Supplementary figures, table and method:

An alternative mode of microRNA target recognition

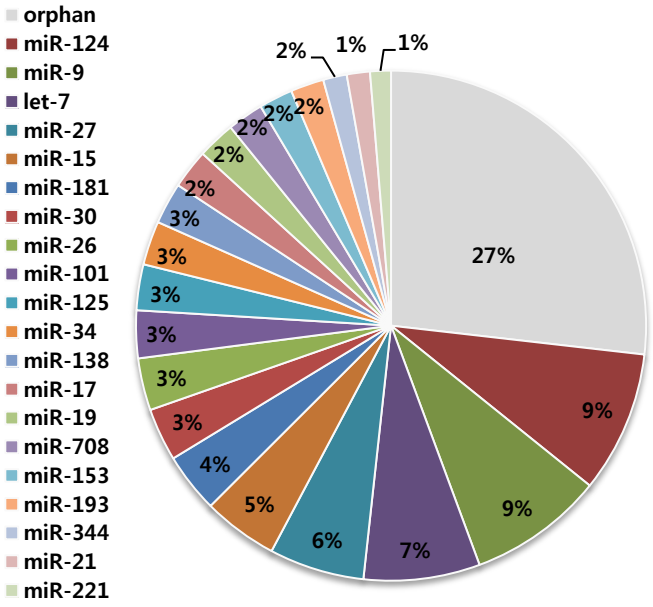
Sung Wook Chi[1-4], Gregory J. Hannon[2] and Robert B. Darnell[1]

[1]Laboratory of Neuro-Oncology, The Rockefeller University, Howard Hughes Medical Institute, New York, New York 10065, USA. [2]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor, New York 11724, USA. [3]Graduate School, Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, 50 Irwon-dong, Kangnam-ku, Seoul 135-710, Korea. [4]Samsung Research Institute for Future Medicine, Samsung Medical Center, 50 Irwon-dong, Kangnam-ku, Seoul 135-710, Korea.

Correspondence should be addressed to S.W.C (swchi@skku.edu) and R.B.D (darnelr@rockefeller.edu).

Supplementary Figure1

a



b

Rank	Number of sites	MEME Expectation	Sequence Logo
1	114	$3.40X10^{-45}$	
2	147	$9.00X10^{-18}$	
3	140	$2.60X10^{-13}$	
4	54	$6.60X10^{-12}$	
5	319	$1.90X10^{-11}$	
6	34	$9.50X10^{-03}$	

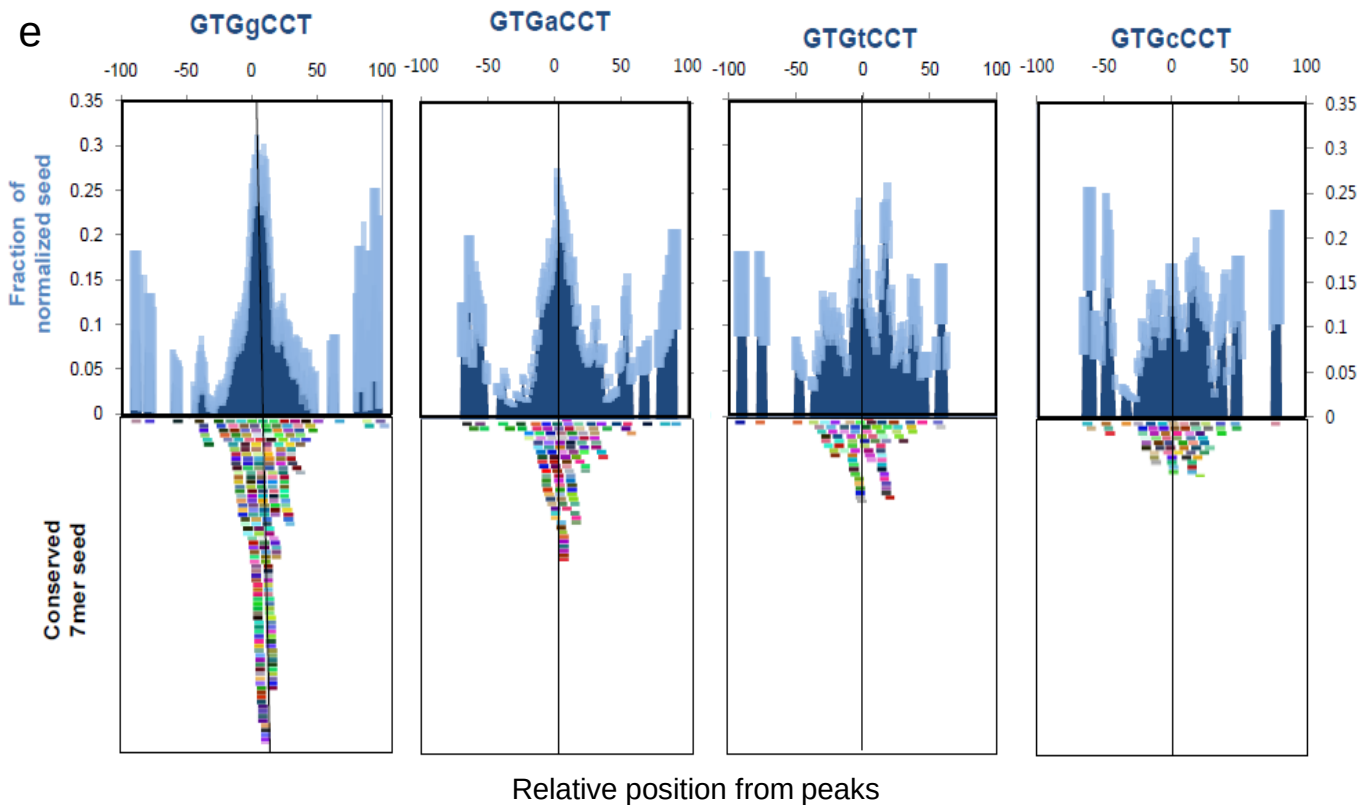
c

rank	motif	regression coefficient	number of clusters	p-value
1	gtggcct	0.708427	75	5.44E-10
2	acgtgag	2.75543	5	8.09E-10
3	tgataggc	2.68713	5	1.32E-09
4	gacctttt	2.01372	5	4.96E-06

d

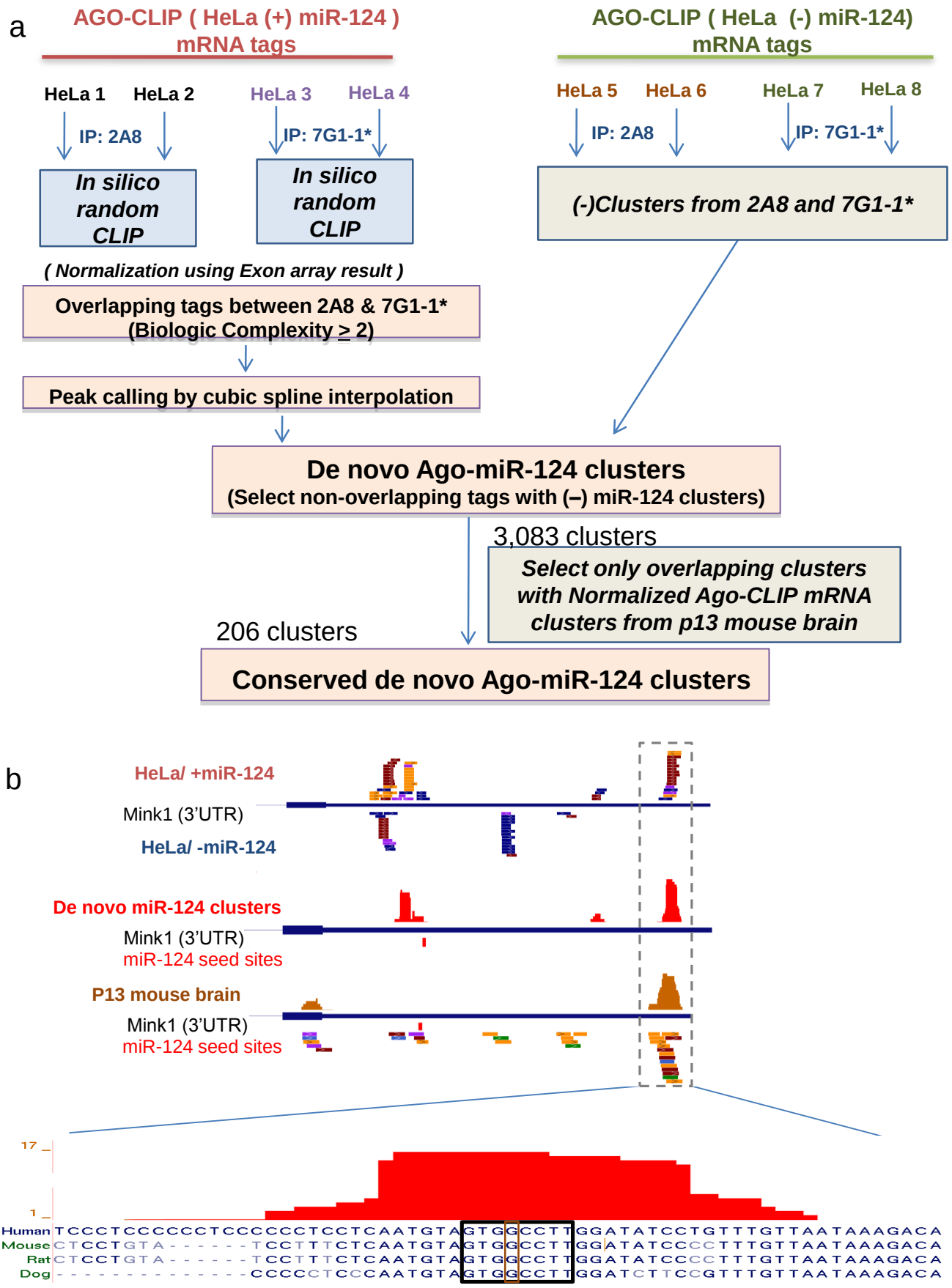
Rank	Conserved motif	obs	exp	obs/exp	p-value
1	TGGCCT	56	19	3.0082207	4.53E-18
2	GTGGCC	48	19	2.5160473	3.55E-11
3	TGTAAA	46	16	2.8122198	2.31E-13
4	TGTACA	44	14	3.0559491	6.13E-15
5	ACAAAA	42	14	2.9067529	4.23E-13
6	GGCCTT	41	12	3.3267258	3.13E-16
7	TGTATA	40	9	4.587933	3.17E-26
8	AAACAA	37	13	2.8036169	5.67E-11
9	AGCTTT	37	12	3.1658176	1.32E-13
10	AAAGTG	35	13	2.7129694	7.62E-10

Supplemental Figure 1

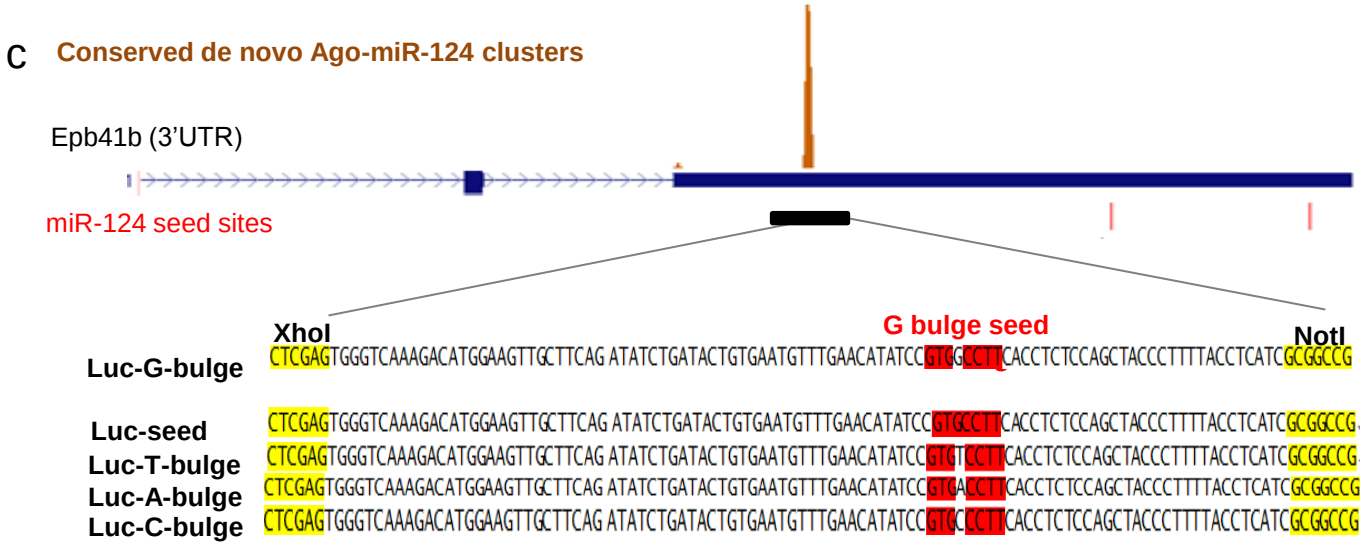


Supplementary Figure 1. Analysis of orphan clusters from Ago CLIP in mouse brain. a. Composition of Ago-mRNA clusters mapped as each of the top 20 Ago-miRNA targets (collapsed by family based on 6-mer seed(2-8)), based on seed matches as described previously[1]. 27% of total Ago mRNA clusters from p13 mouse brain (3073 clusters) are orphan clusters, of which Ago footprint regions (46 nt, BC≥2) have no predicted seed matches among the top 20 Ago-miRNAs families. b. Significantly enriched motifs (E-value < 0.01) are identified in Ago footprint region (62 nt, overlapping region of ≥ 95% clusters) and ranked based on E-value, estimate of the number of expected motifs by MEME analysis. c. We correlated all possible 6 to 8-mer motifs present in the footprint region of orphan Ago-mRNA clusters (64 nt, BC ≥ 2) with cluster height ($\log_2$(tag number in peak height) per cluster), using the MatrixREDUCE linear regression model. The enriched motifs (P<0.01) selected by default parameters (except –max_gap = 0) are shown. G-bulge site to miR-124 is highlighted with red. d. We also searched all possible 6-mer motifs present in the same orphan Ago-mRNA clusters by comparing observed (obs) versus expected (exp; from P13 brain transcripts) frequencies in Ago-mRNA clusters (BC ≥ 2). The enriched motifs (obs/exp > 2.5, *P* < 0.01; chi-square test) are shown in the table and the G-bulge site to miR-124 is highlighted with red. e. The position of conserved 7-mer bulge matches to miR-124 (with different sequence in bulge position; indicated as lower cases, bottom panel; each is represented by a different color) were plotted relative to peak position of all Ago-mRNA clusters (BC ≥ 2) in mouse brain. Top panel; distribution of mir-124 seed matches (plotted relative to cluster peak, normalized to number of clusters; blue graph); pale color indicates standard deviation. The analysis was done as described previously[1].

Supplemental Figure 2

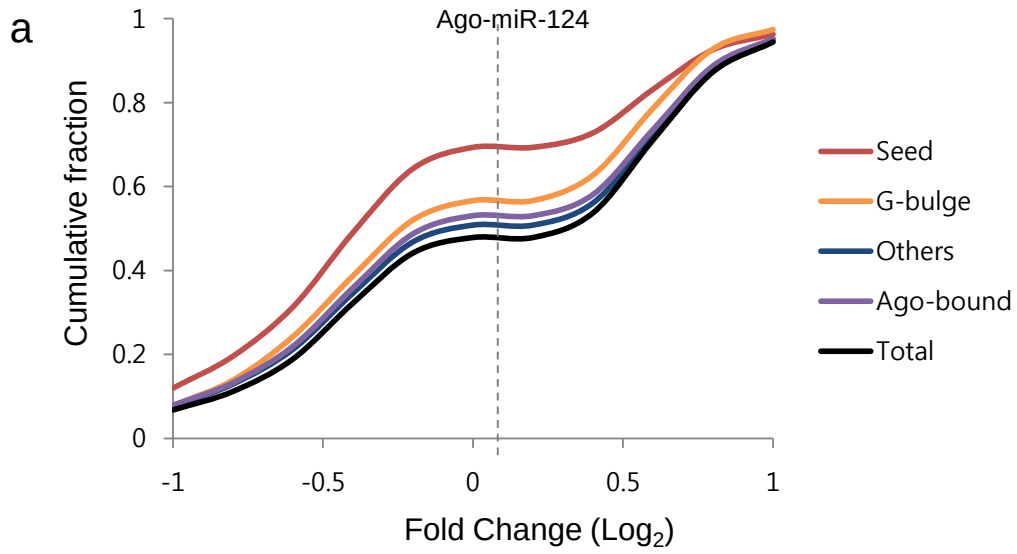


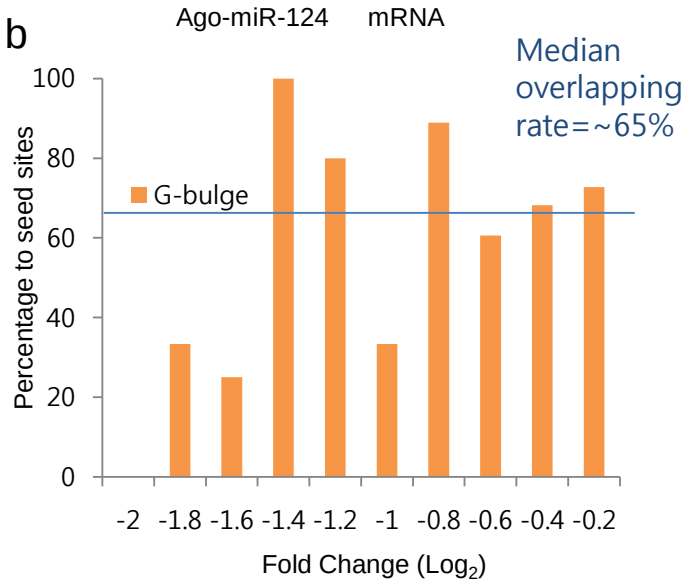
Supplemental Figure 2

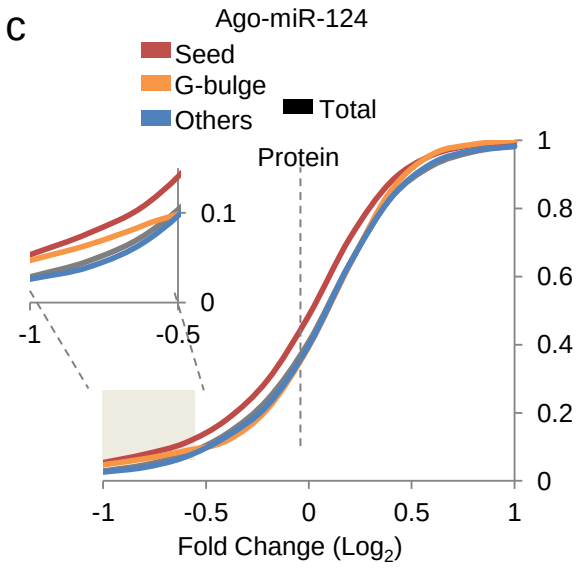


Supplementary Figure 2. Analysis of miR-124 G-bulge sites. a. Normalization of Ago HITS-CLIP tags. Biologic complexity (BC) refers to overlapping tags between experiments. BC is determined by taking data from each Brain HITS-CLIP, normalizing by *in silico* random CLIP for mRNA abundance, and comparing data from at least one of each antibody immunoprecipitation (2A8 and 7G1-1) as described previously[1]. For example, for a cluster to have a biologic complexity of 2, it must have at least one tag from 2A8 and 7G1-1 Ago HITS-CLIP experiments. For the comparison between the tags from mouse brain and from miR-124 transfected HeLa, we only selected tags that were overlapping (conserved de novo miR-124 clusters), to avoid confounding data that might result from the overexpression of miR-124 and the heterogeneity derived from all expressed miRNAs in brain. b. Conserved de novo Ago-miR-124 clusters in 3'UTR of *Mink1* validated by luciferase reporter assays. All tags from Ago HITS-CLIP in p13 mouse brain and miR-124 transfected HeLa are indicated. The locations of miR-124 seed matches are also indicated as red lines below the transcript (indicated as blue line). Dotted box indicated a region containing an orphan cluster (a de novo, miR-124 dependent cluster, but without a canonical miR-124 seed site) that contains a G bulge match; this region was subsequently inserted into the luciferase reporter. Bottom panel shows the sequence conservation of the G-bulge region. For use in luciferase assays (Fig. 2c), to minimize context effects from neighboring sequences only the indicated sequences covering the target regions annealing to miRNA (~100 nt) were used. c. Conserved *de novo* Ago-miR-124 clusters in 3'UTR of erythrocyte membrane protein band 4.1 (*Epb41*, upper panel) as described in A. Black bar indicated the region (104 nt) containing a G bulge match, inserted into the luciferase reporter. Bottom panel shows wild type and mutant sequences used for the luciferase reporter assay.

Supplemental Figure 3







d

	de novo miR-124 G-bulge sites in HeLa (human) expressed in both HeLa and mouse brain	
mouse brain Ago-cluster	**with mouse Ago-cluster**	**without mouse Ago-cluster**
Total number of sites	**8**	**34**
no change (G-bulge)	7	15
bulge position deletion (seed match)	1	0
multiple nucleotide difference	0	10
single nucleotide difference	0	9
position 3	0	1
position 4	0	2
bulge position (5-6)	0	3
position 6	0	1
position 7	0	2

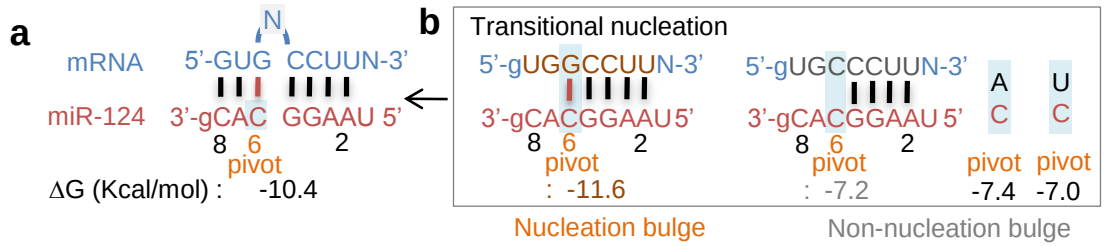
Supplemental Figure 3

e

	with mouse Ago-cluster	without Ago-cluster
no change in G-bulge site	7	15
defective nucleotide change in G-bulge sites	0	19

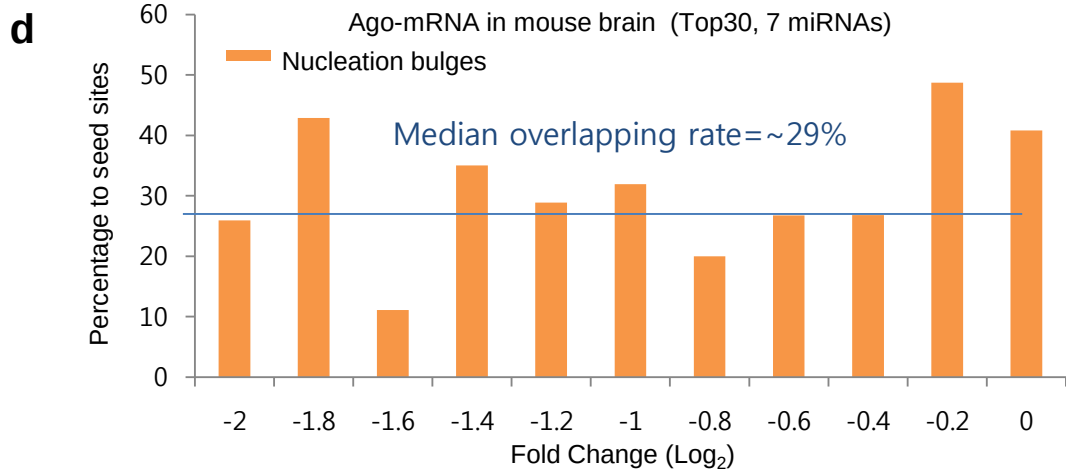
Supplementary figure 3. Validation of miR-124 G-bulge sites. a. Transcripts with G-bulge sites in de novo Ago miR-124 clusters (orange line) showed miR-124 dependent suppression relative to Ago-bound transcripts showing miR-124 dependent expression (purple line) in miR-124 transfected HeLa cells ($P < 0.05$, $n = 2694$). Transcripts harboring de novo Ago miR-124 clusters with seeds (red line) or G-bulge (orange line) showed further miR-124 suppression relative to Ago-bound transcripts (purple line). b. The amount of miR-124 dependant transcripts (identified by microarray experiments in miR-124 transfected HeLa)[2] harboring G-bulge sites in *de novo* Ago miR-124 clusters are calculated as percentage of seed sites in all ranges of repression and plotted. Blue line indicates the median overlap rate (~65%; median ratio of number of given bulge containing targets to the number of seed containing targets throughout all ranges of repression). c. Analysis of cumulative distribution of miR-124 dependant change in proteins levels after miR-124 transfection (identified by SILAC, $n = 860$)[3]. Similar results from miR-124 dependant transcripts were seen when analyzing miR-124 dependent protein suppression determined by the presence of G-bulge sites, but were only evident in smaller numbers of transcripts showing larger changes ($\log_2$ fold change < -0.5; inset). Of note, the published SILAC data is limited to nuclear-localized proteins (only 860 proteins after selecting only brain expressed transcripts) and two technical replicates with some variability ($r^2 = 0.72$) as described[3]. d. Species specific nucleotide changes between two Ago HITS-CLIP datasets (HeLa vs. mouse brain) showed that nucleotide changes in G-bulge sequence could abrogate interaction with Ago *in vivo*. In detail, 239 de novo miR-124 clusters contain G-bulge sites in HeLa cells. Among them, 42 G-bulge sites are located in transcripts also expressed in mouse brain. These 42 sites are valuable since they are confirmed to interact with Ago-miR-124 through G-bulge formation in HeLa (human) and also can be tested for their interaction with Ago-miR-124 in mouse brain. Among 42 sites, 8 sites were also identified to interact with Ago but 34 sites contain no Ago-mRNA clusters in mouse brain. When we looked at mouse sequences aligned to those human G-bulge sites, among 8 sites containing mouse Ago-clusters, 7 sites were conserved (the same G-bulges) and only 1 site was not conserved but changed into a canonical seed site. Interestingly, among 34 G-bulge sites in HeLa (human) containing no mouse Ago-clusters, 19 sites were not conserved, showing single (9 sites) or multiple nucleotide changes (10 sites). Of note, loss of signal in Ago HITS-CLIP data can be interpreted as loss of binding since the sequencing depth of data sets used in this analysis is near saturation[1]. e. Fisher's exact test was performed on a 2 x 2 table derived from d, yielding a significant P-value (0.0098). Based on these observations, we concluded that nucleotide changes in G-bulge sequences abrogate interaction with Ago in vivo, and that positions 6 and 7 in target transcripts are indeed necessary for targeting in G-bulge sites.

Supplemental Figure 4



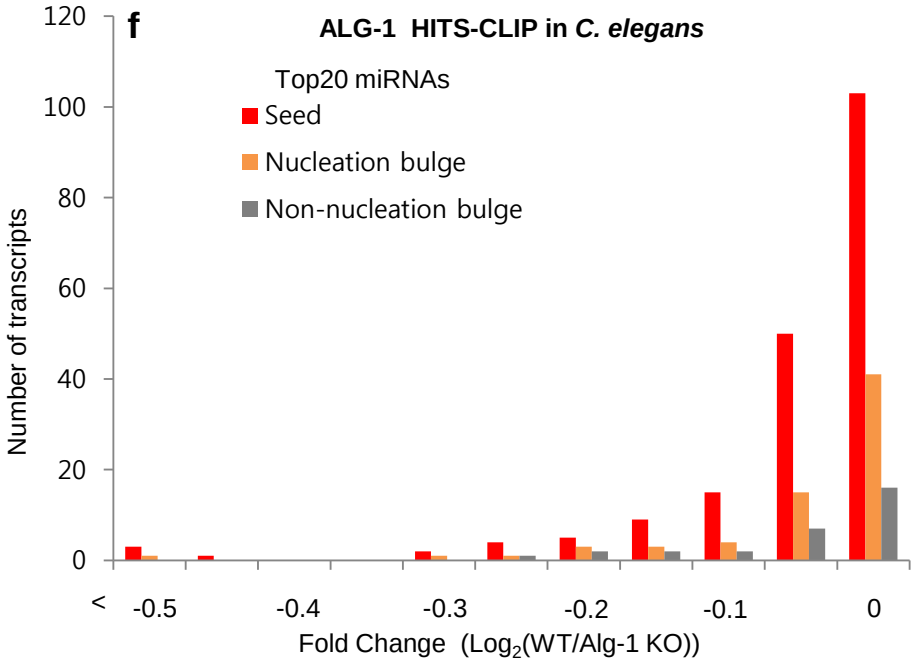
c

Top20 miRNAs (7mers)	Ago footprint regions in 3'UTR: conservation rate	number of conserved motifs	number of motifs	Ago footprint regions in CDS: Conservation rate	number of conserved motifs	number of motifs
Seed	59%	1006	1695	56%	740	1320
Nucleation bulge	32%	258	811	41%	274	674
Non-nucleation bulge	21%	70	341	31%	74	239
Total 7mers	23%	71866	307851	33%	98633	298726

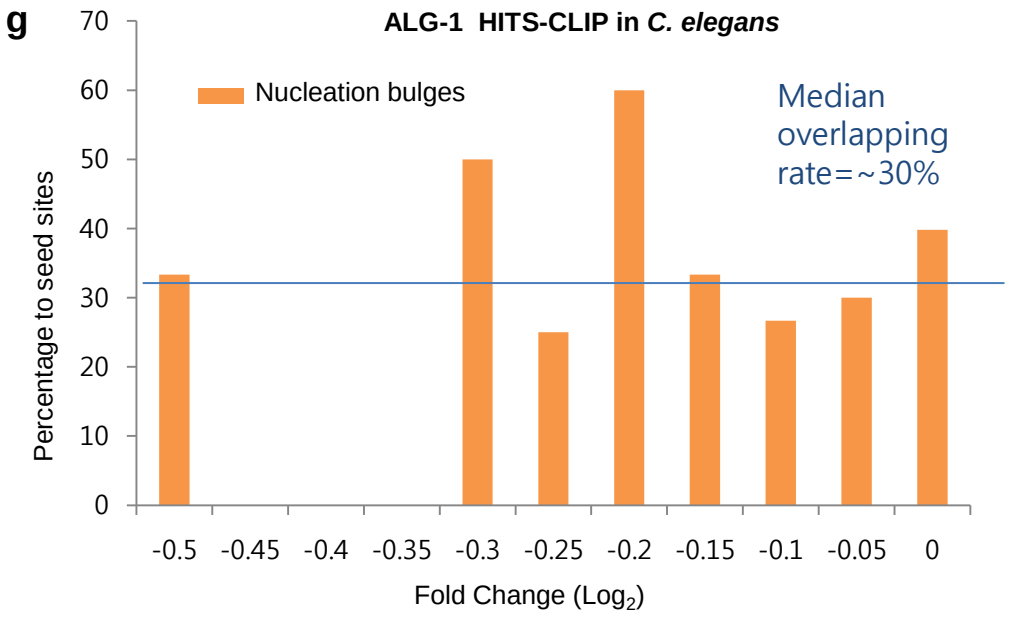


e

Rank	miR name	WT
1	miR-58	303168
2	miR-52	22101
3	miR-71	16512
4	miR-80	10536
5	miR-228	9949
6	miR-1	8799.5
7	miR-79	8381
8	miR-64	5673
9	miR-244	4573
10	miR-229	4108
11	miR-54	3908
12	let-7	3521
13	miR-48	3387
14	miR-65	3199
15	miR-72	3004
16	miR-2	2807
17	miR-86	2275
18	miR-42	2198
19	miR-236	1915
20	miR-239a	1869



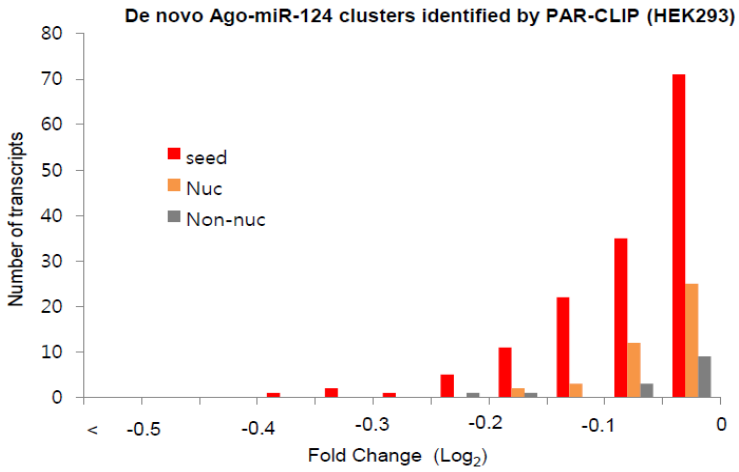
Supplemental Figure 4



Supplementary Figure 4. Identification of evolutionary conserved pivot pairing rules. a. Free energy calculation of miR-124 pairing to a G-bulge site (minimum free energy (ΔG) = -10.4 kcal mol^{-1}, calculated by RNAhybrid program). b. A pivot pairing rule and a transition nucleation model in miR-124. Free energy calculation shows that a G-bulge sequence (left panel) makes a transition nucleation state through 5 consecutive perfect pairing (position 2-6) to yield a very favorable free energy (ΔG = -11.6 kcal mol^{-1}), significantly lower than other bulge sites with inefficient transition states (right panel, only 4 consecutive perfect matches, C in position 6; ΔG = -7.2, A; -7.4, U; -7.0 kcal mol^{-1}). Pivot (position 6 in miRNA, orange), nucleotide sequence in bulge mRNA (blue shadow) and its pairing (orange line) are highlighted. c. Conservation rates of three different sites (seed, nucleation bulge and non-nucleation bulge) including background (total 7-mers) for the top 20 Ago-bound miRNAs were calculated in two different annotations (3'UTR and CDS) of Ago footprint regions (11,463, BC $\geq$ 2; including 3'UTR (~52%), coding sequence (~46%), and 5'UTR (~2%)) identified in brain. d. The amount of transcripts predicted to contain nucleation bulge sites of 7 miRNAs in Top 30 brain miRNAs in Ago-mRNA clusters (analyzed in Fig 3h and 3i) are calculated as percentage to seed sites in all ranges of repression and plotted. Blue line indicates the median overlap rate (~65%) as described in Supplementary Fig. 3b. e. Top 20 Ago-bound miRNAs in C. elegans. The 20 most miRNAs abundantly identified by performing ALG-1 HITS-CLIP[4] in wild type C. elegans are indicated. f. Numbers of the top 20 miRNAs dependant transcripts containing three different sites (seed, nucleation bulge and non-nucleation bulge sites) in wild type region specific clusters (http://yeolab.ucsd.edu/yeolab/Papers_files/NSMB_012010.tar.gz) are plotted in all ranges of repression. Fold Change indicates the $\log_2$ ratio of median probe intensity of transcripts in wild type versus Alg-1 null *C. elegans*, measured by microarray (GSE19138 in Gene Expression Omnibus database)[4]. g. The same analysis performed as in C for Ago-bound miRNAs in *C. elegans*.

Supplemental Figure 5

a



b

Ago-miR-124	Total		Fold change (Log_2) < 0	
	Number of transcripts	% to seed	Number of transcripts	% to seed
7mer-seed	457	100.0	148	100.0
Nuc	141	30.9	42	28.4
Non-Nuc	52	11.4	14	9.5

De novo Ago-miR-124 clusters identified by PAR-CLIP

c

Ago-miR-7	Total		Fold change (Log_2) < 0	
	Number of transcripts	% to seed	Number of transcripts	%to seed
7mer-seed	127	100.0	51	100.0
Nuc	33	26.0	10	19.6
Non-Nuc	1	0.8	0	0.0

De novo Ago-miR-7 clusters identified by PAR-CLIP

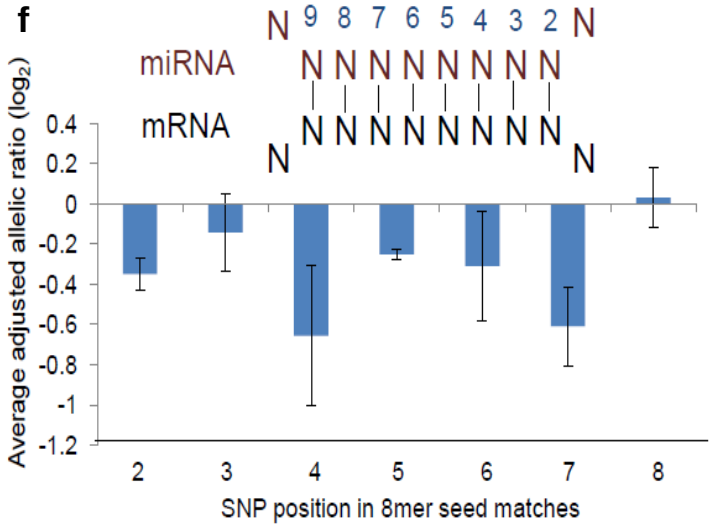
d

	Ago PAR-CLIP	Ago HITS-CLIP
De novo miR-124 clusters (without normalization)	13927	13309
6mer seed match (UGCCUU)	686	439
7mer seed match (GUGCCUU)	289	216
7mer G-bulge (UGGCCUU)	83	239
7mer A-bulge (UGACCUU)	48	34
7mer U-bulge (UGUCCUU)	32	28
7mer C-bulge (UGCCCUU)	30	25

e Normalized de novo miR-124 clusters

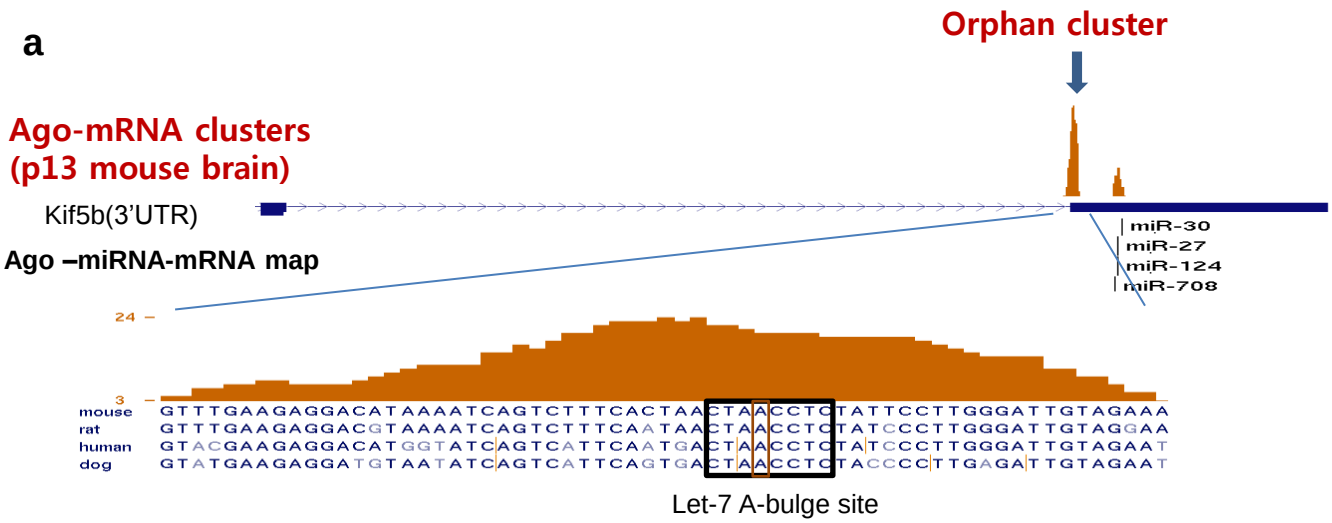
Motif	Type of motif	Number of sites
NNNGCCTT N	Total number of sites with 5mer (GCCTT)	833
GCCTT	5mer (2-6)	0
CGTGCCTTA	seed (1-9)	11
GTGCCTTA	seed (1-8)	58
CGTGCCTT	seed (2-9)	7
TGCCTTA	seed (1-7)	66
GTGCCTT	seed (2-8)	151
GCCTTA	seed (1-6)	89
TGCCTT	seed (2-7)	212
GTG**G**CCTTA	G-bulge (1-8)	39
GTG**G**CCTT	G-bulge (2-8)	125
TG**G**CCTTA	G-bulge (1-7)	15
TG**G**CCTT	G-bulge (2-7)	60

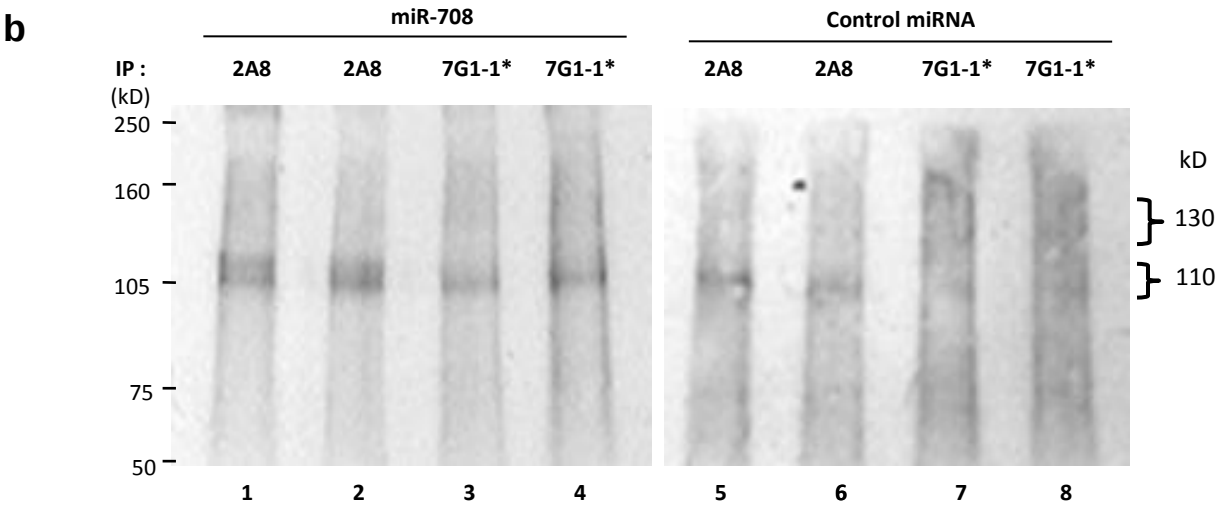
f



Supplementary Figure 5. Validation of pivot pairing rules. a-b. Number of transcripts harboring three different sites (seed, nucleation bulge and non-nucleation bulge sites) in de novo Ago-miR-124 clusters identified by PAR-CLIP[5]. Ago-miR-124 clusters (from Ago2 and miR-124 co-transfected HEK293 cell) defined and published by Hafner et al (mRNA-annotated clusters with at least 5 mRNA-annotated tags of at least 20 nucleotides in length and mapping uniquely to the genome with at most 1 error) are downloaded from the link (http://www.mirz.unibas.ch/restricted/clipdata/RESULTS/index.html) and the clusters overlapping with Ago2 clusters (from Ago2 transfected HEK293) were subtracted to get *de novo* Ago-miR-124 clusters. Level of each target transcripts defined by seed and pivot pairing rules in *de novo* Ago-miR-124 clusters was examined by $\log_2$ ratio (Ago2-miR-124 co-transfected vs. Ago2 transfected HEK293) of microarray data from GEO database (GSE14537) and the number of transcripts of which fold changes are <0 are plotted (a) and counted (b) depending on three different sites (seed, nucleation bulge and non nucleation bulge sites). c. The same analysis performed in (a) except the analysis for de novo Ago-miR-7 clusters (Ago2-miR-7 co-transfected vs. Ago2 transfected HEK293). d. Comparison of Ago HITS-CLIP and PAR-CLIP for the presence of seed and nucleation G-bulge sites in *de novo* Ago-miR-124 clusters. . We applied the same method on both published Ago PAR-CLIP mRNA clusters and Ago HITS-CLIP mRNA clusters without normalized by “in silico random CLIP” algorithm. Focusing on the data set from miR-124 transfected cell line (Ago PAR-CLIP; HEK293+miR-124, Ago HITS-CLIP; HeLa+miR-124), we simply selected de novo miR-124 clusters, which occurred only by miR-124 transfection (excluding Ago-mRNA clusters also observed in control transfected cell lines) and then examined the number of clusters containing canonical seeds or different types of bulge sites. Both methods showed specific enrichment of G-bulge sites against other types of bulges (A, U or C) but Ago HITS-CLIP showed more enrichment of G-bulge sites compared with other Ago PAR-CLIP (8.3 vs. 2.3 fold enrichment, relative to the average number of other bulge types). e. Since nucleation bulges share the same 5-mer seed (position 2-6), we examined whether 5-mer motifs alone (2-6) are sufficient for recognition by Ago-miR-124. All sites containing 5-mer motifs (position 2-6, GCCTT) in normalized de novo miR-124 clusters were examined and divided into three types (5-mer only: blue colored; seed type: red colored; G-bulge type: orange colored). There is no 5-mer in 833 de novo Ago-miR-124 footprints (0% false discovery rate, relative to average target sites per miRNA (~700) estimated previously[1-3], $P = 0.00$). Therefore, we could exclude the possibility that observation of nucleation bulge targets are due to 5-mer seed-mediated targeting. These observations strongly support the conclusion that positions 6-8 facilitate repression in the context of a G-bulge. Of note, in the previous study[1], we have observed evidence that nucleotide in miRNA position 1 could used for recognizing targets in vivo (all 6mer sequences match to miRNA seed position 1-8 were highly enriched in Ago footprint regions). Therefore, sequence matches to miR-124 position 1-6 should be interpreted as 6mer sites and not as 5mer sites. Additionally, this conclusion is supported by the observation that G-bulge sites interacting with miR-124 in human (HeLa) lose interaction with mouse Ago-miR-124 when human G-bulge sites are introduced by not only single nucleotide change in position 2-6 but also in position 6-7 (Supplementary Figure 4f). f. Measurement of mRNA allelic imbalances associated with heterozygous miRNA target sites (AI-Seq)[6] demonstrate that 5mer seed matches are unable to mediate miRNA dependent repression *in vivo*. By analyzing AI-Seq data as described previously[6], the average adjusted allelic ratios (log2) for 8mer seed sites (indicated in upper panel) depending on the position of 31 SNPs relative to pairing with cognate miRNAs (miR-1, miR-122 and miR-133; SNP rs33901248 in 8mer site was not included in this analysis since it had unusual high allelic ration of genomic DNA (9.61)) were calculated. Among them, the average adjusted allelic ratio (log2) for SNPs in position 7 (-0.61) also showed miRNA-dependent repression causing allelic imbalance as well as SNPs in other positions between 2-6, but 5mer seed matches (2-6) could not mediate miRNA-dependent repression. In contrast, the average adjusted allelic ratio (log2) for SNPs in position 8 (0.03), which resulted in remaining 6mer matches (2-7) in both alleles, had no differences, showing effective repression mediated by 6mer seed matches (2-7). Therefore, 5mers have no effect on miRNA-dependent repression *in vivo*. Error bars indicate standard error.

Supplemental Figure 6



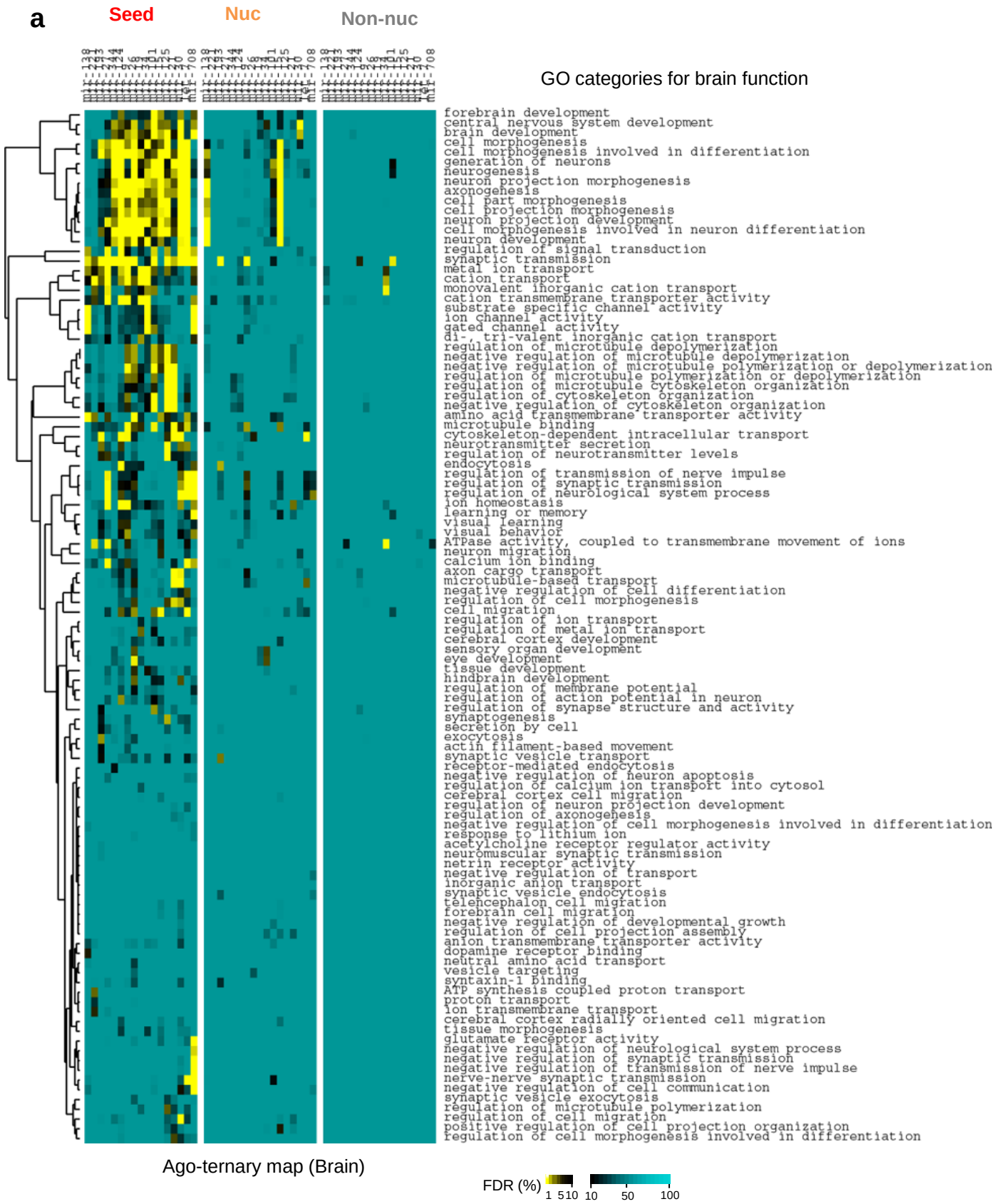


c

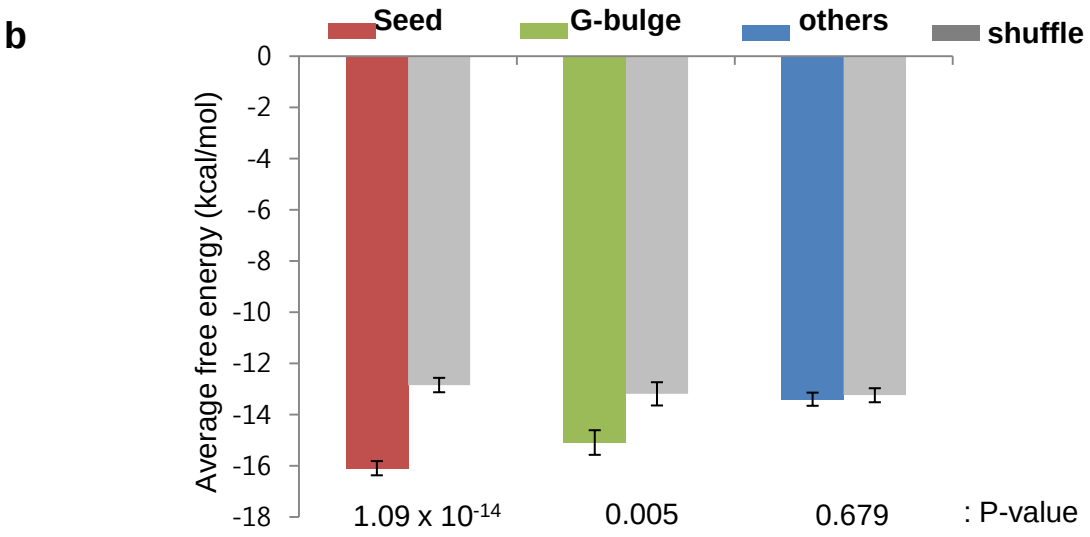
Sample	Transfected miRNA	Antibody	Raw tags (prefiltered)	Aligned tags (hg18)	% align	Unique tags
HeLa1	miR-708	2A8	36,841,031	23,418,057	64	1,246,308
HeLa2	miR-708	2A8	35,323,581	21,539,127	61	1,162,206
HeLa3	miR-708	7G1-1	37,375,073	21,818,310	58	1,490,172
HeLa4	miR-708	7G1-1	36,086,198	21,551,020	60	1,542,616
HeLa5	control	2A8	33,172,073	18,699,503	56	1,080,599
HeLa6	control	2A8	41,050,522	19,445,602	47	2,212,576
HeLa7	control	7G1-1	33,353,630	19,443,341	58	1,250,490
HeLa8	control	7G1-1	30,542,821	14,544,558	48	1,183,258

Supplementary Figure 6. Application of pivot pairing rule to let-7 and miR-708. a. Validation of let-7 A-bulge sites in Kif5b by luciferase reporter assays. Orphan clusters in 3'UTR of Kif5b contains A-bulge site to let-7 and the sequences used for validation by luciferase reporter assay as described in Supplementary Fig 2. b-c. Argonaute HITS-CLIP in HeLa cells transfected with control or miR-708 microRNAs. Argonaute HITS-CLIP in HeLa cells transfected with control or miR-708 microRNAs. b. ^{32}P-labeled RNA crosslinked to Ago immunoprecipitated with 2A8 and 7G1-1* from miR-708 or control miRNA transfected HeLa cells are shown. c. HITS-CLIP sequencing results of RNAs isolated from the 130 kD RNA-protein complex in 8 HeLa experiments with two different miRNA transfections (miR-708 and control miRNA) and two different antibodies (2A8 and 7G1-1). Raw tags (prefiltered) refer to number of Illumina tags prior to default filtering; aligned tags refers to number of tags that could be aligned with genomic sequences from human genome database (hg18), derived from the combination of two alignment results using BLAT and ELAND (Efficient Large-Scale Alignment of Nucleotide Databases; provided by Illumina) program with same criteria used in AGO-CLIP in mouse brain (> 90% identity, Tag start $\leq$3) as described previously[1]. Unique tags are tags with different 5' ends or different degenerative 4 nucleotide barcode introduced in 5' fusion linker.

Supplemental Figure 7



Supplemental Figure 7



Supplementary Figure 7. GO analysis of brain Ago mRNA ternary map and free energy calculation of de novo Ago-miR-124 clusters. a. GO analysis of Ago mRNA ternary map decoded by seed pairing or pivot pairing rules. The GO analysis of target transcripts for each of the Ago top 20 miRNA ternary complexes predicted by canonical seed (8741 clusters, with 6-mers in position 1-8) or pivot pairing rule (1441 clusters, 7-mers) in 11,463 Ago footprint regions (11,463 peaks identified in 11,118 brain Ago-mRNA clusters)[1] is shown. Nucleation bulge-miRNA interactions are estimated as ~15% (1441 clusters, if only Watson-Crick pivot pairing was considered) or ~22% (2162 clusters, if both Watson-Crick and G:U wobble pivot pairing were considered) from the top Ago-miRNA-mRNA interactions (~88% of Ago bound miRNAs present in the top 20 miRNA families[1]). For each category, enrichment was compared with all P13 brain expressed transcripts using DAVID (http://david.abcc.ncifcrf.gov/). False discovery rate (FDR) was calculated for each GO category in biological process and molecular function (index level 5) and is represented as a different color as indicated when it passed the given threshold (count = 2, EASE score = 0.5). Hierarchical clustering of miRNAs and GO categories was performed using Cluster program and visualized as heatmap and tree by Treeview (http://rana.lbl.gov/EisenSoftware.htm). GO categories were divided into two large groups, one with little or no miRNA enrichment, and a second with a large amount of miRNA enrichment in GO categories enriched in brain functions. Since the non-nucleation bulge of miR-181 (GA**T**ATGT; bold indicating bulged nucleotide; match to position 2-7) overlaps with seed match of miR-297b-3p (TATGTAT; match to position 2-8; underline indicating overlapped region, rank 135 among Ago-miRNAs; 6712 tags), non-nucleation bulge of miR-17 (CA**T**CTTT) overlaps with seed match of miR-143 (TCATCT; match to position 3-8, rank 93 among Ago-miRNAs; 15878 tags), and nucleation bulge of miR-153 (TA**T**TGCA) overlaps with nucleation match of miR-669 (AT**T**GCAT, rank 119 among Ago-miRNAs; 9010 tags) and seed match of miR-19 (TTGCACA, rank 27 among Ago-miRNAs; 228275 tags), these miRNAs were not included in the analysis. b. Calculations of free energies from three types of clusters annealing to miR-124. The free energies are calculated by RNAduplex[7] and average values are compared with the shuffled sequences (100 times; average values are indicated as gray bar). P-values from t-test are indicated. Error bars indicate standard error.

Supplementary Table 1

Type of seed match	Motifs	mir-124 seed	Obs /Exp	p-value	Observed number (obs)	Expected number (exp)	seed matched miRNAs
Perfect match	GCCTTA	6mer (1-6)	5.798	0.00	459	79	miR-124
Perfect match	TGCCTT	6mer (2-7)	4.150	0.00	867	209	miR-124
Perfect match	GTGCCT	6mer (3-8)	3.796	1.15×10^{-283}	629	166	miR-124
Perfect match	TGCCTTA	7mer (1-7)	10.212	0	287	28	miR-124
Perfect match	GTGCCTT	7mer (2-8)	8.993	0	374	42	miR-124
Wobble (4)	GC**T**TTA	6mer (1-6)	1.853	4.07×10^{-16}	170	92	none
Wobble (4)	TGC**T**TT	6mer (2-7)	1.186	0.01	241	203	none
Wobble (4)	GTGC**T**T	6mer (3-8)	1.167	0.06	146	125	none
Wobble (4)	TGC**T**TTA	7mer (1-7)	1.879	5.54×10^{-7}	61	32	none
Wobble (4)	GTGC**T**TT	7mer (2-8)	1.369	0.02	52	38	none
Wobble (5)	G**T**CTTA	6mer (1-6)	1.237	0.05	85	69	miR-208
Wobble (5)	TG**T**CTT	6mer (2-7)	1.132	0.07	208	184	miR-743
Wobble (5)	GTG**T**CT	6mer (3-8)	1.165	0.03	192	165	none
Wobble (5)	TG**T**CTTA	7mer (1-7)	1.670	4.40×10^{-4}	46	28	miR-743
Wobble (5)	GTG**T**CTT	7mer (2-8)	1.478	2.74×10^{-3}	58	39	miR-743
Wobble (4,5)	G**TT**TTA	6mer (1-6)	1.032	0.72	130	126	none
Wobble (4,5)	TG**TT**TT	6mer (2-7)	0.919	0.18	257	280	none
Wobble (4,5)	GTG**TT**T	6mer (3-8)	1.219	4.07×10^{-4}	209	171	none
Wobble (4,5)	TGT**T**TTA	7mer (1-7)	1.310	0.03	67	51	none
Wobble (4,5)	GTGTTTT	7mer (2-8)	1.116	0.42	53	47	none
Bulge (A:3-4)	GTGCC**A**T	6mer (3-8)	1.096	0.59	35	32	miR-183
Bulge (A:3-4)	TGCC**A**TT	6mer (2-7)	1.344	0.03	53	39	miR-183, miR-652, miR-685
Bulge (A:3-4)	GCC**A**TTA	6mer (1-6)	1.567	0.01	29	19	miR-652, miR-685
Bulge (G:3-4)	GTGCC**G**T	6mer (3-8)	0.849	0.62	9	11	none
Bulge (G:3-4)	TGCC**G**TT	6mer (2-7)	1.309	0.35	12	9	miR-337-5p
Bulge (G:3-4)	GCC**G**TTA	6mer (1-6)	1.212	0.67	5	4	miR-337-5p
Bulge (C:3-4), (C:4-5), (C:5-6)	GTGCC**C**T	6mer (3-8)	1.170	0.26	52	44	none
Bulge (C:3-4), (C:4-5), (C:5-6)	TGCC**C**TT	6mer (2-7)	1.028	0.85	49	48	none
Bulge (C:3-4), (C:4-5), (C:5-6)	GCC**C**TTA	6mer (1-6)	1.167	0.50	19	16	none
Bulge (A:4-5)	GTGC**A**CT	6mer (3-8)	1.471	0.02	38	26	miR-152, miR-291b-3p, miR-301, miR-130, miR-148, miR-721
Bulge (A:4-5)	TGC**A**CTT	6mer (2-7)	2.581	2.69×10^{-20}	88	34	miR-152, miR-291b-3p, miR-301, miR-130, miR-148, miR-721, miR-17, miR-292-3p, miR-295, miR-294, miR-467, miR-20, miR-93, miR-105, miR-302, miR-106, miR-290-3p, miR-291a-3p
Bulge (A:4-5)	GC**A**CTTA	6mer (1-6)	1.752	2.64×10^{-3}	28	16	miR-17, miR-292-3p, miR-295, miR-294,miR-467, miR-20, miR-93, miR-105, miR-302, miR-106, miR-290-3p, miR-291a-3p
Bulge (T:4-5)	GTGC**T**CT	6mer (3-8)	0.921	0.61	38	41	none
Bulge (T:4-5)	TGC**T**CTT	6mer (2-7)	1.060	0.65	61	58	none
Bulge (T:4-5)	GC**T**CTTA	6mer (1-6)	0.454	0.02	9	20	miR-335-5p
Bulge (G:4-5)	GTGC**G**CT	6mer (3-8)	0.139	1.09×10^{-3}	2	14	none
Bulge (G:4-5)	TGC**G**CTT	6mer (2-7)	0.585	0.12	8	14	none
Bulge (G:4-5)	GC**G**CTTA	6mer (1-6)	0.940	0.90	4	4	none
Bulge (A:5-6)	GTG**A**CCT	6mer (3-8)	2.165	4.19×10^{-14}	91	42	miR-1193, miR-540-3p, miR-1898
Bulge (A:5-6)	TG**A**CCTT	6mer (2-7)	1.826	2.53×10^{-8}	83	45	miR-540-3p, miR-1193, miR-1898, miR-493
Bulge (A:5-6)	G**A**CCTTA	6mer (1-6)	2.112	3.22×10^{-6}	37	18	miR-493, miR-18
Bulge (T:5-6)	GTG**T**CCT	6mer (3-8)	1.240	0.09	60	48	miR-1197
Bulge (T:5-6)	TG**T**CCTT	6mer (2-7)	1.421	2.40×10^{-3}	74	52	miR-1197
Bulge (T:5-6)	G**T**CCTTA	6mer (1-6)	1.830	3.48×10^{-4}	34	19	none
Bulge (G:5-6), (G6-7)	GTG**G**CCT	6mer (3-8)	3.454	2.69×10^{-76}	197	57	none
Bulge (G:5-6), (G6-7)	TG**G**CCTT	6mer (2-7)	3.297	2.69×10^{-72}	203	62	none
Bulge (C:5-6)	GTG**C**CCT	6mer (3-8)	1.170	0.26	52	44	none
Bulge (C:5-6)	TG**C**CCTT	6mer (2-7)	1.028	0.85	49	48	none
Bulge (C:5-6)	G**C**CCTTA	6mer (1-6)	1.167	0.50	19	16	none
Bulge (A:6-7)	GT**A**GCCT	6mer (3-8)	1.899	7.30×10^{-5}	37	19	miR-690
Bulge (A:6-7)	T**A**GCCTT	6mer (2-7)	2.281	6.62×10^{-10}	53	23	miR-690
Bulge (T:6-7), (T:7-8)	GT**T**GCCT	6mer (3-8)	2.779	2.39×10^{-19}	71	26	miR-31
Bulge (C:6-7)	GT**C**GCCT	6mer (3-8)	0.687	0.26	9	13	none
Bulge (C:6-7)	T**C**GCCTT	6mer (2-7)	1.472	0.07	21	14	none

Supplementary Table 1. All possible wobble and bulges in miR-124 seed sequences present in Ago-mRNA clusters. All possible wobble and bulge variants of miR124 seeds were searched in Ago-mRNA footprints (p13 mouse brain) by comparing observed frequency in Ago-mRNA clusters (obs) versus expected frequencies in brain expressed transcripts (exp) as described[1]. It is apparent that G-bulges at position 5-6 or 6-7 are enriched comparable to seed sites (highlighted as red; exp/obs > 3), while C-bulges at 5-6 or 6-7 are not. Note that these miR-124 wobble and bulge sequences could match to a number of other brain expressed miRNAs other than miR-124 (A bulges at position 5-6, which match to a number of such other miRNAs, did not appeared as significant motif in de novo Ago-miR-124 clusters as shown in Fig. 1 and Supplementary Fig. 2b) and that the bulge (T:3-4) was not searched because it is the same as seed (2-8). P-value was calculated based on chi-square test.

Supplementary Method

Ago HITS-CLIP analysis

For bioinformatic analysis of CLIP tags, previous analysis method were used[1]. Briefly, CLIP tags were aligned to human genome (hg18) using BLAT and E-LAND (combination of mapping results from two methods), further visualized and analyzed with UCSC genome browser (http://genome.ucsc.edu/) and Galaxy (http://galaxy.psu.edu). Normalization of tags was performed by *in silico* random CLIP method as described previously[1].

To select conserved de novo miR-124 clusters (overlapping *de novo* clusters both in mouse brain (p13) and miR-124 transfected HeLa, 62nt footprints), genomic coordinations of mouse Ago mRNA clusters (mm8, identified in mouse brain)[1] were converted into genomic coordinations of human (hg18) by using liftOver (http://genome.ucsc.edu/cgi-bin/hgLiftOver) and then they were compared with the genomic coordination of de novo Ago miR-124 clusters (HeLa)[1]. Such conserved de novo Ago clusters were used to estimate how frequently different types of binding sites are used *in vivo* by overcoming the heterogeneity of miRNAs in the brain and the possible artificial Ago-mRNA interactions caused by overexpressing miRNA in HeLa.[1]

For counting conserved motifs which are conserved across more than 4 of 5 species (human, mouse, rat, dog and chicken), Multiz8way results from UCSC genome browser (http://genome.ucsc.edu) were used to calculate conservation rates (number of conserved motifs / number of all motifs). Conservation rates in all 3' UTRs (defined by RefSeq) were calculated for three different sites (seed, nucleation bulge and non-nucleation bulge) of all mouse miRNAs by counting conserved 7-mer motifs, and their distributions were plotted by the number of motifs (Fig. 3c). As a background control, conservation of all 7-mers was also calculated and its distribution was plotted as a ratio in the population.

Ago HITS-CLIP mapping

Ago HITS-CLIP mapping was performed for 7-mer seed matches (position 2-8), nucleation bulge matches (7-mers derived from 6-mer seed matches (position 2-7), in which the nucleotide of the position 5-6 mRNA bulge sequence is complimentary (Watson-Crick base pairing) to position 6 of corresponding miRNA), or non-nucleation bulge matches (7-mers derived from 6-mer seed matches (position 2-7), in which the nucleotide of the position 5-6

bulge sequence is identical to the position 6 of miRNA) of all mouse miRNAs (miRBase, release 15) in all 3'UTR regions (defined by mouse RefSeq track in genome browser) or of the top 20 miRNAs bound to Ago in 11,463 Ago footprint regions (11,463 peaks defined in 11,118 clusters, 62nt) identified in brains (BC $\geq$ 2)[1]. Notably, we noted that seed and bulge sites share 5 nucleotides (position 2-6) and therefore used 7-mer motifs, which could differentiating each seed and bulge sites, to search for nucleation bulges in all analysis. If the nucleation bulges or non-nucleation bulges overlapped with seed matches of other miRNAs (all 6-mer matches in position 1-8), they were not used in the calculations. Orphan clusters (3073 clusters, 27% of total Ago mRNA clusters from p13 mouse brain, BC $\geq$ 2), were selected as Ago HITS-CLIP clusters in which Ago footprint regions (46 nt, estimated Ago bound regions from the peaks of clusters)[1] have no predicted seed matches among the top 20 Ago-miRNAs families. Additional data access for Ago HITS-CLIP can be found at our project websites, http://ago.rockefeller.edu/ and http://wiz.skku.edu/swchi.

Motif analysis

For motif analysis, MEME (http://meme.sdsc.edu) with the 'zoops' model was used, setting motif width from four to eight (others with default parameters). WebLogo 3 (http://weblogo.threeplusone.com/) was used for visualization of motifs. Enriched motifs (E-value < 0.01) were identified in the orphan Ago footprint regions (62 nt; less stringent, $\geq$ 95% of clusters, described as previously)[1] by MEME analysis.

Distribution analysis of motifs relative to cluster peak was analyzed by cubic spline interpolation analysis as described previously[1].The number of motifs in the y-axis of distribution analysis (Fig 1a, right panel) was obtained by counting a given motif in bin sizes of the same motif length at each nucleotide position, sliding down the bin from peaks with increment of one nucleotide

Meta-analysis

Meta analysis was performed for brain expressed transcripts harboring the three different sites (as in c) by obtaining normalized compiled data[8] from microarray experiments using 7 different miRNA transfections (miR-30[9], miR-9[10], miR-181[3,10], let-7[9], miR-34[11], miR-106[12] and miR-16[9,12]; the analysis represents all published data available from the top 30 Ago-bound miRNAs family (based on sequence in position 2-8)) and analyzing Cumulative fraction depending on fold change ($\log_2$ ratio) as described previously[1]. For biological replicates or

experiments for the same miRNA family (defined by 6-mer seed (position 2-7) sequences), median values were used in Fig. 3e. In Fig. 3g transcripts showing robust changes (changes in steady-state levels after miRNA injection; $P < 0.05$, MZdicer+miR430 vs. MZdicer)[13] were selected and their median values from biological replicates were used for the analysis. For the meta-analysis in Fig. 3h, orphan transcripts were identified by analyzing microarray data from miRNA transfections (8 miRNA families; 18 different miRNAs, 20 microarray experiment sets) in a Dicer-null HCT116 cell line[12]. In detail, only transcripts showing miRNA-dependent changes ($P < 0.05$, for siRNA-141, hsa-miR-15a, hsa-miR-16, hsa-miR-192, hsa-miR-215, hsa-miR-17-5p, miR-20, hsa-let-7c, miR-195, miR-106b, miR-15a, miR-15b, miR-16, miR-103, miR-107, miR-16, miR-15a, miR-106b, siRNA-200a and siRNA-200b) were analyzed after selection from the GEO database (GSM156534, GSM156545, GSM156546, GSM156547, GSM156548, GSM156553, GSM156554, GSM156557, GSM156559, GSM156576, GSM156577, GSM156578, GSM156579, GSM156580, GSM156581, GSM156522, GSM156523, GSM156524, GSM156532, GSM156533)[12]. To condense microarray results for the same miRNA family in Fig. 3h, only transcripts without any significant opposite fold change among different miRNA transfections were selected and their median values were used for the analysis. KS test was performed by using total orphan transcripts as control; a null hypothesis that all orphan transcripts are indirect miRNA targets, assuming that all miRNA binding sites are seed matches.

Linear regression analysis

Linear regression analysis for motifs was performed by comparing the frequency of conserved sites (more than 4 of 5 species; seed, nucleation bulge and non-nucleation bulge) in Ago-mRNA clusters (62 nt footprints, BC $\geq$ 2) with the frequency of top 100 miRNAs experimentally determined by mouse brain Ago HITS-CLIP[1]. For this analysis, frequency of conserved sites was calculated as the sum of cluster peak heights (determined by cubic spline interpolation) containing the given motif and used as log 2 value. Frequency of each top 100 Ago-bound miRNAs (~99% of all miRNAs bound to Ago) was calculated as log 10 value.

Gene ontology analysis

Gene ontology (GO) analysis was done using DAVID (http://david.abcc.ncifcrf.gov/). Hierarchical clustering of GO categories was performed using Cluster program and

visualized as heat map and tree by Treeview (http://rana.lbl.gov/EisenSoftware.htm) as described previously[1]

References

1. Chi, S.W., Zang, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479-86 (2009).
2. Lim, L.P. et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769-73 (2005).
3. Baek, D. et al. The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).
4. Zisoulis, D.G. et al. Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. *Nat Struct Mol Biol* **17**, 173-9 (2010).
5. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).
6. Kim, J. & Bartel, D.P. Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression. *Nat Biotechnol* **27**, 472-7 (2009).
7. Hofacker, I.L. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**, 3429-31 (2003).
8. Khan, A.A. et al. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol* **27**, 549-55 (2009).
9. Selbach, M. et al. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58-63 (2008).
10. Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91-105 (2007).
11. He, L. et al. A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130-4 (2007).
12. Linsley, P.S. et al. Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol* **27**, 2240-52 (2007).
13. Giraldez, A.J. et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75-9 (2006).