90	chr13:20869000-20871000	2000	0.00%	0.00%	10.71%	19.64%	0.00%	0.00%	0.00%	0.00%	8.93%	0.00%	17.86%	0.00%	42.86%	28
91	chr3:20186000-20211000	25000	0.00%	8.51%	0.00%	0.00%	4.25%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	87.22%	23.5
92	chr4:123336000-123342000	6000	4.55%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.64%	63.63%	18.18%	22
93	chr10:84029000-84037000	8000	47.62%	0.00%	9.52%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.76%	38.10%	0.00%	0.00%	21
94	chr11:95874000-95880000	6000	0.00%	0.00%	9.52%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	90.48%	0.00%	0.00%	21
All		2489000	7.16%	4.67%	4.36%	0.01%	0.28%	0.51%	1.01%	0.00%	0.00%	0.01%	22.13%	14.19%	45.68%	62465.62
genomic proportion			9.60%	19.26%	7.07%	0.04%	0.75%	2.33%	0.83%	0.00%	0.03%	0.27%	31.19%	5.05%	23.52%	99.96%