JOURNAL OF VIROLOGY, Dec. 2009, p. 12751–12758 Vol. 83, No. 24
0022-538X/09/$12.00 doi:10.1128/JVI.01325-09

Characterization of Viral and Human RNAs Smaller than Canonical MicroRNAs▽§

Zhihua Li,[1,2]† Sang Woo Kim,[1]† Yuefeng Lin,[1]† Patrick S. Moore,[2]† Yuan Chang,[2]† and Bino John[1,2]†*

Department of Computational Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260,[1] and Molecular Virology Program, Hillman Cancer Center, 5117 Centre Avenue, Pittsburgh, Pennsylvania 15213[2]

Received 27 June 2009/Accepted 22 September 2009

Recently identified small (20 to 40 bases) RNAs, such as microRNAs (miRNAs) and endogenous small interfering RNAs (siRNAs) participate in important cellular pathways. In this report, we systematically characterized several novel features of human and viral RNA products smaller than miRNAs. We found that Kaposi sarcoma-associated herpesvirus K12-1 miRNA (23 bases) associates with a distinct, unusually small (17-base) RNA (usRNA) that can effectively downregulate a K12-1 miRNA target, human *RAD21*, suggesting that stable degradation-like products may also contribute to gene regulation. High-throughput sequencing reveals a diverse set of human miRNA-derived usRNAs and other non-miRNA-derived usRNAs. Human miRNA-derived usRNAs preferentially match to 5′ ends of miRNAs and are also more likely to associate with the siRNA effector protein Ago2 than with Ago1. Many non-miRNA-derived usRNAs associate with Ago proteins and also frequently contain C-rich 3′-specific motifs that are overrepresented in comparison to Piwi-interacting RNAs and transcription start site-associated RNAs. We postulate that approximately 30% of usRNAs could have evolved to participate in biological processes, including gene silencing.

Recent studies have identified several classes of small RNAs, such as microRNAs (miRNAs), across a wide range of organisms (9). miRNAs can function as repressors (24) or activators (47) of gene expression. More than 500 human miRNAs are known, and many more are thought to exist (3). Advances in high-throughput sequencing have enabled the discovery of additional animal miRNAs and classes of small RNAs on an unprecedented scale across distantly related species, including animals, fungi, unicellular algae, plants, and viruses (9). In addition to miRNAs, recently identified metazoan small RNAs include endogenous small interfering RNAs (siRNAs) that mediate mRNA degradation (6, 13, 33, 45), hairpin RNAs that are thought to repress target transcripts in flies (32), Piwi-interacting RNAs (piRNAs) that modulate spermatogenesis in mammals (14, 22), piRNA-like repeat-associated siRNAs that regulate chromatin structure of retrotransposons in insect germ lines (21), transcription start site-associated RNAs (TSSa-RNAs), and the functionally uncharacterized piRNA-like 21U-RNAs harboring 5′ uridines in worms (36).

The various small-RNA classes frequently share sequence characteristics and recruit similar protein partners. For example, the 5′ ends of miRNAs (41, 49) and piRNAs (14) prefer uridines that likely bind to argonaute/Piwi proteins (44, 46). The argonaute (Ago) proteins bind miRNAs and siRNAs, while the Piwi proteins, a subgroup of the Ago family, bind piRNAs (9). Ago/Piwi-bound small RNAs are part of a larger ribonucleoprotein complex (RNP). The small RNAs in RNPs act as adaptors that recognize complementary nucleic acid targets and thus guide the RNPs to their target molecules (e.g., mRNAs) to direct processes, such as transcript cleavage (miRNAs, siRNAs, and piRNAs) and translational repression (miRNAs). Transcript cleavage by siRNAs and piRNAs also generates numerous small secondary RNAs that are believed to also regulate cellular processes (4, 33).

Small RNAs are generally distinguished based on their protein partners, mode of biogenesis, and mechanisms of maturation. For instance, while piRNAs exclusively associate with the Piwi proteins and are thought to originate from single-stranded RNAs, metazoan miRNAs interact with the Ago proteins and are processed by Dicer from imperfect hairpin structures formed by their RNA precursors. As a consequence of distinct mechanisms that mediate small-RNA biogenesis, the sequence lengths are simple, yet powerful distinguishing features of various metazoan small-RNA classes, including 21U-RNAs (21 nucleotides [nt]), endogenous siRNAs (~22 nt), miRNAs (~22 nt), repeat-associated siRNAs (~25 nt), and piRNAs (~28 nts).

RNA products shorter than 20 nucleotides are currently not studied in detail due to the hypothesis that they are transient degradation products and because of technical challenges in their detection and analysis. Surprisingly, even degradation-like RNA products can evolve biological function, as exemplified by the weakly expressed small RNA (~25 nt) that originates from a small nucleolar RNA (snoRNA) (ACA45), which functions like a miRNA (8) and is Dicer dependent. In this

* Corresponding author. Mailing address: 3080 BST3, 3501 Fifth Avenue, Pittsburgh, PA 15260. Phone: (412) 648-8607. Fax: (412) 648-3163. E-mail: bino@pitt.edu.

§ Supplemental material for this article may be found at http://jvi.asm.org/.

† Author contributions: Z.L. carried out the majority of the wet-lab experiments, conducted the majority of computational analyses, and wrote the manuscript. S.W.K. improved the protocols used to detect small RNAs and generated some sequencing data. Y.L. tested multiple methods of RNA classification based on probabilistic and support-vector machine approaches. P.S.M., Y.C., and B.J. conceived the project, supervised the project, and contributed to the writing of the manuscript.

▽ Published ahead of print on 7 October 2009.

study, we examine the abundance, expression patterns, and sequence characteristics of unusually small (~16-base) RNAs (usRNAs) that are below the canonical size of miRNAs, which generally would otherwise be ignored as insignificant transient degradation products.

MATERIALS AND METHODS

Northern blot analysis. Total RNA (50 μg) was separated on 18% denaturing polyacrylamide gels and electro-transferred to Brightstar positively charged nylon membranes (Ambion, Austin, TX). LNA-DNA mixed oligonucleotide probes (Sigma-Aldrich, St. Louis, MO) were labeled with the nonradioactive digoxigenin (DIG), using an end tailing kit (Roche Applied Science, Indianapolis, IN). Hybridization was carried out overnight in Ultrahyb buffer (Ambion) at 42°C, and the blots were washed for two rounds in 2× SSC/0.5% SDS (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and one round in 5× SSC/0.5% SDS at 68°C. DIG signals were detected with alkaline phosphatase-conjugated anti-DIG antibody (catalog no. 11093274910; Roche). The probe sequence 5′-GCTTAC ACCCAGTTTCCTGTAAT-3′ (LNAs are underlined) was used to detect K12-1.

Preparation of small RNA libraries. BCP-1 cells (12) were maintained in RPMI 1640 medium (Mediatech Inc, Herndon, VA) supplemented with 20% fetal bovine serum (Sigma-Aldrich). Total RNA was isolated from BCP-1 cells using Trizol (Invitrogen) and size fractionated on a 20% polyacrylamide-urea gel (National Diagnostics, Atlanta, GA) along with synthesized RNA size markers (Sigma-Aldrich). The gel was stained with 5 μg/ml ethidium bromide (Bio-Rad, Hercules, CA), and the RNA fractions ranging approximately from 8 to 30 nucleotides were excised. The gel slices were crushed and soaked in 0.4 M NaCl, frozen, thawed, and passed through Nanosep 100K filters (Pall Corp., Ann Arbor, MI). The RNA was precipitated with 20 μg glycogen (Ambion), dissolved in water, and ligated to the 3′ adaptor (5′-AMP-CTGTAGGCACCATCAAT-d dC-3′; IDT Corp., Coralville, IA) using T4 RNA ligase (Ambion). The ligated RNA was size fractionated, excised on 18% polyacrylamide-urea gel, eluted from the gel slices, and ligated to the 5′ adaptor (ATCGTrArGrGrCrArCrCrUrGrA rArA, where r indicates ribonucleotide; IDT Corp.). The reaction mixture was phenol extracted and precipitated. The resulting ligated RNAs were reverse transcribed based on a primer (ATTGATGGTGCCTACAG) complementary to the 3′ adaptor using SuperScript II (Invitrogen, Carlsbad, CA). The reverse-transcribed cDNA was amplified by PCR based on 5′ (ATCGTAGGCACCTG AAA) and 3′ (ATTGATGGTGCCTACAG) primers complementary to the adaptors using Platinum *Taq* high-fidelity polymerase (Invitrogen, Carlsbad, CA). PCR products were separated on a 4% MetaPhor agarose gel (Lonza, Rockland, ME) and purified using the Qiaex II gel extraction kit (Qiagen, Valencia, CA). A second round of PCR was performed to add 454 sequencing primers to the PCR products, and the resulting products were sequenced by 454 Life Sciences (Branford, CT).

Western blots. Double-stranded RNAs (dsRNA) were synthesized commercially by Proligo (Paris, France). The sequences for miR-K12-1 dsRNA were 5′-AUUACAGGAAACUGGGUGUAAGC-3′ (sense) and 5′-UUACACCUG UUUCCUGCAACCC-3′ (antisense). Sequences for ds-us-K12-1 were 5′-AUU ACAGGAAACUGGGU-3′ (sense) and 5′-CCAGUUUCCUGUAACCC-3′ (antisense). The Block-iT fluorescent oligo (Invitrogen) was used as the irrelevant control sequence. Block-iT oligo is not known to regulate genes via the RNA interference pathway since its sequence is designed to have negligible similarity to any known gene. For transfection of oligos into HEK293 cells, 6×10^5 cells were plated in six-well plates and transfected the following day with 100 nM dsRNA duplexes using Lipofectamin2000 (Invitrogen). Cells were harvested after posttransfection (48 h), and Western blot analyses were performed using standard procedures. Rabbit anti-RAD21 antibody was obtained from Abcam (Cambridge, MA); mouse anti-tubulin antibody was purchased from Sigma-Aldrich (St. Louis, MO); IRDye 800CW goat anti-mouse immunoglobulin G (IgG) was obtained from Rockland Immunochemicals (Gilbertsville, PA); Alexa Fluor 680 goat anti-rabbit IgG was obtained from Invitrogen. Membranes were scanned, and images were analyzed using the Odyssey infrared imaging system (Li-CoR Biosciences, Lincoln, NE). The ratio of expression levels of RAD21 to tubulin was normalized to that of the RNA-transfected control sample.

Real-time RT-PCR. After transfection of dsRNAs as noted above, total RNA was purified using Trizol reagent (Invitrogen) and then treated with RNase-free DNase I (Ambion). Reverse transcription (RT) was performed using a Superscript first-strand synthesis kit (Invitrogen) and random primers. Real-time PCR was carried out using SYBR greenER qPCR SuperMix Universal (Invitrogen) on a SmartCycler system (Cepheid, Sunnyvale, CA). The primers used for amplification of *RAD21* were 5′-GCACACTCCTGGTTTGGAAC-3′ (sense) and 5′-A ACAGTCACATGATTTCTGATGC-3′ (antisense). The house-keeping GAPDH gene was used as a control. The primers for *GAPDH* were 5′-CGGAGTCAAC GGATTTGGTCGTAT-3′ (sense) and 5′-AGCCTTCTCCATGGTGGTGAAG AC-3′ (antisense). BCBL1 cDNA was used as a reference sample to generate the standard curves for both *RAD21* and *GAPDH*, from which the expression levels in the experimental samples were extrapolated. The ratio of expression levels of *RAD21* to *GAPDH* was normalized to that of the RNA-transfected control sample.

Immunoprecipitation and Northern blot analysis of Ago-RNA complexes. HEK293 cells were transfected with ds-K12-1 and ds-us-K12-1 as described above. Cells were lysed after 48 h, using buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.25% NP-40, and 1.5 mM $MgCl_2$. The antibodies used were anti-Ago1-4B8 (8), anti-Ago2-11A9 (8), anti-Ago3-4B1-F6 (51), anti-Ago4-1B7-G11 (51), and irrelevant antibody rat IgG 2b (BD Biosciences Pharmingen, San Diego, CA; 1 μg/ml). After overnight incubation at 4°C, protein A/G Plus agarose beads (Santa Cruz Biotechnology, Inc., Santa Cruz, CA) were added to the lysate at a concentration of 30 μl/ml. After another 6 h of incubation at 4°C, beads were pelleted, washed four times with washing buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5% NP-40, and 1.5 mM $MgCl_2$, and then washed once with phosphate-buffered saline. RNA was extracted from beads by Trizol (Invitrogen) and then detected by Northern blotting. The probe used against us-K12-1 was an LNA-DNA mixed oligo (5′-ACCCAGTTTCCTGTAA T-3′; LNAs are underlined). Northern blot bands were quantified by ImageJ (http://rsbweb.nih.gov/ij/). The ratio between the quantified intensities of us-K12-1 and K12-1 for each Ago was used as an estimate for the relative affinity of each Ago protein. The ratios were further normalized using the relative affinity of Ago1 to enable easy data interpretation, resulting in a relative affinity value of 1.0 for Ago1.

Luciferase assays. HEK293 cells (ATCC) were maintained in Dulbecco's minimal essential medium (Mediatech Inc, Herndon, VA) supplemented with 10% fetal bovine serum (Sigma-Aldrich, St. Louis, MO) and were plated at a density of 4×10^5/ml in 24-well plates. Cells were transfected the following day using Lipofectamin2000 with firefly luciferase reporter plasmids (200 ng) and *Renilla* luciferase control plasmids (100 ng), along with either 10 nM ds-K12-1 or ds-us-K12-1 and irrelevant control oligos. Luciferase activity was measured 48 h after transfection using the dual luciferase assay system (Promega, Madison, WI). The ratio of the firefly and *Renilla* activities was normalized by the average activity (firefly/*Renilla*) of the control RNA. For dose-response analysis, the amount of plasmids was kept constant, and desired amounts of oligos (ds-K12-1 or ds-us-K12-1) were mixed with control oligos to maintain the concentration of total oligos at 10 nM. The strategy was used to help ensure that equal amounts of plasmid-RNA-lipid complexes were formed in different wells, and thus, transfection efficiency was stable across samples.

Ago-association data and motif detection. The Ago immunoprecipitation data (8) contained only sequences with a minimum length of 17 bases. In order to compare Ago1 and Ago2 data, Ago1 reads were multiplied by a normalization factor (2.08) to accommodate the difference in total number of reads between the samples. Motifs were detected using MEME and MAST (http://meme.sdsc.edu). Two-tailed Fisher's exact test was used to estimate statistical significance of the observed preferences of motifs for usRNAs versus TSSa-RNAs and piRNAs.

To search for additional position-specific motifs and since readily usable software was not available for such a search, we implemented concepts used for entropy-based motif representation (5, 39). Specifically, with respect to the 5′ and 3′ ends of a given sequence, we evaluated the positional preference of every possible sequence substring (*k*-mers), where *k* is the sequence length. The information content of every position for a set (S_k) of all RNA sequences containing a given *k*-mer was determined. The *i*nformation *c*ontent (IC) at a position (*j*) is calculated using the base (*b*) frequency (*F*) at position *j*, based on Shannon entropy. Specifically, IC is defined (39) as $\mathrm{IC}j(S_k) = \log_2 4 - [\Sigma_{b\in\{A,U,G,C\}} - F_{bj}(S_k) \times \log_2 F_{bj}(S_k)]$.

The positional preference of a given base is evaluated as the *r*elative *i*nformation *c*ontent of the base, $\mathrm{RIC}_{bj}(S_k)$, the product of $\mathrm{IC}_j(S_k)$ and $F_{bj}(S_k)$ (5, 39). To avoid artifacts due to small sample size and to limit our analysis to a set of prevalent motifs, sequence sets (S_k) that contain less than a predefined number of sequences, N^0, were not considered for the analysis. We further required the bases corresponding to a specific *k*-mer at a given position to have the highest RIC scores and the corresponding RIC scores to be above a predefined threshold, RIC^0 ($\mathrm{RIC}^0 = 0.2$). The threshold N^0 was set to 1% of the number of sequences in S_k. If the resulting threshold was less than 100, N^0 was set to 100. To avoid spurious motifs due to close sequence family members, we used BLAST to curate representative RNAs that shared less than 80% sequence identity. The

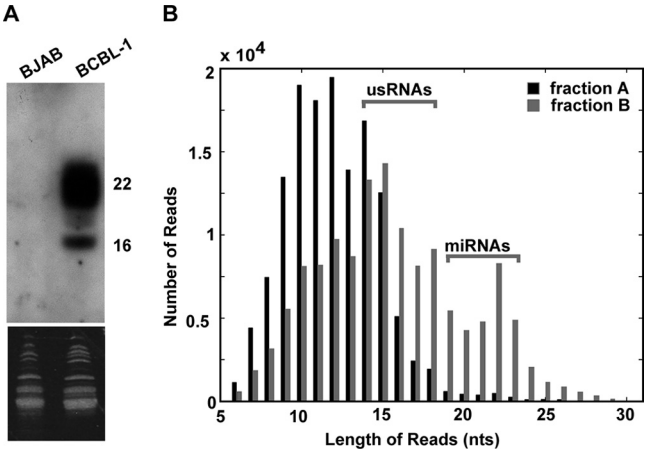

FIG. 1. Detection of us-k12-1 and other putative small RNAs. (A) Northern blotting detected K12-1 (~22 nt) and us-K12-1 (~16 nt) in KSHV-positive cells (BCBL-1) but not in KSHV-negative cells (BJAB). Ethidium bromide staining (bottom) confirmed equal loading. (B) The number of sequence reads for various sizes of RNAs. Numerous usRNAs were also found in fraction B, likely due to RNA cross-hybridization and nonuniform RNA migration.

representative set of sequences of 8,374 usRNAs, 23,438 human piRNAs (38), and 11,549 TSSa-RNAs (k = 1 to 7) were used to search for motifs.

RESULTS

Human cells contain numerous stable usRNAs smaller than miRNAs. While studying miRNAs of Kaposi sarcoma-associated herpesvirus (KSHV), we observed that the K12-1 KSHV miRNA (~22 nt) was consistently associated with a usRNA (us-K12-1) having a unique and distinct signal (Fig. 1A). To identify the sequence of us-K12-1 and its functional potency and to investigate whether similarly sized usRNAs were present, we used high-throughput sequencing to determine the composition of two small RNA fractions (Materials and Methods), comprised of 8 to 18 bases (fraction A) and 19 to 30 bases (fraction B). High-throughput sequencing yielded 533,491 (293,762 in A and 239,729 in B) reads, corresponding to 69,094 distinct RNA sequences longer than five nucleotides. Visual inspection (Fig. 1B) and deconvolution analysis (see the supplemental material) of resulting sequence-length distributions revealed a peak for 15- to 18-base-long RNA products conforming to a narrow distribution similar to the distribution of 19- to 23-base-long RNAs containing siRNAs and miRNAs (see Fig. S1 in the supplemental material). The length distribution analysis also revealed a peak for ~9- to 14-base-long RNAs that were not considered for further analysis due to the difficulty in reliably mapping their genomic coordinates and because of limits in validating and studying the expression patterns of such small RNA products.

Removal of RNAs matching the highly expressed tRNAs, rRNAs, snoRNAs, and miRNAs did not eliminate the characteristic peaks corresponding to usRNAs (see Fig. S2 in the supplemental material). Moreover, the average cellular abundance of usRNAs was similar to that of canonical small RNAs (19 to 23 nt) containing Dicer-dependent products, suggesting that these usRNAs are stably expressed in cells (see the supplemental material). To gain further insights into the stability of usRNAs, we transfected us-K12-1 and K12-1 separately into HEK293 cells and monitored their expression levels (see Fig. S3 in the supplemental material). The half-life of us-K12-1 was unusually long (~72 h) compared to that of mammalian mRNAs (35) yet shorter than that of the K12-1 miRNA, which has a half-life over 96 h. The unexpected stability of us-K12-1 suggested that usRNAs might have protein partners that stabilize them. We also note that although time point 0 used identical concentrations of the two RNAs, the signal intensity of us-K12-1 was significantly lower than that of K12-1, highlighting the technological limits in detecting us-K12-1. Thus, it is likely that us-K12-1 is more abundant than its concentration estimated from Northern blots.

We also compared usRNAs from BCP1 to those of a small RNA library (8 to 18 bases) from an unrelated tissue sample (normal breast). We found that 24.8% (298) of the 15- to 18-base-long usRNAs from the breast library matched identically (100%) to those of the BCP1 library, substantiating that numerous usRNAs are generated repeatedly across various cell types. These observations led us to believe that a fraction of usRNAs are not transient degradation products but rather stable molecules which could have evolved to participate in biological processes.

Numerous miRNA-derived usRNAs preferentially associate with Ago2. Sequencing revealed that the us-K12-1 sequence is identical to the first 17 bases of K12-1 miRNA and revealed the presence of other miRNA subproducts. We hypothesized that relevant miRNA subproducts might associate with Ago effector proteins and therefore used raw data from a recent Ago immunoprecipitation study (8) to determine whether miRNA-derived usRNAs interact with Ago proteins. Of 205 Ago-associated miRNAs, 62 (30%) generate usRNAs that also associate with Ago in HEK293 cells. Similar to us-K12-1, these cellular usRNAs are almost always processed from positions 1 to 18 of the 5′ termini of miRNAs (Fig. 2A), indicating 5′-dependent and position-specific processing from mature or precursor miRNAs. We also found that, compared to Ago1, miRNA-derived usRNAs (17 to 18 nt) preferentially associate with Ago2 (Fig. 2B), the only human Ago with endonuclease activity, which is required for siRNA function. In comparison, miRNAs (21 to 23 nt) manifest a reverse trend (0.29-fold Ago2 versus Ago1 binding).

To examine whether us-K12-1 manifests Ago association patterns similar to those of the general population of usRNAs, we immunoprecipitated Ago-RNA complexes from us-K12-1- or K12-1-transfected cells using antibodies (2, 48, 51) against Ago1, Ago2, Ago3, and Ago4. Since Ago antibodies are expected to have different affinities of binding to the target Ago proteins, we realized that the amounts of RNA precipitated by different antibodies might not be identical. To minimize such bias caused by the antibodies, the ratio of the estimated binding affinities of each Ago protein to us-K12-1 and K12-1 was used to quantitate the pattern of Ago protein binding to us-K12-1 and K12-1. Consistent with the observed Ago association pattern using all usRNAs, calculation (see Materials and Methods) of relative affinities of Ago proteins with respect to K12-1 in the binding of us-K12-1 revealed that the relative affinity of Ago2 is more than those of Ago1, Ago3, and Ago4 (Fig. 2C). Taken together, these observations indicate that

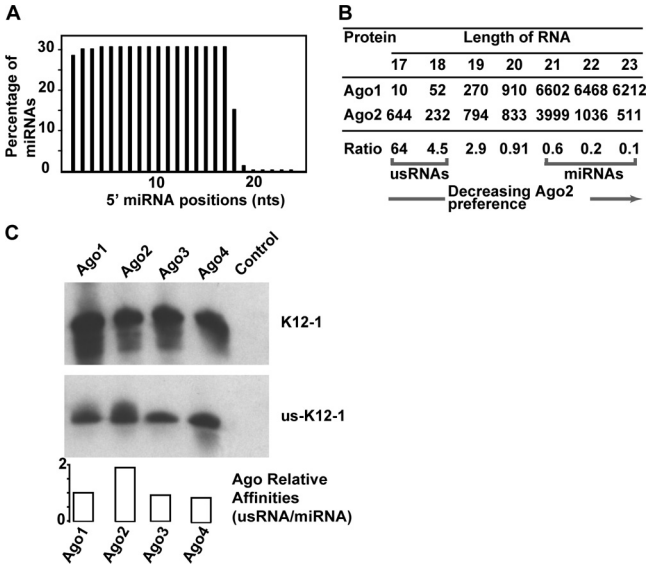

FIG. 2. Characteristics of Ago-associated miRNA-derived usRNAs. (A) Ago-associated miRNA-derived usRNAs primarily match to 5′ positions 1 to 17 of miRNAs and rarely map to 3′ ends of miRNAs; percentage of miRNAs that match to usRNAs (17 to 18 nt) is plotted at each miRNA position. (B) Number of sequence reads associated with Ago1 and Ago2 and their ratios (Ago2 to Ago1) indicate the preferential association of Ago2 with 17- to 18-base-long miRNA-derived usRNAs. (C) Northern blot of Ago-associated K12-1 (top) and us-K12-1 (bottom) based on immunoprecipitation of Ago 1 to 4. Control, irrelevant antibody. The estimated Ago relative affinities with respect to K12-1 in binding us-K12-1 are indicated (see Materials and Methods). Relative affinities are normalized to that of Ago1.

these miRNA-derived usRNAs might be functionally viable and may contribute to gene regulation to some extent.

us-K12-1 can regulate human RAD21. Using a miRNA target prediction method (19), we identified two potential miRNA K12-1 binding sites in the 3′ untranslated region (UTR) of human *RAD21* mRNA (Fig. 3A) encoding a ds break repair protein involved in the G_2/M checkpoint (11, 17). The first site in *RAD21* is strongly complementary to both K12-1 and us-K12-1 (positions 1 to 15), suggesting a possible RNA interference-mediated regulatory site. In contrast, the complementarity in the second region is limited (positions 2 to 8), consistent with a canonical miRNA binding site (7). To test whether *RAD21* is a direct target of us-K12-1 and K12-1, we transfected HEK293 cells with us-K12-1 or K12-1 ds oligonucleotides. Both miR-K12-1 and us-K12-1 were able to knock down the endogenous expression level of RAD21, both at the protein (Fig. 3B) and mRNA (Fig. 3C) levels. To investigate which of the two predicted target sites are responsible for the gene silencing effect, we used HEK293 cells transfected with a firefly luciferase reporter fused to the wild-type (WT) human *RAD21* 3′ UTR or to UTRs having mutations at the first (M1), second (M2), or both (M1 + 2) target sites. Transfection efficiencies were normalized using a *Renilla* luciferase reporter. Both us-K12-1 and K12-1 inhibit the WT reporter activity by 62% (Fig. 3D). While mutations in M1 markedly reduced repression by us-K12-1 and K12-1 miRNA, mutations at the M2 site reduced miRNA K12-1 activity (62% to 26%) but had little effect on us-K12-1-mediated repression (62% to 71%). When both target sites were mutated (M1 + 2), both us-K12-1 and miR-K12-1 inhibition of reporter activities were abolished. To further study the regulation, we performed dose-response assays for both K12-1 and us-K12-1 using the WT reporter. We found that the activity profiles of K12-1 and us-K12-1 were

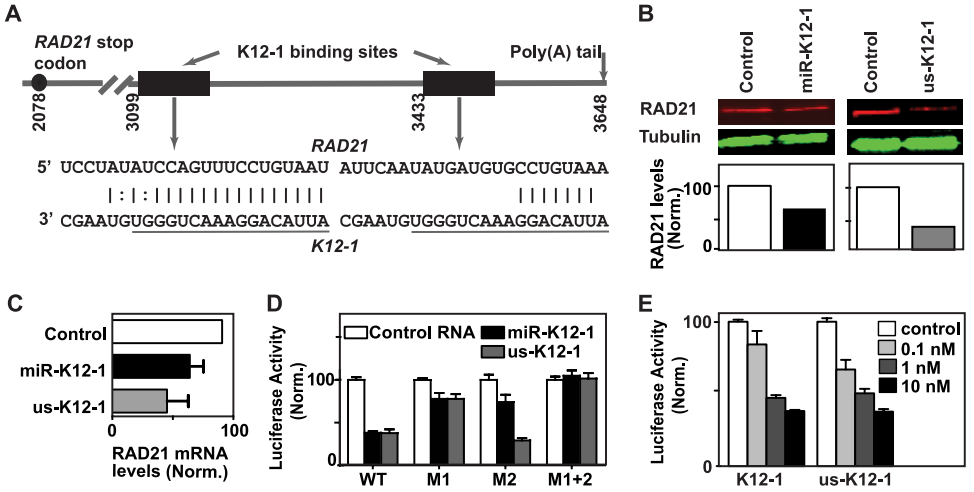

FIG. 3. Both miR-K12-1 and us-K12-1 can downregulate *RAD21*. (A) The 3′ UTR of the *RAD21* mRNA contains two putative binding sites for *K12-1* and us-K12-1 (underlined) at positions 3099 (site 1) and 3433 (site 2). (B) Immunoblots of endogenous RAD21 indicate that both miR-K12-1 and us-K12-1 downregulate the protein level of RAD21, in comparison to control RNA (Control). Tubulin levels confirm equal loading. For clarity, RAD21 levels are also normalized (Norm.) based on Tubulin signals (bottom). (C) Real-time RT-PCR of RAD21 transcripts demonstrate that miR-K12-1 and us-K12-1 also downregulate RAD21 mRNA levels. (D) Normalized firefly luciferase activity for WT or mutated (M1, M2, and M1 + 2) UTRs after transfection of K12-1 miRNA or us-K12-1. Note that although mutations in either binding site affect miR-K12-1 activity, us-K12-1 activity is largely unaffected by mutations in the canonical miRNA binding site (site 2). (E) Normalized firefly luciferase activity for WT UTR after transfection of either K12-1 miRNA or us-K12-1 at various concentrations (0.1 to 10 nM). All error bars represent standard error of the mean ($n \geq 3$).

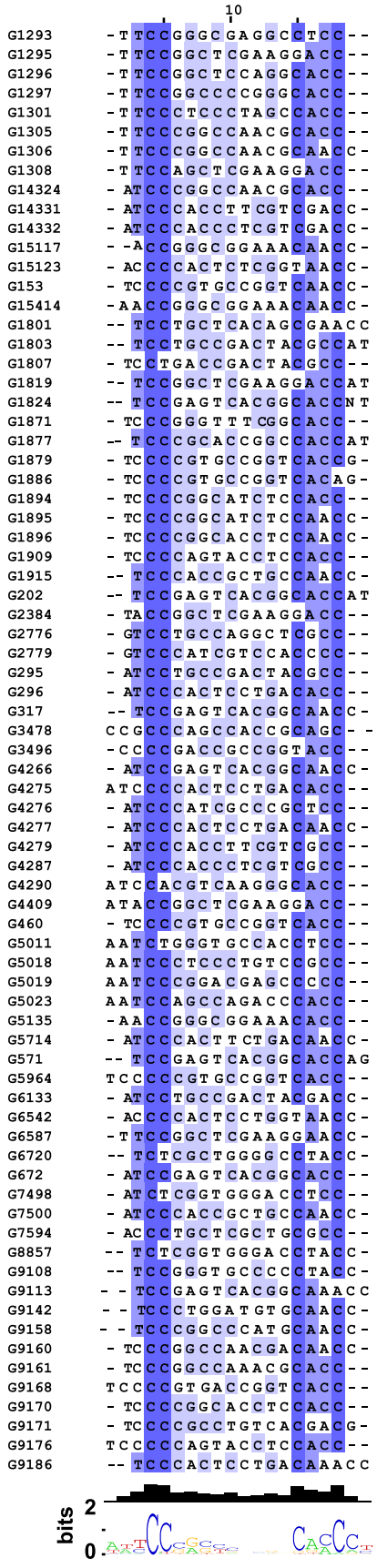

FIG. 4. Ago-associated non-miRNA-derived sequences containing the 5′ and 3′ C-rich motifs. The sequences (17 to 18 nt) are highly conserved at the 5′ and 3′ end. The WebLogo representation (bottom) of the motif reveals several similarities with that of the CACCA-containing RNAs, particularly with respect to the strong preferences for C's at the 3′ end, C's at 5′ positions 4 and 5, and a subdued yet relevant preference for A in CACC at the 3′ end (see also Fig. 5).

similar, and both RNAs regulated RAD21 in a dose-dependent manner (Fig. 3E). As noted earlier, since the endogenous concentration of us-K12-1 is likely much higher than its abundance estimated from Northern blots, us-K12-1 could regulate gene expression to a considerable extent.

usRNAs contain position-specific motifs. Next we analyzed the sequence characteristics of usRNAs to determine if they contain sequence motifs that may provide some insights into their underlying cellular processes. Since Ago-associated usRNAs are more likely to have biological activities, we initially focused on Ago-immunoprecipitated small RNAs from HEK293 cells and found an unusual CC signature motif (Fig. 4) (E value = 10^{-93}) present in 15.7% (75/479) of non-miRNA-derived usRNAs associated with Ago, which is not present in miRNA-derived usRNAs or in other small RNA classes (piRNAs, miRNAs, and TSSa-RNAs). Among the 75 Ago-associated sequences containing the CC signature motif, 35 RNAs contained a CACC subsequence, of which the majority (71.4%) harbored the CACC motif in a position-specific manner at the 3′ end (±2 bases).

Detection of the position-specific CACC motif in Ago-associated non-miRNA-derived usRNAs led us to expand the motif search to the complete usRNA population in BCP1 cells. We also included piRNAs and TSSa-RNAs as controls in our analysis. We did not detect any strongly represented sequence motifs in piRNAs and TSSa-RNAs, except for the known 5′ preference for uridine in piRNAs. However, a highly position-specific CACCA motif, which is similar to the motif present in Ago-associated usRNAs in HEK293 cells, was present in usRNAs from BCP1 cells; 50 percent (158/315) of CACCA-containing usRNAs from BCP1 cells embedded CACCA precisely at position 1 from the 3′ end, in contrast to proportions of 2.8% (18/632) for piRNAs and 9.5% (6/63) for TSSa-RNAs ($P = 1 \times 10^{-68}$ and 5×10^{-10}, respectively).

In addition to the CACCA motif, we find three nonredundant motifs in the 3′ ends of non-miRNA-derived usRNAs from BCP1 cells, which are also generally overrepresented in usRNAs (Fig. 5). Sixty-one percent (191/313) of the usRNAs encoding the subsequence CGCCA embedded CGCCA precisely at position 1 from the 3′ end, in contrast to 8.7% (15/173, $P = 3 \times 10^{-32}$) of piRNAs and 8.0% (6/75, $P = 5 \times 10^{-18}$) of TSSa-RNAs. Two less prominent motifs, GCCGC and GGGGG, are located approximately 10 bases away from the 3′ end in usRNAs. While the GCCGC motif is unique to usRNAs, the GGGGG motif is also weakly manifested in TSSa-RNAs, in which it is also located approximately 10 bases from the 3′ end. GGGGG-like motifs are overrepresented in promoter regions (23) and are preferentially located toward the transcription start sites of protein-coding genes. Thus, it is possible that the GGGGG motif-containing usRNAs represent transient degradation products or are TSSa-RNA-like prod-

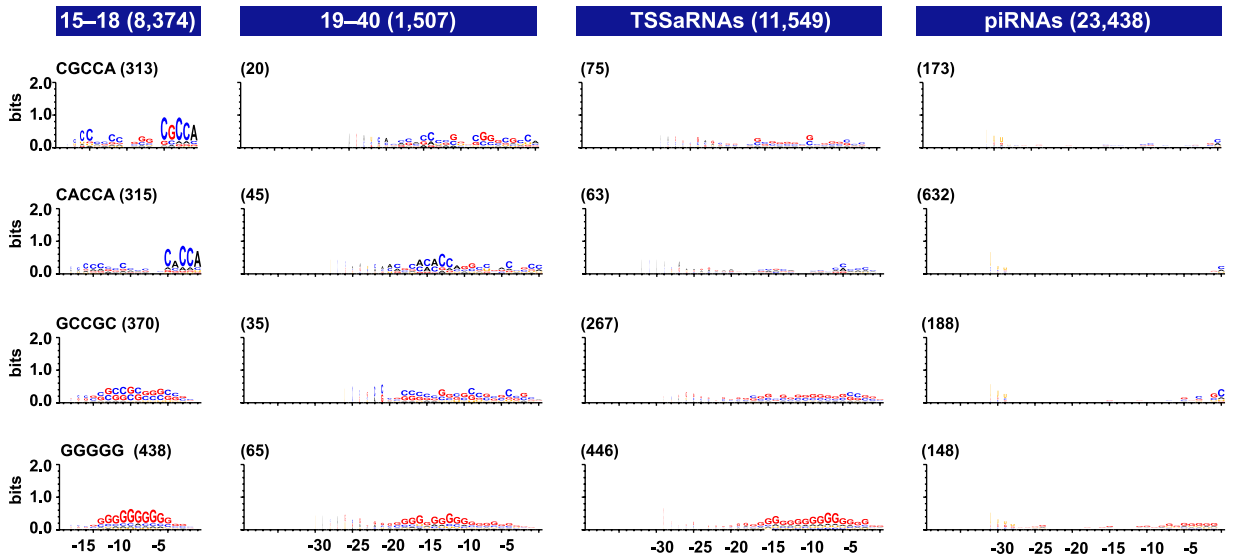

FIG. 5. The motifs CGCCA, CACCA, GCCGC, and GGGGG are preferentially located at specific 3′ positions of usRNAs and do not manifest such positional preferences in RNAs longer than 18 bases generally. The total number of RNAs (e.g., 23,438 piRNAs) used in the analysis and the total number of motif-containing RNAs (e.g., 313 CGCCA-containing RNAs) are indicated in parentheses.

ucts, while RNAs that harbor the consistently C-rich motifs, CGCCA, CACCA, and GCCGC, likely correspond to biologically relevant usRNAs.

DISCUSSION

The identification of a viral RNA product (usRK12-1) that is shorter than canonical miRNAs led us to investigate the characteristics of usRNAs. We found that a subset of miRNAs produce usRNAs specifically from the 5′ ends, which are also associated with Ago proteins and manifest a stronger preference for Ago2 than Ago1. The existence of miRNA-derived usRNAs is also supported by previously uninterpreted bands that are similar to those for us-K12-1 in Northern blots of several miRNAs (e.g., JC virus miR-J1 [40] and miR-122 in Huh7 cells [20]). Cumulatively, the results above and the observations that us-K12-1 can repress the RAD21 mRNA and that the majority of miRNA-derived usRNAs maintain the same 5′ seed region as the parental miRNAs suggest that miRNA-derived usRNAs may also contribute to gene regulation.

In addition to miRNA-derived usRNAs, we found hundreds of 15- to 18-base-long non-miRNA-derived usRNAs. We also found that many usRNAs harbor motifs that are both overrepresented and highly preferred at precise locations on their 3′ ends. Intriguingly, all three position-specific motifs that are overrepresented in usRNAs and the motif detected in Ago-associated 15- to 18-base-long RNAs are C rich. The absence of such motifs in the large class of TSSa-RNAs and piRNAs suggests that these motifs are not simply increasing the stability (e.g., exonuclease resistance) of RNA transcripts. Independent detection of similar motifs in Ago-associated non-miRNA-derived usRNAs further highlights the importance of such motifs. Other known small RNA motifs include a hexanucleotide element that affects nuclear localization of a miRNA (18), the terminal U in piRNAs and 21U-RNAs that may be relevant for interactions with Ago (14, 36), and the seed regions of miRNA family members, which determine their target specificities (27). Thus, it is likely that the motifs that we identified are not limited to a single function but rather reflect the diverse pathways in which small RNAs participate.

Comparison of usRNAs from an additional data set of RNAs from normal breast tissues with BCP1 usRNAs revealed the presence of hundreds of completely identical sequences in both libraries. The large (25%) overlap between the two unrelated libraries diminishes the possibility that such overlapping usRNAs are transient products and indicates that they likely are generated by a process that repeatedly, reproducibly, and accurately produces the same sequences in various tissues. Since miRNA-derived usRNAs possess 5′ ends that are identical to those of mature miRNAs that are produced by Dicer, the specific cellular process that generates them is likely dependent on the miRNA pathway and occurs arguably after the processing of the primary transcripts of miRNAs by Drosha. In contrast, non-miRNA-derived usRNAs are generated likely by other cellular mechanisms. One such possible mechanism is the stress-response pathway that induces tRNA-derived small RNAs (~30 to 50 nt) termed sitRNAs and tiRNAs that were found to occur in stress-induced cells (28, 50). The stress response pathway also overlaps with the miRNA pathway. When cells are subjected to stress, Ago proteins accumulate in cell structures known as stress granules, a process that requires miRNAs (26). Additional evidence linking small RNAs to stress responses was discovered in plants responding to biotic and abiotic stress factors (29, 31, 42). Taken together with the discovery of tRNA-derived small RNAs under stress and the reasoning that small RNAs provide an effective mechanism to fine-tune gene expression, we speculate that cellular stress response pathways might also play a role in the biogenesis of usRNAs.

The aforementioned observations and the recent finding that a degradation-like product from the ACA45 snoRNA

functions as a miRNA raise the possibility that many stably expressed usRNAs participate in biological processes. While miRNA-derived usRNAs might participate in fine-tuning gene expression, the cellular roles of non-miRNA-derived usRNAs remain unclear and require further study. Extrapolating the results of expression profiling of non-miRNA-derived usRNAs (see the supplemental material), we estimate that ~30% of non-miRNA-derived usRNAs that are generated across tissues could have evolved biological roles. In conclusion, the discovery of the diverse group of usRNAs and the recent discoveries of longer, stable, degradation-like products, such as ACA45-derived small RNA (8), and many other snoRNA-derived RNAs (43) that are Dicer dependent point to a biologic role for many of the previously ignored small RNAs that derive from long functional gene transcripts.

ACKNOWLEDGMENTS

We thank Gunter Meister and Mikiko C. Siomi for generously providing various Ago antibodies. We also thank members of the laboratories of Y.C./P.S.M. and B.J. for helpful discussions, Jason Boles for computer systems administration, Guodong Liu for help with preparing supplemental data, and Elane Fishilevich for comments on the manuscript.

This project was supported by funds from the University of Pittsburgh (B.J.), an NIGMS/NIH grant (B.J.), the American Cancer Society (B.J.), NCI/NIH grants (P.S.M. and Y.C.), and, in part, a grant from the Pennsylvania Department of Health (P.S.M. and Y.C.).

The funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Reference deleted.
2. **Beitzinger, M., L. Peters, J. Y. Zhu, E. Kremmer, and G. Meister.** 2007. Identification of human microRNA targets from isolated argonaute protein complexes. RNA Biol. **4:**76–84.
3. **Berezikov, E., G. van Tetering, M. Verheul, J. van de Belt, L. van Laake, J. Vos, R. Verloop, M. van de Wetering, V. Guryev, S. Takada, A. J. van Zonneveld, H. Mano, R. Plasterk, and E. Cuppen.** 2006. Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. Genome Res. **16:**1289–1298.
4. **Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G. J. Hannon.** 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell **128:**1089–1103.
5. **Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner.** 2004. WebLogo: a sequence logo generator. Genome Res. **14:**1188–1190.
6. **Czech, B., C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde, M. Dus, N. Perrimon, M. Kellis, J. A. Wohlschlegel, R. Sachidanandam, G. J. Hannon, and J. Brennecke.** 2008. An endogenous small interfering RNA pathway in Drosophila. Nature **453:**798–802.
7. **Doench, J. G., C. P. Petersen, and P. A. Sharp.** 2003. siRNAs can function as miRNAs. Genes Dev. **17:**438–442.
8. **Ender, C., A. Krek, M. R. Friedlander, M. Beitzinger, L. Weinmann, W. Chen, S. Pfeffer, N. Rajewsky, and G. Meister.** 2008. A human snoRNA with microRNA-like functions. Mol. Cell **32:**519–528.
9. **Farazi, T. A., S. A. Juranek, and T. Tuschl.** 2008. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Development **135:**1201–1214.
10. Reference deleted.
11. **Gachet, Y., S. Tournier, J. B. Millar, and J. S. Hyams.** 2001. A MAP kinase-dependent actin checkpoint ensures proper spindle orientation in fission yeast. Nature **412:**352–355.
12. **Gao, S. J., L. Kingsley, M. Li, W. Zheng, C. Parravicini, J. Ziegler, R. Newton, C. R. Rinaldo, A. Saah, J. Phair, R. Detels, Y. Chang, and P. S. Moore.** 1996. KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. Nat. Med. **2:**925–928.
13. **Ghildiyal, M., H. Seitz, M. D. Horwich, C. Li, T. Du, S. Lee, J. Xu, E. L. Kittler, M. L. Zapp, Z. Weng, and P. D. Zamore.** 2008. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science **320:**1077–1081.
14. **Girard, A., R. Sachidanandam, G. J. Hannon, and M. A. Carmell.** 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature **442:**199–202.
15. Reference deleted.
16. Reference deleted.
17. **Hoque, M. T., and F. Ishikawa.** 2002. Cohesin defects lead to premature sister chromatid separation, kinetochore dysfunction, and spindle-assembly checkpoint activation. J. Biol. Chem. **277:**42306–42314.
18. **Hwang, H. W., E. A. Wentzel, and J. T. Mendell.** 2007. A hexanucleotide element directs microRNA nuclear import. Science **315:**97–100.
19. **John, B., A. J. Enright, A. Aravin, T. Tuschl, C. Sander, and D. S. Marks.** 2004. Human MicroRNA targets. PLoS Biol. **2:**e363.
20. **Jopling, C. L., M. Yi, A. M. Lancaster, S. M. Lemon, and P. Sarnow.** 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science **309:**1577–1581.
21. **Klenov, M. S., S. A. Lavrov, A. D. Stolyarenko, S. S. Ryazansky, A. A. Aravin, T. Tuschl, and V. A. Gvozdev.** 2007. Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. Nucleic Acids Res. **35:**5430–5438.
22. **Lau, N. C., A. G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D. P. Bartel, and R. E. Kingston.** 2006. Characterization of the piRNA complex from rat testes. Science **313:**363–367.
23. **Lee, J., Z. Li, R. Brower-Sinning, and B. John.** 2007. Regulatory circuit of human microRNA biogenesis. PLoS Comput. Biol. **3:**e67.
24. **Lee, R. C., R. L. Feinbaum, and V. Ambros.** 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75:**843–854.
25. Reference deleted.
26. **Leung, A. K., J. M. Calabrese, and P. A. Sharp.** 2006. Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. Proc. Natl. Acad. Sci. USA **103:**18125–18130.
27. **Lewis, B. P., C. B. Burge, and D. P. Bartel.** 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell **120:**15–20.
28. **Li, Y., J. Luo, H. Zhou, J. Y. Liao, L. M. Ma, Y. Q. Chen, and L. H. Qu.** 2008. Stress-induced tRNA-derived RNAs: a novel class of small RNAs in the primitive eukaryote Giardia lamblia. Nucleic Acids Res. **36:**6048–6055.
29. **Liu, H. H., X. Tian, Y. J. Li, C. A. Wu, and C. C. Zheng.** 2008. Microarray-based analysis of stress-regulated microRNAs in Arabidopsis thaliana. RNA **14:**836–843.
30. Reference deleted.
31. **Lu, S., Y. H. Sun, R. Shi, C. Clark, L. Li, and V. L. Chiang.** 2005. Novel and mechanical stress-responsive microRNAs in Populus trichocarpa that are absent from Arabidopsis. Plant Cell **17:**2186–2203.
32. **Okamura, K., W. J. Chung, J. G. Ruby, H. Guo, D. P. Bartel, and E. C. Lai.** 2008. The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature **453:**803–806.
33. **Pak, J., and A. Fire.** 2007. Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science **315:**241–244.
34. Reference deleted.
35. **Ross, J.** 1995. mRNA stability in mammalian cells. Microbiol. Rev. **59:**423–450.
36. **Ruby, J. G., C. Jan, C. Player, M. J. Axtell, W. Lee, C. Nusbaum, H. Ge, and D. P. Bartel.** 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell **127:**1193–1207.
37. Reference deleted.
38. **Sai Lakshmi, S., and S. Agrawal.** 2008. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. Nucleic Acids Res. **36:**D173–D177.
39. **Schneider, T. D., and R. M. Stephens.** 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. **18:**6097–6100.
40. **Seo, G. J., L. H. Fink, B. O'Hara, W. J. Atwood, and C. S. Sullivan.** 2008. Evolutionarily conserved function of a viral microRNA. J. Virol. **82:**9823–9828.
41. **Stark, A., P. Kheradpour, L. Parts, J. Brennecke, E. Hodges, G. J. Hannon, and M. Kellis.** 2007. Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes. Genome Res. **17:**1865–1879.
42. **Sunkar, R., and J. K. Zhu.** 2004. Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell **16:**2001–2019.
43. **Taft, R. J., E. A. Glazov, T. Lassmann, Y. Hayashizaki, P. Carninci, and J. S. Mattick.** 2009. Small RNAs derived from snoRNAs. RNA **15:**1233–1240.
44. **Takeda, A., S. Iwasaki, T. Watanabe, M. Utsumi, and Y. Watanabe.** 2008. The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. Plant Cell Physiol. **49:**493–500.
45. **Tam, O. H., A. A. Aravin, P. Stein, A. Girard, E. P. Murchison, S. Cheloufi, E. Hodges, M. Anger, R. Sachidanandam, R. M. Schultz, and G. J. Hannon.** 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature **453:**534–538.
46. **Tolia, N. H., and L. Joshua-Tor.** 2007. Slicer and the argonautes. Nat. Chem. Biol. **3:**36–43.
47. **Vasudevan, S., Y. Tong, and J. A. Steitz.** 2007. Switching from repression to activation: microRNAs can up-regulate translation. Science **318:**1931–1934.

48. **Weinmann, L., J. Hock, T. Ivacevic, T. Ohrt, J. Mutze, P. Schwille, E. Kremmer, V. Benes, H. Urlaub, and G. Meister.** 2009. Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. Cell **136:**496–507.
49. **Xie, Z., E. Allen, N. Fahlgren, A. Calamar, S. A. Givan, and J. C. Carrington.** 2005. Expression of Arabidopsis MIRNA genes. Plant Physiol. **138:**2145–2154.
50. **Yamasaki, S., P. Ivanov, G. F. Hu, and P. Anderson.** 2009. Angiogenin cleaves tRNA and promotes stress-induced translational repression. J. Cell Biol. **185:**35–42.
51. **Zuma-Makai, A., H. Oguri, T. Mituyama, Z. R. Qian, K. Asai, H. Siomi, and M. C. Siomi.** 2008. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. Proc. NaH. Acad. Sci. USA **105:** 7964–7969.