1164–1177 *Nucleic Acids Research, 2013, Vol. 41, No. 2* *Published online 26 November 2012*
doi:10.1093/nar/gks1098

Profiling of T helper cell-derived small RNAs reveals unique antisense transcripts and differential association of miRNAs with argonaute proteins 1 and 2

Sumanth Polikepahad[1] and David B. Corry[1,2,*]

[1]Department of Medicine and [2]Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX 77030, USA

Received June 9, 2012; Revised October 15, 2012; Accepted October 21, 2012

ABSTRACT

RNA interference mediated through antisense transcripts is a fundamentally important mechanism regulating gene expression that remains incompletely understood. Here, we have used next-generation sequencing to determine from mouse CD4+ T cells the functional implications of antisense transcripts binding to argonaute (AGO) proteins that mediate RNA interference and post-transcriptional gene silencing. This effort identified 90 new microRNAs (miRNAs) and six endogenous hairpin RNA-derived small interfering RNAs (siRNAs) mapping to distinct introns. Unexpectedly, 69 miRNAs were expressed as non-canonical isomiRs as the dominant AGO-binding transcript, with extensive 3′ terminal nucleotide modifications. Furthermore, differential expression analysis between AGO1- and AGO2-bound miRNAs suggested preferential binding of isomiRs ending with 3′ adenine residues to AGO1 and 3′ uridine residues to AGO2. Analysis of the putative targets of all miRNAs suggested a striking preference for regulating transcription and transcription factors with additional evidence of a functional division of labor between AGO proteins in this regard. We further provide evidence that multiple mitochondrial genomic loci serve as the source of endogenous *cis*-natural antisense transcripts. These findings imply diversity in AGO protein function based on differential miRNA binding and indicate that RNA interference-based gene regulation is more complex than previously recognized.

INTRODUCTION

The argonaute proteins comprise an evolutionarily conserved family of RNA-binding proteins that are involved in the regulation of transcribed RNAs through RNA interference. The two major classes of argonaute proteins, AGO and PIWI, both bind to 19–35 nt short RNAs and regulate expression of genes and transposons in somatic and gonadal tissues, respectively. AGO proteins bind to microRNAs (miRNAs) or small interfering RNAs (siRNAs) within the RNA-induced silencing complex (RISC) to suppress target gene expression either by endonucleolytic cleavage or translational repression through a process termed post-transcriptional gene silencing (PTGS). AGO proteins are essential for embryological development and cellular differentiation, and in plants and primitive animals are required for host defense (1). The number of argonaute family proteins varies widely among species, ranging from 27 in *Caenorhabditis elegans* to eight in mice and humans (2).

AGO proteins further play crucial roles in the biogenesis of mature miRNAs. After Drosha cleavage of precursor miRNAs, the miRNA/miRNA* duplex binds to AGO proteins, a process termed 'RISC loading' to form the RISC complex. Numerous factors influence loading of small RNAs into AGO proteins, including Dicer processing, the structure of the small-RNA duplex, terminal (5′ and 3′) nucleotides, thermodynamic properties of the duplex and individual strands and the structure of the AGO protein itself (3). In *Drosophila*, Dicer 1 processes miRNA precursors, whereas Dicer 2 processes siRNA precursors. Accordingly, miRNAs differentially load into AGO1 and siRNAs into AGO2. However, notable exceptions to this rule in which some Dicer 1 produced miRNAs preferentially bind to AGO2 suggested the presence of a

*To whom correspondence should be addressed. Tel: +1 713 798 8740; Fax: +1 713 798 5780; Email: dcorry@bcm.edu

post-dicer processing mechanism that influenced AGO loading (3). The degree of base pairing in duplex small RNAs further appears to influence the loading of small RNAs into AGO proteins. siRNAs, which form fully complementary duplexes, preferentially bind AGO2, whereas miRNA duplexes that contain unpaired bulges preferentially load into AGO1. AGO1 further favors a terminal U and AGO2, 5′C residues (4). Furthermore, strict strand selection of miRNAs is observed with miRNA mature/guide strands binding to AGO1 and miRNA*/passenger strands binding to AGO2 (3,5). This preferential binding is primarily attributed to the structure of the miRNA/miRNA* and siRNA duplexes where the strand with the least 5′ thermodynamic stability is selected as the guide strand (5).

Mice and humans possess four biochemically distinct AGO proteins, AGO1-4. Although functional differences among these proteins remain largely speculative, AGO2 likely possesses a unique slicer, i.e. endonucleolytic, activity required to cleave target mRNAs (6). Differential loading of miRNAs into AGO proteins has been reported from transformed human cells (7), but mammalian AGO proteins are nonetheless believed to be functionally redundant (3). In addition to miRNAs and siRNAs, next-generation sequencing (NGS) technologies have enabled the discovery of many additional classes of non-coding RNAs (ncRNAs) from animals and plants with potential to participate in PTGS, including promoter-associated RNAs (paRNAs), transcription start site-associated RNAs (TSSaRNAs), 5′ capped promoter-associated small RNAs (PASRs), intron-derived miRNAs, i.e. mirtrons, short hairpin RNAs (shRNAs), antisense miRNAs (asmiRNAs), animal endogenous siRNAs (endo-siRNAs), natural antisense transcripts (NATs) and many others (8). Although the molecular mechanisms and AGO-associated functional pathways of most of these ncRNAs remain undefined, their manifold existence confirms the highly complex nature of PTGS.

In this study, we have identified multiple RNA transcripts that bind to AGO1 and AGO2 proteins in mouse splenic CD4+T cells. By combining co-immunoprecipitation of AGO proteins with NGS and advanced bioinformatics approaches, we show the differential AGO binding of several known miRNAs and their isomiRs, in addition to the discovery of many novel miRNAs, hpRNA derived endo-siRNAs and novel NATs deriving from the mitochondrial genome.

MATERIALS AND METHODS

Isolation of CD4+T cells and argonaute bound RNA co-immunoprecipitation

Spleens were isolated from eight (two control, three AGO1 and three AGO2) 5 weeks old C57BL/6 mice and CD4+T cells were extracted by using immunomagnetic selection. Twenty million cells per each mouse were used for co-immunoprecipitation with Rabbit IgG (Millipore, Billerica, MA, USA) control, anti-AGO1 (Millipore) and anti-AGO2 (Wako, Richmond, VA, USA) monoclonal antibodies.

Preparation of deep sequencing libraries

Total RNA including small RNAs were extracted by using miRNAEasy kit (Qiagen, Valencia, CA, USA) and quality assessed by 260/230 nm absorbance ratios. Small-RNA transcripts of 15–40 nucleotides were gel purified after running 5 μg of total RNA on 15% [Tris-borate-EDTA (ethylene diamine tetraacetic acid) buffer (TBE)]-urea polyacrylamide gels. A synthetic Illumina-specific 26-residue adapter RNA oligonucleotide (5′-GUU CAG AGU UCU ACA GUC CGA CGA UC-3′) was ligated to the 5′-end of the small RNAs. The ligated small RNA was gel purified to remove free adapter. A synthetic Illumina-specific 22-residue 3′-adapter with inverted dideoxythymidine (idT) added at the 3′-end (5′-p UCG UAU GCC GUC UUC UGC UUG idT-3′) was ligated to the 5′-ligated small RNA and gel purified. The resultant RNA library was reverse transcribed and amplified by PCR for 19 cycles using adapter-specific primers. The PCR products were then sequenced using Illumina-Genome Analyzer IIx, at the Institute for Molecular Design Center—University of Houston.

Read processing, mapping and aligning and post-alignment processing of deep sequenced libraries

Raw sequences were uploaded to NCBI Sequence Read Archives (SRA) and are available through accession number SRA056111. The quality of raw sequences was evaluated using FASTX-Toolkit version 0.6, generating quality statistics and box-whisker plots to summarize the base call quality score range at each position in all eight sequence files (Supplementary Figure S1). We considered base quality as poor if the median score was <28 or lower quartile was <20. Based on these criteria, read quality was very high for positions 1–30 (with median scores ranging between 30 and 40), but declined thereafter in all libraries with each additional nucleotide. As expected, we observed an overrepresentation of adapter sequences at positions 31–40, which were removed from all full-length reads by using a custom 'perl' script that searches for and removes a minimum of 8 nt of perfectly matched adapter sequence from the sequenced read. The region that was most frequently deleted was the 31–40 segment, confirming the over-representation of adapter sequences. A tag file for each library was then created by collapsing unique reads and calculating tag counts for each read using the same 'perl' script.

The collapsed reads were first mapped to the known mouse miRNA database by using standalone miRanalyzer version 0.2 which uses Bowtie 2.0, an ultra-fast memory-efficient short read aligner (9,10), using the following parameters: –best –strata –e –k10 –l17 –n1. For non-microRNA reads, we used Bowtie 2.0 directly for mapping to other transcript classes. This is mainly because miRanalyzer is mainly made for identifying miRNAs and hence has a maximum read length limitation of 28, precluding mapping of longer reads to longer transcripts. We first mapped these reads to mouse genome [NCBI Build 37, University of California at Santa Cruz (UCSC) mm9] using the following parameters: –best –k10 –v1. Reads mapping 10 or more times to the genome were

discarded. The mapped reads were aligned to the mouse transcriptome by using BEDTools suite (11). For miRNA and other transcript mapping, we selected a minimum length of 17 nt. Reads aligning to more than one RNA class or more than one type of transcript in each class were considered ambiguous and are not shown.

miRBASE v. 18 indexes were downloaded from the miRanalyzer website http://bioinfo2.ugr.es/miRanalyzer_dev/DB/mature18.zip; mouse protein coding genes and repeat sequences, UCSC genome browser; ncRNA, ENSEMBL browser and RFam database; PASR, deepBase database. Visualization of the mapped reads was performed on the UCSC genome browser. RNA folding of pre-miRNA and hpRNAs to predict hairpin structures and calculate minimum free energy (MFE) levels was performed on the RNAfold server (12). Multiple alignments of the precursor miRNAs, canonical isomiRs and detected miRNAs, and multiple species seed sequences of novel miRNAs (candidate_21) were performed by the clustalw2 program.

Differential expression analysis

We first checked for the consistency of the data among biological replicates in each group by performing unsupervised clustering analysis that measures Euclidean distances among the samples, which clearly revealed that replicates from each group cluster together (Supplementary Figure S2). Next, differential expression analysis between AGO1 and AGO2 samples was performed by the *Empirical analysis of digital gene expression data in R* (EdgeR) package version 2.4.6 (13). By using this package, first the tags with >5 counts per million (CPM) were removed in each library, the resultant reads were normalized by the Trimmed Mean of M component method, normalization factors were calculated, effective library sizes were estimated by using normalization factors, fitted according to the Negative Binomial model, estimates of dispersions were made by using the quantile-adjusted conditional maximum likelihood method and differential expression tests were performed. The resultant *P*-values were adjusted for false discovery rate (FDR) by using Benjamini and Hochberg's approach and only adjusted *P*-values with <0.05 were considered statistically significant for the miRNA differential expression analysis. To generate heatmaps, the log count values of each miRNA or isomiR were first moderated by the variance stabilizing transformation (VST) method outlined in the DESeq version 1.6.1 package manual and heatmaps were created by the heatmap2 function of the gplots2.10.1 package (14). The VST data were also used to calculate the Euclidean distances among samples and sample clustering, which showed similarity among different samples (14).

Target prediction and gene ontology enrichment

The targets of differentially expressed miRNAs were predicted by using TargetScanMouse 6.0. Predicted targets that were targeted by at least two miRNAs and with <−0.10 total context score (2582 targets in AGO1 and 1549 targets in AGO2) were targeted for GO enrichment, which was performed by the 'functional annotation tools' function of Database for Annotation Visualization, and Integrated Discovery (DAVID) bioinformatics tool (15). The enrichment was performed using the medium classification stringency option against the total mouse gene population background (15).

Novel miRNA discovery

Novel miRNAs were discovered by using miRanalyzer version 0.2 (9,16). This program uses stringent criteria for putative miRNA discovery by employing several prediction models. For accuracy, only those miRNAs were considered novel that were predicted by at least three different prediction models and expressed in at least three of the six samples (three AGO1 and three AGO2).

Statistical analysis

Data are presented as means ± SEM. Quantitative real time PCR data are presented using the $\Delta\Delta C_t$ relative expression method. The U6RNA was used as an endogenous control and data were normalized to the control IgG sample. Differential expression analysis of isomiRs and primary-isomiRs was performed by the EdgeR package as described earlier with *P*-values adjusted using Benjamini and Hochberg's approach to correct for FDR (13). For determining the significance of 3′ and 5′ terminal nucleotide occurrences in differentially expressed isomiRs, chi-square goodness-of-fit test with expected terminal nucleotide frequency was calculated from all 6141 isomiRs using Graphpad Prism version 5.01 software. Gene ontology (GO) enrichment analysis was performed by the DAVID bioinformatics tool with *P*-values calculated as previously described (15). For all analyses, a *P*-value of <0.05 was considered significant.

RESULTS

AGO1 and AGO2 bind distinct miRNAs and other small RNAs

Although the specificity and lack of cross-reactivity between AGO1 and AGO2 antibodies have been reported (17–19; http://condense.it/Millipore_AGO1 and http://condense.it/Wako_AGO2), we conducted additional studies to confirm the specificity of anti-AGO antibodies. First, we performed qRT-PCR for *let-7f* using small RNAs derived from AGO1- and AGO2-specific immunoprecipitates of CD4+ T cells as compared with non-specific IgG immunoprecipitates (Figure 1A). This analysis indicated a marked differential binding of *let-7f* to AGO1 and AGO2 as compared with the relatively weak and non-specific binding in association with rat IgG. Second, the length distribution analysis of the sequenced and trimmed reads selected for mapping revealed predominant binding of both AGO1 and AGO2 to transcripts of between 21 and 23 nucleotides; again, only random, low-grade transcript binding was observed with IgG (Figure 1B). Third, the vast majority of transcripts binding to AGO1 (74.4%) and especially AGO2 (97.2%) were miRNAs, comprising a total of 682

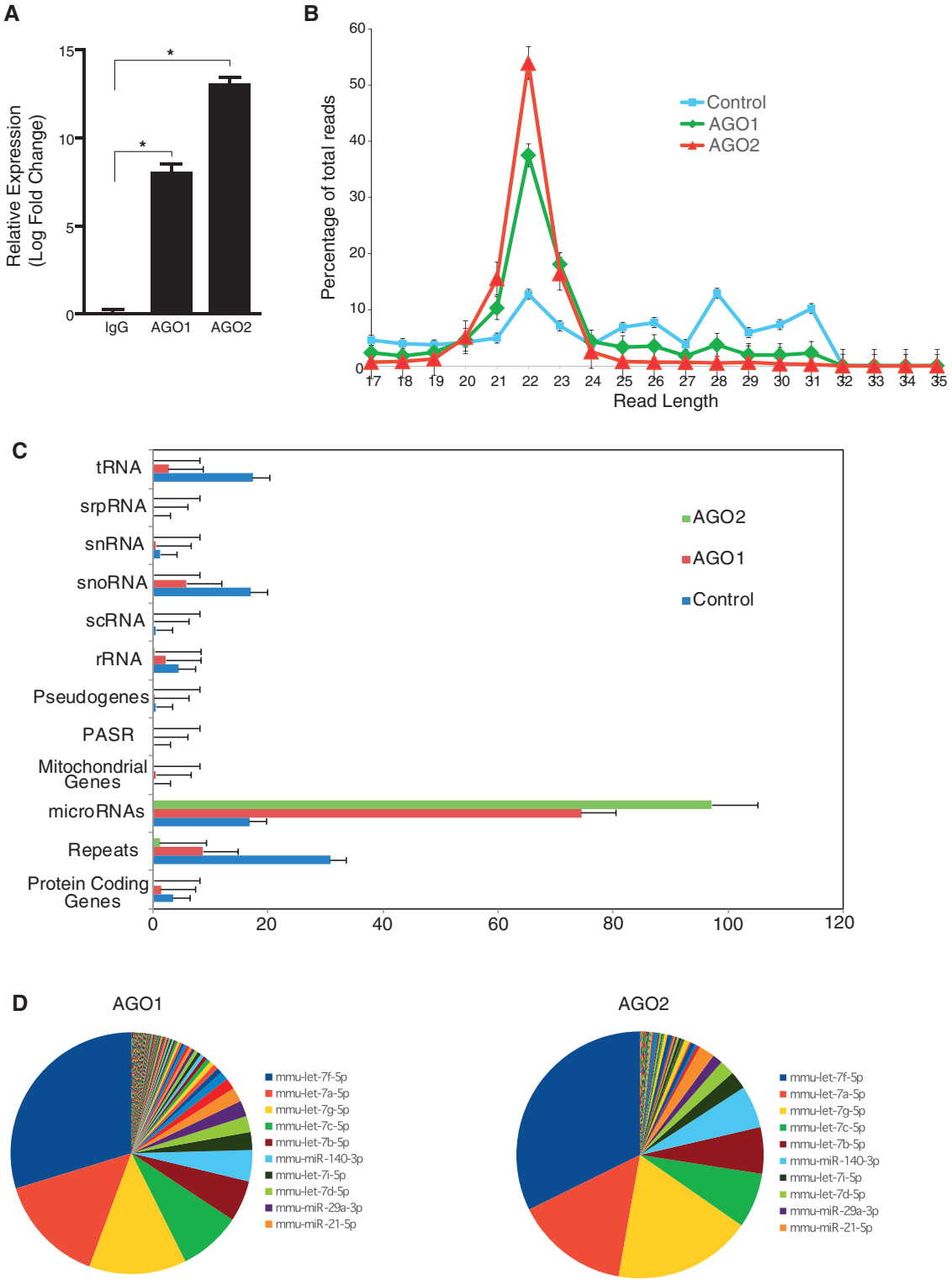

Figure 1. AGO1 and AGO2 bind predominantly to miRNAs from T helper cells. (**A**) Quantitative RT-PCR of miRNA let-7f from the co-immunoprecipitated small-RNA fraction of mouse splenic CD4+T cells with data normalized first to U6RNA within each group and then to IgG controls. The data were analyzed by $\Delta\Delta C_t$ method and $-\Delta\Delta C_t$ values are shown. $n = 3$, $^*P < 0.05$. (**B**) Size distribution of RNA sequences binding to AGO1, AGO2 and control IgG plotted as a percent of total reads. (**C**) Percentage of mapped reads plotted as a function of transcript types. (**D**) Piecharts showing the distribution of average percentages of various types of miRNAs in AGO samples. Legend includes the top 10 most highly expressed miRNAs.

mature miRNAs, with 591 binding to AGO1 and 637 to AGO2 (Figure 1C and D; Supplementary Table S1). In contrast, although substantial numbers of transcripts were found in association with control immunoglobulin, no dominant (i.e. comprising >50% of all transcripts) transcript class was apparent. Repeat elements (29.6%) were the most abundant transcript class found in association with control rat IgG with miRNAs comprising an average of 14.8%. Thus, these data suggest that miRNAs specifically and preferentially associate with both AGO1 and AGO2, in marked contrast to the non-specific transcript binding to control immunoglobulin.

These findings support the specificity of the anti-AGO antibodies in comparison to IgG control, but do not fully address the specificity of the anti-AGO antibodies themselves. Thus, we next analyzed the specificity of transcript binding to AGO1 and AGO2. For this purpose, we performed sample clustering, which calculates the Euclidian sample to sample distance after VST of the count data after removing <5 CPM reads from each sample (the resulting data of which was used for differential expression analysis), by using the *dist* function of the DESeq package. During this clustering, the dispersion of count data was estimated in the 'blind' mode, i.e. the variance-stabilized transformation was not 'aware' of the experimental design, to avoid bias. This analysis clearly demonstrated that biological replicates of the AGO1 or AGO2 immunoprecipitations cluster together, strongly indicating that these antibodies lack cross reactivity (Supplementary Figure S2). Notably, the most abundant miRNAs for both groups were of the let-7 family (Figure 1D; Supplementary Table S2), which parallels observations from whole mouse and human lung (20) and D. Corry *et al*. (unpublished data).

To further analyze the transcript binding specificity of AGO1 and AGO2, we determined the differential association of distinct transcript species with the argonaute proteins. Although AGO1 and AGO2 both predominantly bound miRNAs, AGO2 binding of miRNAs was far more robust ($P < 1.40^{-7}$; Supplementary Table S1). In contrast, AGO1 bound a greater diversity of small RNA species, including tRNA, srpRNA, scRNA, snoRNA and snRNA, with greater avidity relative to AGO2 ($P < 0.01$ for each transcript). Together, these studies suggest that, although similarities in transcript binding are clear, AGO1 and AGO2 are functionally non-redundant.

Dominant expression of non-canonical isomiRs for some miRNAs

Several recent NGS-based studies have suggested that the miRNA-generating arm of the pre-miRNA produces several sequence variants of the 'canonical' or 'reference' miRNA (miRBASE sequences for each miRNA), known as isomiRs (21,22). We identified 92 019 isomiRs for 682 known miRNAs in at least one of the six T-cell samples. Most were of low copy number and after filtering the data for low copy (<5 CPM) and potentially spurious counts (variable counts among the replicate samples), we revised the isomir count to 6140. There were 214 miRNAs having at least two isomiRs, with let-7 family members occupying 7 of the top 10 most abundant, and let-7f-5p having the highest number of isomiRs (589) followed by miR-140-3p (551) (Figure 2A). The sequences, raw counts and percent abundance of all isomiRs are shown in Supplementary Table S3.

Next, we determined if the highest expressed isomiR for each miRNA was the same as the canonical miRNA. For this analysis, we first isolated the most abundant isomiR for each miRNA in each of the six libraries and then selected only those isomiRs that were expressed in either all three replicates of each of the two groups, i.e. AGO1 and AGO2, or expressed in all six libraries. This yielded 130 miRNAs, of which 69 isomiRs (~53%) were non-canonical in at least one of the two groups (Supplementary Table S3). Among these 69, the most highly expressed isomiR differed for each AGO sample (Supplementary Table S4). Most of these isomiRs occupied >50% of the fraction, indicating their predominant expression. Some miRNAs, such as let-7b and let-7i were some of the most highly expressed of all miRNAs. Intriguingly, no canonical sequences were found for 17 miRNAs. Even where the canonical miRNA was the primary isomiR, the non-canonical miRNA fraction remained robust. For example, at least 40% of the let-7 miRNA fraction consisted of non-canonical isomiRs. Furthermore, all modifications (i.e. edits) of the highly expressed isomiRs occurred exclusively at the 3′ termini. Isolated 5′ modifications were completely absent from our libraries, and although 5′ modifications in combination with 3′ modifications were detected, they were exceptionally rare (total of three examples found). Additional modifications included single and multiple nucleotide extensions, deletions and substitutions. The most common (~28%) of the 3′ modifications was a single U extension. Nine isomiRs showed 3′ substitutions of nucleotides, representing 3′ edits that do not map to the mouse genome (Supplementary Table S4). Exemplary alignments are shown in Figure 2B.

Preferential binding of isomiRs with 3′A to AGO1 and 3′U to AGO2

To determine if isomiRs exhibit preferential binding to AGO proteins, we performed differential expression analysis. Comparing isomiRs associating with AGO1 and AGO2, only isomiRs that showed >3 log-fold difference and adjusted *P*-value <0.05 were considered to be significantly differentially associated. Of 6140 isomiRs, 614 (10%) corresponding to 127 miRNAs, showed differential association, with 392 showing preference to AGO1 and 222 to AGO2 (Figure 3A and B**).** Fold changes in expression and all *P*-values are shown in Supplementary Table S5. Of these 392 isomiRs binding to AGO1, 53.4% possessed adenine as the 3′ terminal nucleotide and 58.8% of 222 isomiRs binding to AGO2 had uracil as the 3′ terminal nucleotide (Figure 3C). As for 5′ nucleotides, both AGO1- and AGO2-binding isomiRs had uracil as the predominant nucleotide (Figure 3D). These observations suggest that loading of isomiRs within argonaute proteins in part depends on terminal adenine and uracil

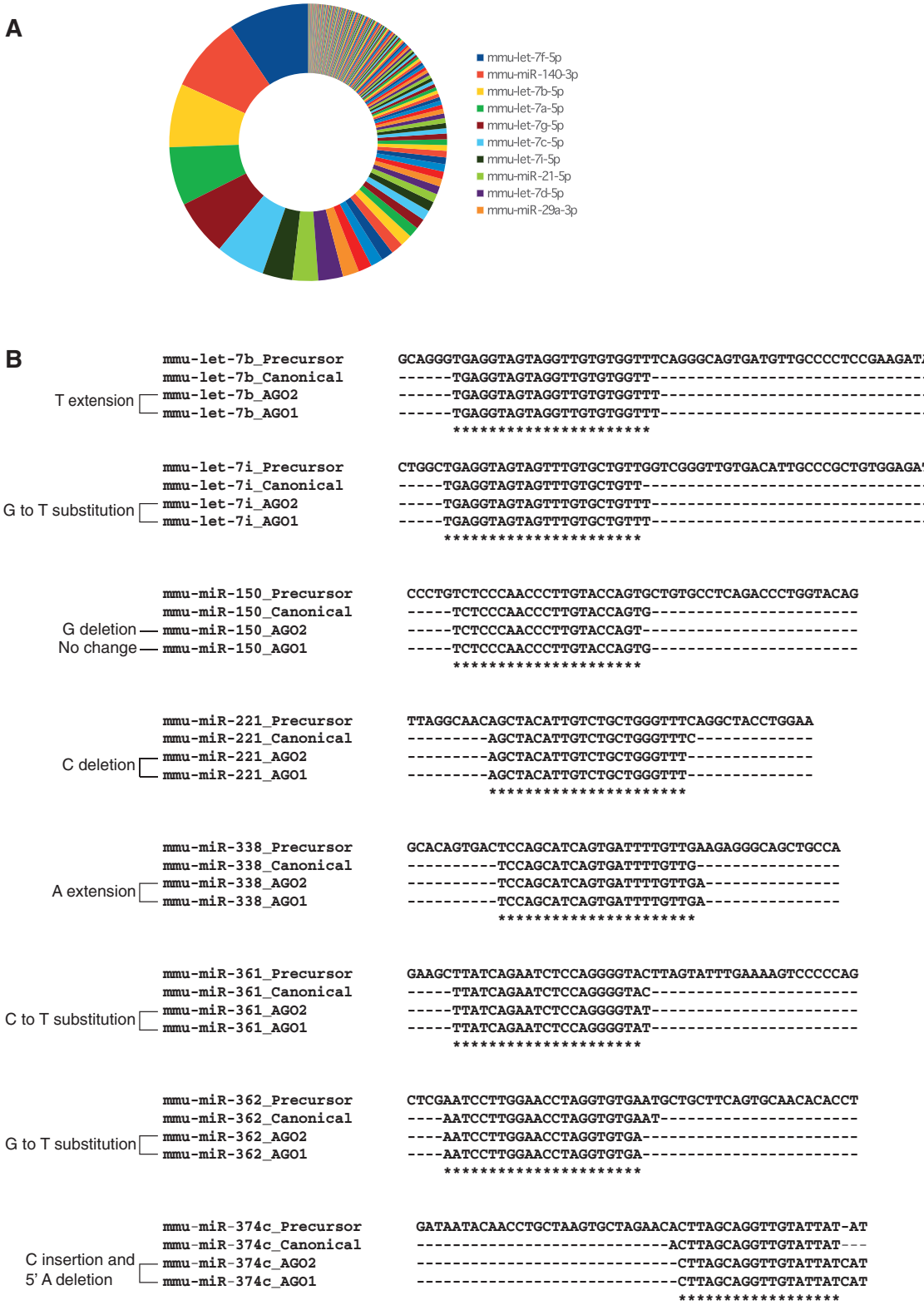

Figure 2. Distribution of isomiR types and 3′ modifications of primary isomiRs. (**A**) Relative abundance of isomiRs after filtering for low copy number and spurious reads. The legend indicates the top 10 miRNAs with the most abundant isomiRs. (**B**) Representative 3′-end modifications of the most abundant isomiRs, including extensions, deletions and substitutions.

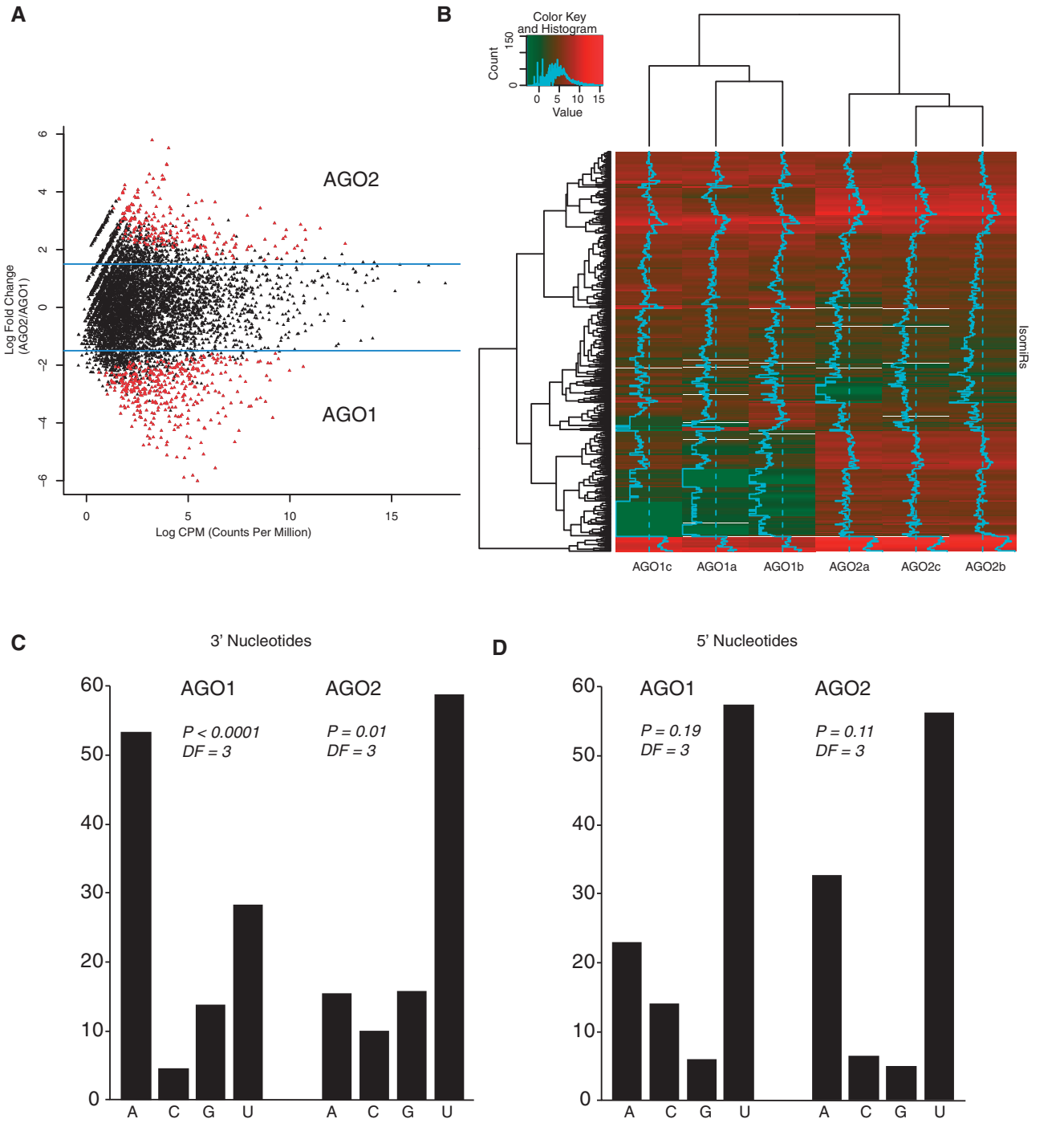

Figure 3. Differential expression of isomiRs and associated 3′ terminal nucleotides between AGO1 and AGO2. (**A**) Smearplot showing the differential expression of isomiRs between AGO1 and AGO2 samples. Reads that are at least 1.5 log-fold different with adjusted *P*-value < 0.05 are indicated in red. (**B**) Heatmap and hierarchial clustering of the differentially expressed isomiRs indicating clearly distinct AGO1 and AGO2 groups. Cyan colored lines depict the deviation of the absolute count value for each isomiR from the mean count. The color key index plots the normalized value against the isomiRs counts having the same value. (**C**) 3′-end nucleotides of differentially expressed isomiRs between AGO1 and AGO2 samples. The four types of nucleotides on the *x*-axis are plotted against the percentage of isomiRs ending with the corresponding nucleotide in all the differentially expressed isomiRs with at least 1.5 log-fold change. Chi-squared goodness-of-fit test was performed with expected frequencies of 3′-end nucleotides calculated from all the 6140 isomiRs. Chi-squared test, AGO1 $P < 0.0001$, DF = 3 and AGO2 $P = 0.01$, DF = 3. (**D**) 5′-end nucleotides of differentially expressed isomiRs between AGO1 and AGO2 samples. The four types of nucleotides on the *x*-axis are plotted against the percentage of isomiRs ending with the corresponding nucleotide in all the differentially expressed isomiRs with at least 1.5 log-fold change. Chi-squared test, not significant.

expression, with cytosine and guanine expression being of far less importance. The findings further indicate that AGO1 and AGO2 bind distinct isomiRs based in part by terminal nucleotide sequence.

Functional sorting of miRNAs between AGO1 and AGO2

Next, to begin to understand possible functional differences in AGO proteins, we assessed the differential binding of miRNAs to AGO1 and AGO2. For this purpose, we extracted isomiRs that were of highest expression in all six libraries and further filtered the data for low copy number (<5 CPM) and spurious counts. The final dataset yielded 295 miRNAs of 682 total. Differential expression analysis of these miRNAs yielded 74 miRNAs (55 AGO1; 19 AGO2) that were at least 1.5-fold different and with adjusted *P*-value <0.05 (Figure 4A and B; Supplementary Table S6). Of note, the 3′ terminal nucleotide bias remained.

We then predicted the targets of these 74 miRNAs by TargetScanMouse 6.0. To avoid non-specific targets, we selected targets with total context scores of <−0.10 and targeted by at least two miRNAs, after which GO enrichment by the DAVID algorithm was performed under medium stringency against total mouse gene population background. Targets belonging to both AGO1 and AGO2 were very significantly enriched for the term 'transcription regulation' (Table 1) with the majority of targets belonging to the categories of transcription factor, promoter binding gene, DNA binding, transcriptional co-activator and transcriptional co-repressor (Supplementary Table S7). To determine if these miRNAs regulate common genes, we identified the common genes of the top annotation cluster with the lowest *P*-value between the AGO1 and AGO2 groups and found that ~129 genes were shared between these two groups (34% of 390 AGO1; 57% of 228 AGO2) (Supplementary Table S7). These findings indicate that distinct sets of miRNAs acting through two different AGO proteins co-regulate transcription by employing more than one-third of the same genes.

Mitochondrial cis-NATs

Cis-NATs are sense–antisense transcripts transcribed from the same genomic locus, but from both directions (23). We found several sense–antisense transcripts from all T-cell libraries that overlap by at least 18 nt and which derive from mitochondrial loci encoding eight distinct mitochondrial genes and which associated with both AGO1 and AGO2 (Supplementary Table S8). Five of eight *cis*-NAT sequences are apparently transcribed from regions encoding five distinct NADH dehydrogenase subunits (Mt-ND2-6). Two additional *cis*-NAT sequences derive from tRNA encoding regions (Mt-Tq and Mt-Tp) and one from the cytochrome-b gene (Table 2; Figure 5; Supplementary Figure S3). The reads mapping to the opposite strand of these functionally important genes are complementary to ESTs and/or GenBank mRNAs, the complete sequences and functions of which remain unknown. It should be noted that these reads map only to these locations and nowhere else in the mouse genome.

Hairpin-precursor derived endogenous T-cell siRNAs

We have further discovered sequences that arise from long double-stranded stem loop precursors at six potential locations (*Vpas13a, Arhgap17, Cpne5, Spata6, Tcf20 and Aftph*) (Figure 6A). All sequences derive from intronic regions and are candidate endogenous-siRNAs based on previously published criteria (Supplementary Table S9) (24). Most of these sequences derive from sample AGO2c, which contained the highest number of usable reads. In agreement with a prior observation, this suggests that endogenous siRNAs are profoundly rare transcripts that are only to be found in the most comprehensively sequenced samples (24). Despite their low copy number, two candidate siRNAs appear in more than one sample at the exact location. The *Vps13a*-derived siRNA is particularly interesting because of its high read count and appearance in five of the six samples at the same location. All putative siRNAs satisfy the following criteria: (i) mapping only to one genomic locus; (ii) mapping sense and antisense to proximal intronic regions (Figure 6B); (iii) the read length ranges from 18 to 24 nt. Moreover, these sequences fail to meet criteria for miRNAs as the pre-miRNA stem loop structures do not form perfectly complementary double stranded stems. Also unlike miRNAs, endogenous siRNAs exhibit no species conservation, i.e. the sequences are only found in the mouse (Figure 6B). Endogenous siRNAs further failed to map to spliced ESTs, suggesting that they are distinct from NATs (data not shown).

Many novel miRNAs discovered

Finally, by using standalone miRanalyzer software, we have discovered 90 novel miRNAs that are expressed in at least three of the six samples. The complete list is shown in Supplementary Table S10. A representative novel miRNA having species conservation within the seed sequence is shown in Figure 7.

DISCUSSION

More than 90% of the eukaryotic genome is believed to be transcribed to RNA, but <2% of such transcripts encode functional proteins (8). Recent discoveries of the important roles played by several types of ncRNAs including miRNAs, piRNAs and siRNAs have clearly suggested pivotal roles in multiple biological processes. Moreover, advanced bioinformatics approaches combined with increasing power of deep sequencing technologies have made it possible to discover novel types of ncRNAs that markedly enrich the ncRNA databases. In this study of mouse T helper cells, we have identified three types of antisense ncRNAs, including miRNAs, *cis*-NATs and endogenous siRNAs, including many novel sequences. Moreover, we have identified isomiRs that demonstrate differential binding to AGO1 and AGO2, a property that appears in part to be due to editing of the 3′-end. Together, our findings point to an extraordinarily complex role played by ncRNAs in regulating T-cell function, especially regarding transcription.

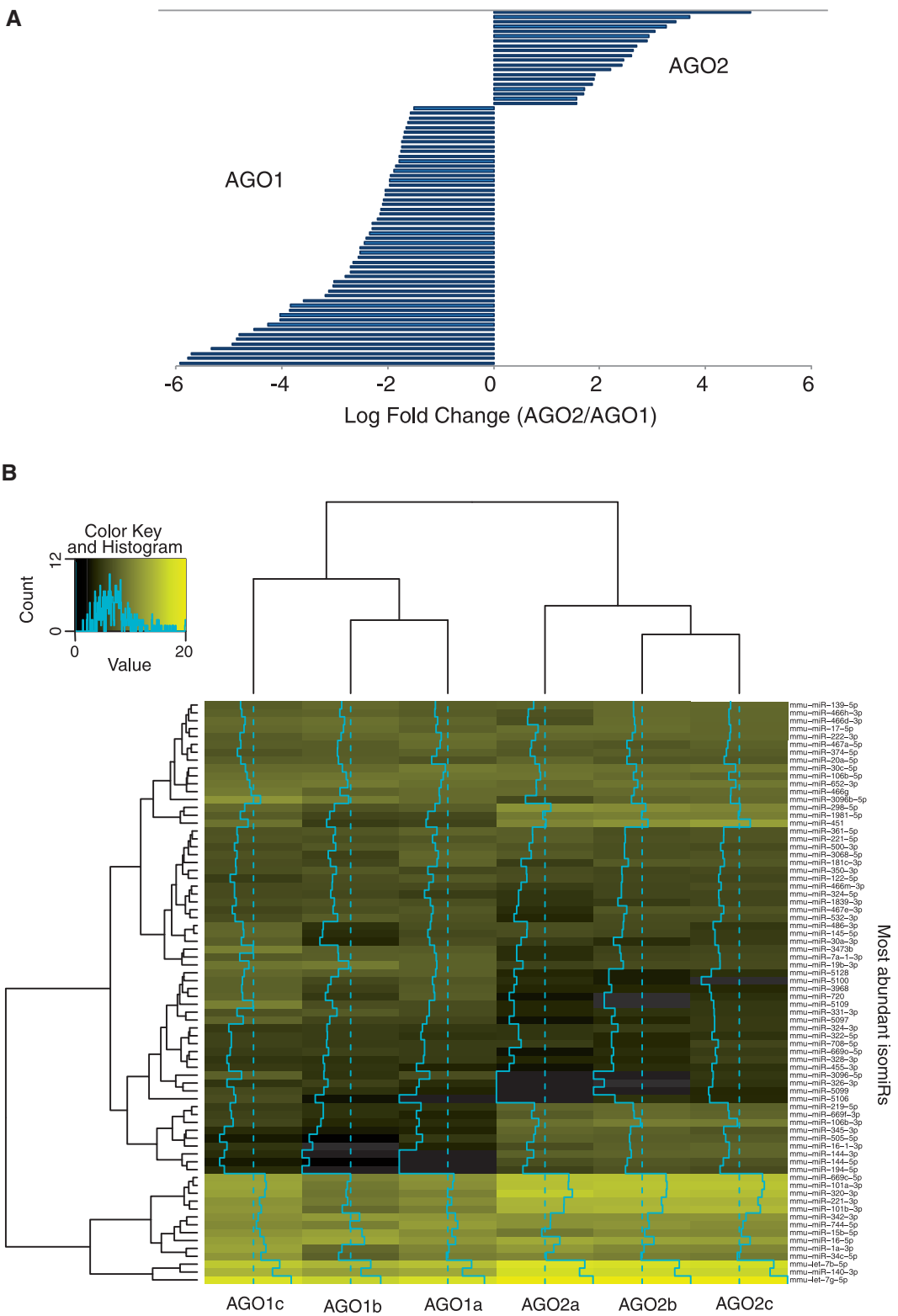

Figure 4. Differential expression of the most abundant isomiRs. (**A**) Log-fold change in expression comparing AGO2- and AGO1-associated isomiRs. A positive fold change indicates that the isomiR is binding to the AGO2; negative to AGO1. Only reads with at least 1.5-fold difference and adjusted *P*-values <0.05 were considered statistically significant. (**B**) Heatmap and hierarchial clustering analysis of the same isomiRs. Cyan lines depict the deviation of the count value for each isomiR from the mean count. The color key index plots the normalized value against the isomiR counts having the same value.

Table 1. GO enrichment of the targets of the same isomiRs

Targets of AGO1 bound miRNAs			
Annotation cluster 1		Enrichment score: 26.37	
Category	Term	Count	*P*-value
SP_PIR_KEYWORDS	Transcription regulation	390	2.8E-39
Biological process	Regulation of transcription	491	1.5E-35
Biological process	DNA binding	400	1.1E-29
SP_PIR_KEYWORDS	Nucleus	726	3.4E-29
Molecular function	Transcription regulator activity	295	1.2E-27
SP_PIR_KEYWORDS	Transcription	394	2.5E-27
Biological process	Transcription	390	3.1E-27
SP_PIR_KEYWORDS	DNA binding	327	5.8E-26
Biological process	Regulation of RNA metabolic process	331	1.2E-23
Biological process	Regulation of transcription, DNA-dependent	324	1.1E-22
Molecular function	Transcription factor activity	201	9.2E-22
Molecular function	Sequence-specific DNA binding	143	3.9E-15
Targets of AGO2 bound miRNAs			
Annotation cluster 1		Enrichment score: 13.44	
Category	Term	Count	*P*-value
SP_PIR_KEYWORDS	Transcription regulation	228	1.0E-21
Biological process	Regulation of transcription	294	1.4E-19
Molecular function	DNA binding	243	1.0E-17
SP_PIR_KEYWORDS	Transcription	232	1.5E-15
SP_PIR_KEYWORDS	Nucleus	421	3.0E-15
SP_PIR_KEYWORDS	DNA binding	192	1.3E-14
Biological process	Transcription	232	1.9E-14
Molecular function	Transcription regulator activity	171	1.5E-13
Molecular function	Transcription factor activity	118	2.2E-11
Biological process	Regulation of transcription from RNA polymerase II promoter	100	2.3E-11
Biological process	Regulation of RNA metabolic process	190	8.5E-11
Biological process	Regulation of Transcription, DNA-dependent	186	2.3E-10
Molecular function	Sequence specific DNA binding	83	7.3E-8

'Functional annotation tools' feature of DAVID bioinformatics program was used to enrich the targets. The top clusters with highest enrichment score in both groups are shown. The fourth column depicts the count percentage of the total number of targets in each category. SP_PIR_KEYWORDS: keywords of the Swiss-Prot Protein Information Resource.

Table 2. Genomic loci of mitochondrial *cis*-NATs

Chromosome locus	Possible source of sense	Possible source of antisense	Figure
chrM:3772-3842	AK138995	Mt-Tq-201 (tRNA)	Supplementary Figure S3A
chrM:3914-4951	*NADH Dehydrogenase 2*	AK501316	Supplementary Figure S3B
chrM:9459–9806	*NADH Dehydrogenase 3*	AK141612	Supplementary Figure S3C
chrM:10 167–11 544	*NADH Dehydrogenase 4*	AA467310 or AA433514 or CB597336	Supplementary Figure S3D
chrM:11 742–13 565	*NADH Dehydrogenase 5*	AK138272 or CB597336	Supplementary Figure S3E
chrM:14 145–15 288	*Cytochrome b*	BF785953	Supplementary Figure S3F
chrM:15 356–15 422	AK154943 or AK131592	Mt-Tp-201 (tRNA)	Figure 5A
chrM:13 552–14 070	AK140457 or BC055066 or BC081549 or AK142712 or AK154943 or AK159184	*NADH Dehydrogenase 6*	Figure 5B

Sense and antisense pairs located in eight mitochondrial genomic loci and the coordinates and possible source of the reads are shown.

Tissue based differential expression and function of primary-isomiRs was first suggested by Morin *et al.* (25). Wyman *et al.,* further reported that 3′ modification of miRNAs is an evolutionarily conserved process that is functionally relevant in diverse tissues and disease states (22). They also suggested that the post-transcriptional addition of adenine to generate 3′NTAs (non-templated additions), here termed 3′ substitutions, is a physiologically regulated process and several members of the nucleotidyl transferase family, including PAPD4, PAPD5, TUT1, MTPAP, etc., might regulate this process (22). In this study, the majority of the 3′ modifications were extensions, i.e. the modified isomiR aligns to the genome perfectly, yet still differ from the canonical sequence. Such modifications

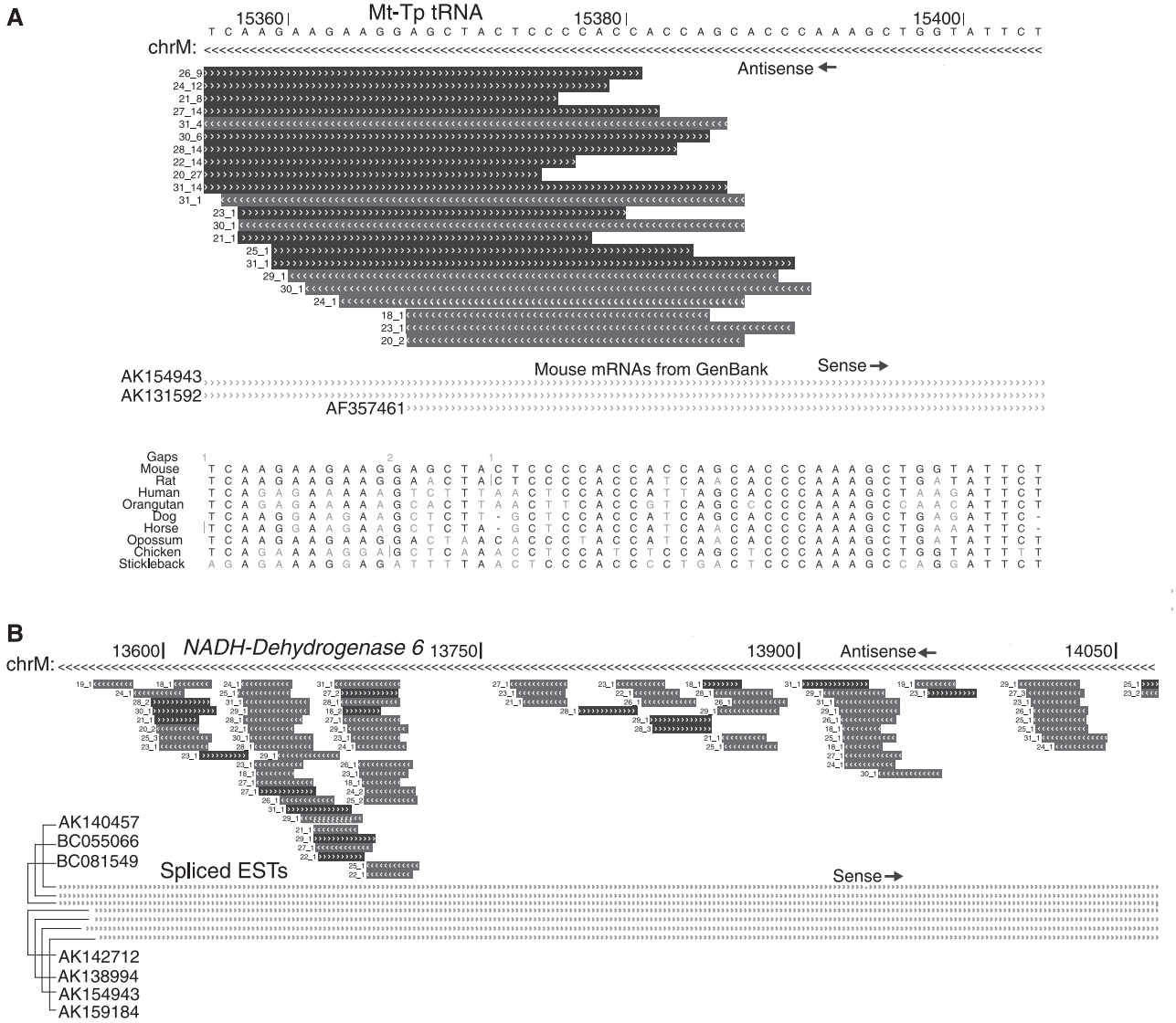

Figure 5. Representative AGO-associated *cis*-NATs. (**A**) Representative reads mapping sense and antisense to the same genomic locus encoding the Mt-Tp tRNA gene. Reads shown in dark shading align sense with respect to the genomic locus; light shading, antisense. Multiple alignments with various vertebrate species are also shown. The annotation of the reads follows length_copy number format. (**B**) Representative reads mapping sense and antisense to the mitochondrial 'NADH-Dehydrogenase 6' locus.

can be caused by either differential processing by DICER during miRNA synthesis or nucleotidyl transferases post-maturation or both (26). We also identified nine cases of 3′ substitutions, in which the sequences align imperfectly with the genome. These modifications are most likely caused by the post-maturational 3′ addition of nucleotides by nucleotidyl transferases (26). Regarding their function, isomiRs are differentially expressed among distinct human cancers including gastric cancers and metastatic prostate cancer, murine cardiomyocytes and in pre-eclampsia and psoriasis (27–31). Furthermore, Cloonan *et al.*, reported that isomiRs function cooperatively with their corresponding canonical miRNAs during target regulation in various human tissues, confirming their functional importance (32). Our studies indicate that a large number of miRNAs from CD4 T cells are expressed primarily as non-canonical isomiRs. This observation may be particularly relevant because of the need of these T cells to differentiate into specialized effector subtypes (Th1, Th2, Th17, etc.) that are required to defend against a wide variety of pathogens (33).

The differential binding affinity of isomiRs to AGO proteins was first shown for miR-142-5p in human Jurkat T cells (34). Burroughs *et al.* also showed differential binding of isomiRs to AGO proteins in human THP-1 cell lines for the most abundant isomiRs having 3′ nucleotide variations and less 5′ variation (7). However, in this study we showed that isomiRs deriving from 127 distinct miRNAs associate differentially between AGO1 and AGO2, suggesting that differential AGO binding might be more prevalent in normal somatic cells than previously thought. Moreover, we have shown that isomiRs ending with 3′adenine preferentially bind to AGO1 and those ending with 3′ uracil, to AGO2 (Figure 3C). Such

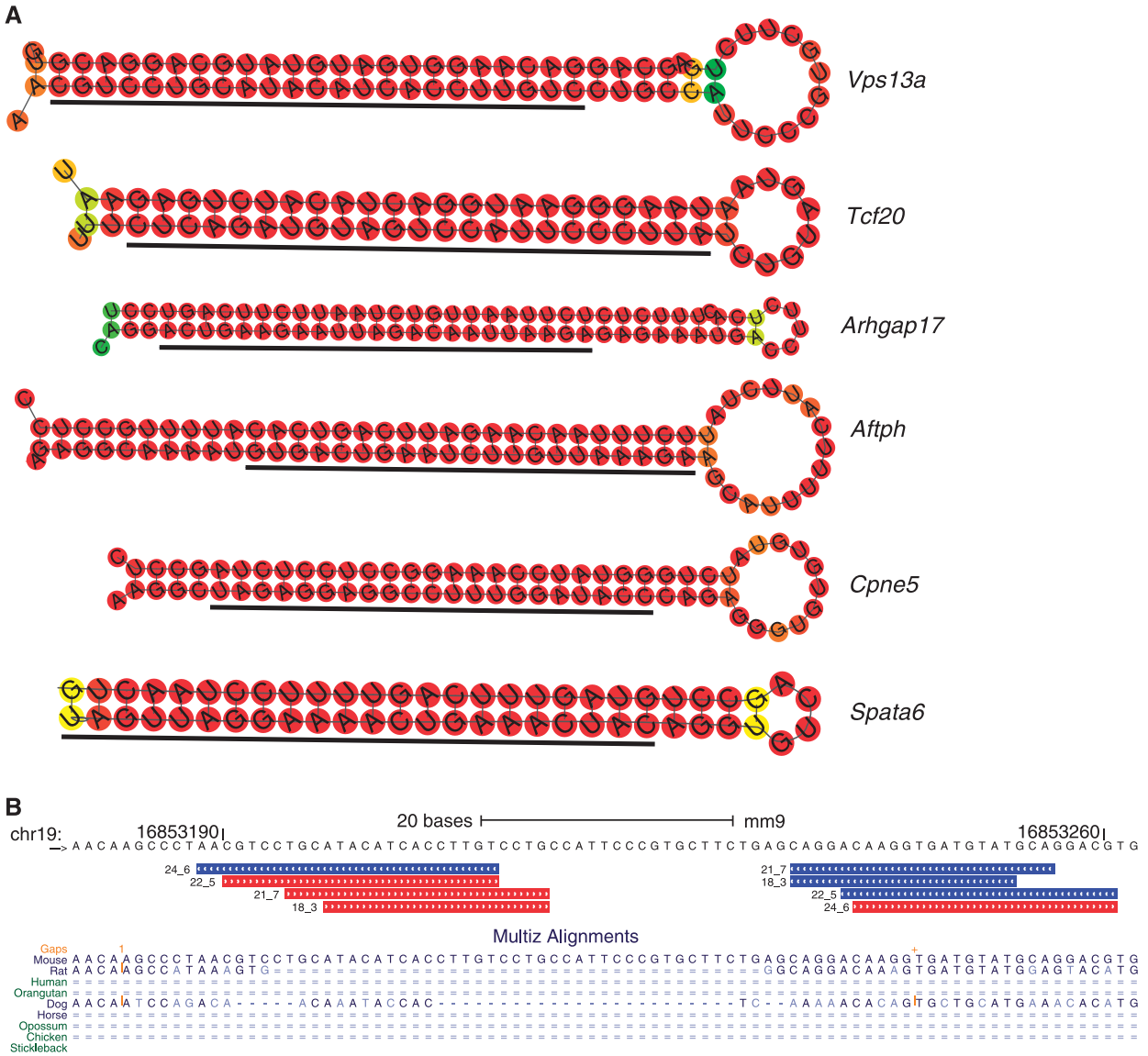

Figure 6. Hairpin precursors of candidate endogenous siRNAs and the representative genome alignment. (**A**) Folding diagrams of candidate endogenous siRNA precursors. Except for *Spata6*, the junction of stem and loop is constituted by A = U base pairing suitable for Dicer cleavage. The color key index depicts 0 as lowest base pair probability and 1 highest. Underscored regions indicate most likely mature sequences. (**B**) Alignment of the candidate endo-siRNAs derived from *Vps13a* introns. The same reads that map in sense orientation to the genome map antisense to a very proximal locus. Alignments for various species are also shown.

3′ nucleotide bias for AGO proteins has previously been observed in *C. elegans* (35,36), whereas this study reports such bias in mammalian cells for the first time. Although further study is required, terminal nucleotide modifications potentially explain the differential sorting of miRNA isomiRs between AGO proteins and further support the notion of distinct functions of AGO proteins. Our laboratory and others have further shown that post-transcriptional editing of miRNAs can significantly alter the target repertoire of miRNAs, again suggesting distinct functions for isomiRs (20,26). Thus, analysis of AGO-associated transcripts confirms that miRNA isomiRs are extensively edited post-transcriptionally and imply differential functions for distinct isomiRs deriving from the same canonical sequences and AGO proteins.

Little is currently known of the biochemical properties of mammalian AGO proteins. miRNAs of plants and lower organisms are known to sort differentially to AGO proteins, but this is currently thought be a primitive trait that is largely lost in higher eukaryotes (3). Our findings from T cells indicate, in contrast, that many miRNAs in fact differentially bind to AGO1 and AGO2. Deeper analysis of the putative targets of these distinct miRNA classes indicates that only approximately a third of these targets overlap despite having a predominant focus on the regulation of transcription factors. Together, these findings indicate that isomiRs work cooperatively with AGO1 and 2 to regulate extremely fundamental cellular processes, especially transcription, but imply the existence of a division of labor with regard to the regulation of distinct genes.

```
Locus: chr17:24,857,717-24,857,742

  Candidate_21  ACTCTCTCACTCTGCATGGTTACACA

         Mouse  aCTCTCTC-actctgcatggttacaca
           Rat  aCTCTCTC-actctgcatggttacact
         Human  -CTCTCTCggctctgcatagttgcact
     Orangutan  -CTCTCTCagctctgcatagttgcact
           Dog  aCTCTCTCagctctgcatggtcatgct
         Horse  aCTCTCTCagctctgcatggttacact
      Opossum  --------------acttagccgtcat
       Chicken  gCTgTaTCaggaatctaaaaccccatt


Precursor:Candidate_21 length: 97
Mature:Candidate_21 length: 26
mfe: -34.5 kcal/mol

position  52
target 5' C        UUC  A     A       A 3'
           GUGUGACU   GU CGGAG GAGAGAG
           CACAUUGG   UA GUCUC CUCUCUC
miRNA  3' A             C     A       A 5'


Candidate_21: Precursor
```

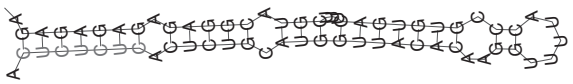

Figure 7. Features of a candidate novel miRNA**.** miRNA Candidate_21 genomic coordinates, mature sequence, seed region, species conservation of the seed region, MFE value after folding, multiple alignments from other species and the precursor RNA structure are shown. The gray highlighted region of the precursor RNA is the seed sequence.

In addition to miRNAs, our findings demonstrate the existence of other classes of antisense transcripts in T cells, including NATs and endogenous siRNAs. The existence and differential expression of antisense transcripts deriving from mitochondrially encoded genes has previously been demonstrated from porcine and murine brain (37,38). Similarly, Burzio *et al.,* showed the existence of long antisense transcripts deriving from mitochondrially encoded rRNA 16 in tumors and normal proliferating human cells (800 and 2000 bp, respectively). These mtncRNAs have inverted repeats and also showed differential expression between tumor and normal cells (39). NATs are critically important regulators of genomic imprinting, RNA interference, alternative splicing and X-chromosome inactivation (23), functions that are no less important in T cells. However, T-cell NATs all derive from unique loci including five distinct subunits of the NADH dehydrogenase complex, which is a crucial enzyme regulating cellular respiration. Several NADH dehydrogenase subunits have been critically linked to several neurological disorders (40–43). A critical extension of the current studies is to determine if the unique *cis*-NATs discovered here may regulate T cell and therefore immune function.

The lack of RNA dependent RNA polymerase (RdRP) that is essential for siRNA mediated gene regulation and the marked sensitivity of mammalian cells to the long dsRNA precursors of siRNAs that resemble viral genomes have suggested that endogenous siRNAs are not functional in higher eukaryotes (24). However, large numbers of hairpinRNA (hpRNA)-derived endo-siRNAs have been demonstrated in mouse oocytes (44). Interferon-based immune responses to long dsRNA are suppressed in oocytes, providing a means for the physiological expression of these transcripts in haploid cells. However, Smalheiser *et al.,* reported finding eight endogenous siRNA candidates in mouse hippocampus (24). Our studies confirm the existence of endogenous, AGO-binding siRNAs in mouse somatic T cells, suggesting that such transcripts are expressed physiologically, albeit at extremely low levels.

Much further work is required to establish the functional significance of distinct isomiRs, NATs and siRNAs, both generally in mammals and specifically in T cells. We note further that the complete lack of species conservation of siRNAs and NATs implies that these transcripts regulate morphology and function in a species-specific manner. Additional studies are further required to determine the redundant and unique functions of the other known mammalian AGO proteins. More complete sequencing of AGO-related samples from diverse species and tissues combined with specialized techniques for manipulating AGO proteins *in vitro* and *in vivo* will provide insight into these critical questions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–10 and Supplementary Figures 1–3.

FUNDING

Clayton Foundation for Research and the National Institutes of Health (NIH) [HL095382 to D.B.C.]. Funding for open access charge: NIH [HL095382 to D.B.C.].

Conflict of interest statement. None declared.

REFERENCES

1. Joshua-Tor,L. and Hannon,G.J. (2011) Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harb. Perspect. Biol.*, **3**, a003772.
2. Ender,C. and Meister,G. (2010) Argonaute proteins at a glance. *J. Cell Sci.*, **123**, 1819–1823.
3. Czech,B. and Hannon,G.J. (2011) Small RNA sorting: matchmaking for Argonautes. *Nat. Rev. Genet.*, **12**, 19–31.
4. Czech,B., Malone,C.D., Zhou,R., Stark,A., Schlingeheyde,C., Dus,M., Perrimon,N., Kellis,M., Wohlschlegel,J.A., Sachidanandam,R. *et al.* (2008) An endogenous small interfering RNA pathway in Drosophila. *Nature*, **453**, 798–802.
5. Cenik,E.S. and Zamore,P.D. (2011) Argonaute proteins. *Curr. Biol.*, **21**, 21.
6. Liu,J., Carmell,M.A., Rivas,F.V., Marsden,C.G., Thomson,J.M., Song,J.J., Hammond,S.M., Joshua-Tor,L. and Hannon,G.J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437–1441.
7. Burroughs,A.M., Ando,Y., de Hoon,M.J., Tomaru,Y., Suzuki,H., Hayashizaki,Y. and Daub,C.O. (2011) Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol.*, **8**, 158–177.
8. Rother,S. and Meister,G. (2011) Small RNAs derived from longer non-coding RNAs. *Biochimie*, **93**, 1905–1915.

9. Hackenberg,M., Sturm,M., Langenberger,D., Falcon-Perez,J.M. and Aransay,A.M. (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res.*, **37**, W68–W76.
10. Langmead,B. (2010) Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinformatics*, **11**, 1–24.
11. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
12. Hofacker,I.L. (2009) RNA secondary structure analysis using the Vienna RNA package. *Curr. Protoc. Bioinform.*, **12**, 1–16.
13. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
14. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
15. Dennis,G. Jr, Sherman,B.T., Hosack,D.A., Yang,J., Gao,W., Lane,H.C. and Lempicki,R.A. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.*, **4**, 3.
16. Hackenberg,M., Rodriguez-Ezpeleta,N. and Aransay,A.M. (2011) miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic Acids Res.*, **39**, W132–W138.
17. Janowski,B.A., Huffman,K.E., Schwartz,J.C., Ram,R., Nordsell,R., Shames,D.S., Minna,J.D. and Corey,D.R. (2006) Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat. Struct. Mol. Biol.*, **13**, 787–792.
18. Leung,A.K., Young,A.G., Bhutkar,A., Zheng,G.X., Bosson,A.D., Nielsen,C.B. and Sharp,P.A. (2011) Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. [Erratum appears in Nat. Struct. Mol. Biol. 2011 Sep;18(9):1084]. *Nat. Struct. Mol. Biol.*, **18**, 237–244.
19. Wu,C., So,J., Davis-Dusenbery,B.N., Qi,H.H., Bloch,D.B., Shi,Y., Lagna,G. and Hata,A. (2011) Hypoxia potentiates microRNA-mediated gene silencing through posttranslational modification of Argonaute2. *Mol. Cell. Biol.*, **31**, 4760–4774.
20. Polikepahad,S., Knight,J.M., Naghavi,A.O., Oplt,T., Creighton,C.J., Shaw,C., Benham,A.L., Kim,J., Soibam,B., Harris,R.A. *et al.* (2010) Proinflammatory role for let-7 microRNAS in experimental asthma. *J. Biol. Chem.*, **285**, 30139–30149.
21. Morin,R.D., O'Connor,M.D., Griffith,M., Kuchenbauer,F., Delaney,A., Prabhu,A.-L., Zhao,Y., McDonald,H., Zeng,T., Hirst,M. *et al.* (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.*, **18**, 610–621.
22. Wyman,S.K., Knouf,E.C., Parkin,R.K., Fritz,B.R., Lin,D.W., Dennis,L.M., Krouse,M.A., Webster,P.J. and Tewari,M. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res.*, **21**, 1450–1461.
23. Zhang,Y., Liu,X.S., Liu,Q.R. and Wei,L. (2006) Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res.*, **34**, 3465–3475.
24. Smalheiser,N.R., Lugli,G., Thimmapuram,J., Cook,E.H. and Larson,J. (2011) Endogenous siRNAs and noncoding RNA-derived small RNAs are expressed in adult mouse hippocampus and are up-regulated in olfactory discrimination training. *RNA*, **17**, 166–181.
25. Morin,R.D., O'Connor,M.D., Griffith,M., Kuchenbauer,F., Delaney,A., Prabhu,A.L., Zhao,Y., McDonald,H., Zeng,T., Hirst,M. *et al.* (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. [Erratum appears in Genome Res. 2009 May;19(5):958]. *Genome Res.*, **18**, 610–621.
26. Burroughs,A.M., Ando,Y., de Hoon,M.J., Tomaru,Y., Nishibu,T., Ukekawa,R., Funakoshi,T., Kurokawa,T., Suzuki,H., Hayashizaki,Y. *et al.* (2010) A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness. *Genome Res.*, **20**, 1398–1410.
27. Guo,L., Li,H., Liang,T., Lu,J., Yang,Q., Ge,Q. and Lu,Z. (2012) Consistent isomiR expression patterns and 3' addition events in miRNA gene clusters and families implicate functional and evolutionary relationships. *Mol. Biol. Rep.*, **39**, 6699–6706.
28. Humphreys,D.T., Hynes,C.J., Patel,H.R., Wei,G.H., Cannon,L., Fatkin,D., Suter,C.M., Clancy,J.L. and Preiss,T. (2012) Complexity of murine cardiomyocyte miRNA biogenesis, sequence variant expression and function. *PLoS One*, **7**, e30933.
29. Joyce,C.E., Zhou,X., Xia,J., Ryan,C., Thrash,B., Menter,A., Zhang,W. and Bowcock,A.M. (2011) Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome. *Hum. Mol. Genet.*, **20**, 4025–4040.
30. Li,S.C., Liao,Y.L., Ho,M.R., Tsai,K.W., Lai,C.H. and Lin,W.C. (2012) miRNA arm selection and isomiR distribution in gastric cancer. *BMC Genomics*, **13(Suppl 1)**, S13.
31. Watahiki,A., Wang,Y., Morris,J., Dennis,K., O'Dwyer,H.M., Gleave,M. and Gout,P.W. (2011) MicroRNAs associated with metastatic prostate cancer. *PLoS One*, **6**, e24950.
32. Cloonan,N., Wani,S., Xu,Q., Gu,J., Lea,K., Heater,S., Barbacioru,C., Steptoe,A.L., Martin,H.C., Nourbakhsh,E. *et al.* (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.*, **12**, R126.
33. Dong,C. (2006) Diversification of T-helper-cell lineages: finding the family root of IL-17-producing cells. *Nat. Rev. Immunol.*, **6**, 329–333.
34. Siomi,M.C. and Siomi,H. (2004) Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Nucleic Acids Symp. Ser.*, **52**, 59–60.
35. Frank,F., Sonenberg,N. and Nagar,B. (2010) Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, **465**, 818–822.
36. Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
37. Michel,U., Stringaris,A.K., Nau,R. and Rieckmann,P. (2000) Differential expression of sense and antisense transcripts of the mitochondrial DNA region coding for ATPase 6 in fetal and adult porcine brain: identification of novel unusually assembled mitochondrial RNAs. *Biochem. Biophys. Res. Commun.*, **271**, 170–180.
38. Smalheiser,N.R., Lugli,G., Thimmapuram,J., Cook,E.H. and Larson,J. (2011) Mitochondrial small RNAs that are up-regulated in hippocampus during olfactory discrimination training in mice. *Mitochondrion*, **11**, 994–995.
39. Burzio,V.A., Villota,C., Villegas,J., Landerer,E., Boccardo,E., Villa,L.L., Martinez,R., Lopez,C., Gaete,F., Toro,V. *et al.* (2009) Expression of a family of noncoding mitochondrial RNAs distinguishes normal from cancer cells. *Proc. Natl Acad. Sci. USA*, **106**, 9430–9434.
40. Crimi,M., Papadimitriou,A., Galbiati,S., Palamidou,P., Fortunato,F., Bordoni,A., Papandreou,U., Papadimitriou,D., Hadjigeorgiou,G.M., Drogari,E. *et al.* (2004) A new mitochondrial DNA mutation in ND3 gene causing severe Leigh syndrome with early lethality. *Pediatr. Res.*, **55**, 842–846.
41. Johns,D.R., Neufeld,M.J. and Park,R.D. (1992) An ND-6 mitochondrial DNA mutation associated with Leber hereditary optic neuropathy. *Biochem. Biophys. Res. Commun.*, **187**, 1551–1557.
42. Lertrit,P., Noer,A.S., Jean-Francois,M.J., Kapsa,R., Dennett,X., Thyagarajan,D., Lethlean,K., Byrne,E. and Marzuki,S. (1992) A new disease-related mutation for mitochondrial encephalopathy lactic acidosis and strokelike episodes (MELAS) syndrome affects the ND4 subunit of the respiratory complex I. *Am. J. Hum. Genet.*, **51**, 457–468.
43. Lin,F.H., Lin,R., Wisniewski,H.M., Hwang,Y.W., Grundke-Iqbal,I., Healy-Louie,G. and Iqbal,K. (1992) Detection of point mutations in codon 331 of mitochondrial NADH dehydrogenase subunit 2 in Alzheimer's brains. *Biochem. Biophys. Res. Commun.*, **182**, 238–246.
44. Okamura,K., Chung,W.J. and Lai,E.C. (2008) The long and short of inverted repeat genes in animals: microRNAs, mirtrons and hairpin RNAs. *Cell Cycle*, **7**, 2840–2845.