Supplementary Methods

Ago CLIP

CLIP was performed as described using a monoclonal Ago antibody, 2A8, generously supplied by Z. Mourelatos[1]. Experiments using the 7G1-1* antibody were undertaken after recognizing that this monoclonal antibody, obtained from batches generated by the Iowa Developmental Hybridoma (IDH) bank (Cat # 7G1-1), were a mixture of clones of anti-FMRP antibody and anti-Ago antibody, an observation confirmed by mass-spectrometry sequencing of immunoprecipitated products in WT and FMRP KO brain (Supplementary Fig. 1), and confirmed independently (Z. Mourelatos, personal communication). Attempts to subclone independent hybridoma clones from a different frozen batch of cells obtained from the IDH were not successful (as subsequent batches from IDH harbored only the original 7G1-1 clone). CLIP with the 2A8 antibody was done in sibling P13 mice from freshly dissected neocortex. CLIP with the 7G1-1* antibody done in WT P13 mouse brain was in the presence of 200ug of blocking peptide (KHLDTKENTHFSQPN; mapped to FMRP amino acid 354 to 368), which has been demonstrated to block FMRP immunoprecipitation by 7G1-1[2]. In addition, one experiment was done in sibling FMRP KO P13 mouse brain, although only tags that were also present in independent experiments were considered (see Supplementary Fig. 5). 7G1-1 and 7G1-1* were obtained from the Developmental Studies Hybridoma Bank, developed under the auspices of the NICHD and maintained by the University of Iowa, Department of Biological Sciences, Iowa City, IA 52242. For immunoblot analysis, the anti-Ago 7G1-1* antibody was used to detect Ago proteins; results were confirmed independently with an independent anti-Ago antibody (Upstate:04-085). Other antibodies used (Supplementary Figure 1) were generously supplied by others; anti-Ago1 and anti-Ago2 antibodies were from T. Tuschl, anti-FMRP (2F5) antibody was from J. Fallon.

Exon arrays in P13 mouse brain.

For estimating the level of brain transcripts, total RNAs from three P13 mouse brains were extracted using Trizol and RNAeasy kit and mRNA was amplified and labeled by the method provided by Affymetrix. Mouse MoEx 1.0ST Arrays were used for measuring signal intensity of each exon in the samples. To process the signals from the array, quantile normalization and PM-GCBG (signal adjustment based on the

background with similar GC content) were applied. The IterPLIER method was used for selecting appropriate "core exon" probes to estimate gene-level intensities. The presence of transcript in the P13 brain was determined by the p-value derived from DABG method (detection above background, $P<0.05$) and genes with more than 100 in final total probe intensity were selected. Finally, median $\log_2$ values from three biological replicates are used to estimate level of transcripts in the P13 brain. Analyses were performed by using Affymetrix Power Tools.

Bioinformatic analysis of CLIP tags.

CLIP tags were aligned to mm8 genome using BLAT and to miRNAs from miRBASE (http://microrna.sanger.ac.uk) using BLAST, further visualized and analyzed with UCSC genome browser (http://genome.ucsc.edu/) and Galaxy (http://galaxy.psu.edu). General bioinformatics analysis including *in silico* random CLIP were performed as Python script utilizing BioPython (http://biopython.org); linear regression analysis for motifs by MatrixREDUCE[45]; microarray data for miR-124 targets from GEO database (http://www.ncbi.nlm.nih.gov/geo); microarray analysis by BioConductor (http://www.bioconductor.org/); statistical tests by Scipy (http://www.scipy.org/).

Normalization of CLIP tags using in silico random CLIP algorithm.

Details are given in Supplementary Figs. 4-5.

Ranking of Ago-miRNA CLIP tags.

The ranking of miRNAs is based on the frequency of each ~22 nt sequences in Ago-miRNA tags. Where multiple members of family of miRNAs have identical seeds, we did analyses of the family in aggregate, summing the frequency of each member.

Normalization of Ago-mRNA clusters along transcripts (Fig. 1h).

To normalize for differences in the length of transcripts in the CDS and 3' UTR, the number of clusters at each position was divided by the total number of transcripts at that position (blue graph). The standard deviation of fraction in each position was estimated based on the assumption that prediction result is binomial distribution (drawn as light blue graph). To determine whether the different length of transcripts biased the position of clusters, the position of a cluster within each transcript was randomly redistributed. These control clusters ("control for mRNA length") were also

normalized and replotted (red graph, with standard deviation determined as above, drawn in pink).

Distribution of tags relative to cluster peaks (Fig. 2).

We plotted the distribution of Ago-mRNA tags within clusters. To compensate for differences in the number of CLIP tags present in different clusters, we randomly chose 30 tags from each of the 61 most robust clusters (single peak, BC = 5, and $\geq$ 30 tags per peak; Fig. 2a) or 135 robust clusters (BC=5, $\geq$ 30 tags per peak; Fig. 2c). The peak of each cluster was determined by cubic spline interpolation analysis (see next section), and the distribution of tags relative to the peak was plotted (brown graph, Fig. 2a and 2c).

For the distribution of miRNA binding sites in 135 robust clusters (Fig. 2c), conserved seed sites (position 2-7) of top 30 miRNAs and bottom 30 miRNAs (Supplementary Fig.2) were searched in Ago CLIP cluster regions and plotted as the relative position from the peaks. If width of an Ago CLIP cluster was more than 100nt from the peak, we restricted the search within +/- 100nt window from the peaks.

To determine the distribution of conserved miR124 seed sites (6mers in position 2-8) in Ago-mRNA clusters (BC $\geq$ 2; Fig. 2d), we first determined the number of clusters as a function of relative position from the peak (plotted as fraction value; green graph, Fig. 2d). Because of the variability in the width of AGO CLIP cluster, the number of predicted conserved mir-124 target sites in each relative position from the peak of the cluster was divided by the number of clusters in each position and indicated as fraction (red graph, Fig. 2d). Standard deviation of fraction in each position was estimated based on the assumption that prediction result is binomial distribution. The 884 conserved miR-124 seed sites noted in Fig. 2d correspond to the 1,393 raw 6-mer sites identified in Supplementary Table 3 after collapsing overlapping 6-mer sites into single seed sites.

Cluster peak analysis using cubic spline interpolation.

Based on the position of tags in genome, the number of unique tags in each nucleotide position was calculated for all AGO-mRNA CLIP clusters. Cubic spline interpolation method was applied to interpolate tag density in the clusters by using Scipy (http://www.scipy.org/). By determining the derivative of the function at each point of the interpolation, locating the point where the derivative = 0, and confirming

that the derivative changes from positive to negative around this point, the location and number of peaks per cluster were determined. Excess kurtosis was determined using Scipy (http://www.scipy.org/).

Prediction of miRNA binding sites

Based on the observation that enriched 6-8 mer seed sequences were present in the Ago footprint region (Supplementary Fig. 9, and data not shown), we identified all 6-mer sequences in miRNA seed position 1-8 to identify candidate 6-8 mer miRNA binding sites. For selecting conserved sites, Multiz8way results from UCSC genome browser (http://genome.ucsc.edu) were used to search for seed sites that were conserved across more than 4 of 5 species (human, mouse, rat, dog, chicken). Predicted miRNA binding sites were downloaded from databases available from the following websites, using precompiled batch results and default parameters set by the developers of each algorithm; miRBase (http://microrna.sanger.ac.uk/ , miRanda algorithm, Sep 2008 version), TargetScan 4.1 (http://www.targetscan.org/ downloading mm8 bed format files from the www site), PicTar4 (predictions in human were obtained from UCSC genome browser and converted to mouse by the liftOver program), PITA (http://genie.weizmann.ac.il/pubs/mir07/index.html, TOP prediction), and RNA22[3] (http://cbcsrv.watson.ibm.com/rna22.html, precompiled data for the 3'UTR of mouse transcripts (with parameters: G=0 M=14 E=-25Kcal/mol) were mapped to mouse genome by ELAND (Efficient Large-Scale Alignment of Nucleotide Databases; provided by Illumina)).

Estimates of false positive rates and specificity

We estimated false positive rates of the Ago ternary map using three different approaches.

1) We estimated a false positive rate from data in Fig. 2c by comparing the number of conserved seed matches of the bottom 30 and top 30 Ago-miRNA in the 62nt Ago footprint region (134 robust clusters; 8/118 ~ 6.8%). This data also illustrates the improvement in specificity achieved by confining the analysis to the Ago footprint region; the false positive rate for the entire 200nt cluster window in Fig. 2C is 41/171 ~ 24%.

A similar method was to compare the number of conserved seed matches in the bottom 20 and top 20 Ago-miRNAs in the 62nt Ago footprint region (BC$\geq$2,

Supplementary Figure 13C). The false positive rate was estimated as ~10% (~3.37 $\log_2$ fold ratio (top 20 vs bottom 20) = ~10 fold increase).

2) We also estimated false positive rate from the data in Fig. 2d by calculating the number of clusters harboring conserved miR-124 seeds expected by chance as follows. We identified 13,272 conserved miR-124 seed sequences in 52,164,693 nt transcribed in P13 brain (from microarray and RefSeq data; ~0.0016/62 nt Ago footprint). We also identified 1393 clusters harboring conserved miR-124 seeds in the Ago footprint region (62 nt) of 11,118 total Ago-mRNA clusters (BC > 2; Fig. 2d). We calculated the number of miR-124 clusters expected by chance (among p13 brain transcriptome) in the footprint region of these as 11,118 clusters x 0.0016 miR-124 seeds/random Ago footprint = ~175 clusters by chance. Therefore the estimated false positive rate of Ago-miR-124 clusters is 175/1393 ~ 12.6%.

A similar method was applied to the top20 miRNA seed matches in 62nt Ago footprint region (BC≥2, Supplemental Table 3, Supplementary Figure 13D). The false positive rate was estimated by comparing the number of conserved seed matches from the top 20 Ago-miRNA with the expected number by chance (among the P13 brain transcriptome) in the Ago footprint region (~3.02 $\log_2$ fold ratio (observed vs expected) = ~8 fold increase = ~12.5% false positive rate).

False positive rates from other seed based methods combining proteomic approaches were estimated[4,5]. They were estimated as ~40%[6] or ~66%(based on the reports that two-thirds of the predicted targets appeared to be nonresponsive to miR-223 loss in neutrophils[4].) This higher number of false positives is also in agreement with findings from two other approaches using purification of miRNA-protein complexes (>50%[7] or ~70%[8];based on reports about ~30% of seed-containing true positives). Therefore we concluded that the Ago-miR-124 ternary map (with a false positive rate of ~6.8-12.6%) achieved a minimum of a 3-fold, and a maximum of a 10 fold improvement in false-positives relative to other approaches (~40-70%).

We also analyzed observed vs. expected frequencies for each seed match from the top 30 miRNAs in Ago-mRNA clusters (BC≥2; Supplementary Table 3). Some apparent false-positives are in this group, as estimated by observed vs. expected seed enrichment (e.g. miR-125, miR-708, miR-324-3p); this is also apparent in the erratic shape of the curves in Supplemental Figure 13. However, even within the top

20 robust miRNA set, we do not know the rules of miRNA binding, and this may impact upon the false positive set. For example, miR-125 has no enrichment of the observed vs. expected 6-mer seed frequency, but if we include seed conservation, the enrichment becomes statistically significant. In addition, even without including seed conservation, if we examine observed vs. expected miR-125 8-mer seed frequency, then we again see statistically significant seed enrichment in Ago-mRNA clusters. Moreover, this data makes biologic sense, since miR-125 is expressed in the brain, and the GO targets identified match the known biology (with roles in cytoskeletal regulation; Figure 5 and Supplementary Figure 14).

3) False positive rate and specificity of Ago-miR-124 ternary cluster map were assessed by comparing it with meta-analysis of mir-124 microarray data (Fig 4 and Supplementary Figure 11), especially with two sets – decreased transcripts enriched in direct targets (703 true set, negative fold change) and increased transcripts enriched in indirect targets (575 false set, positive fold change). The false positive rate was estimated as ~27% and specificity was 92.5% (See Supplementary Figure 11 for detail).

Estimates of false negative rates and sensitivity

The issue of how to estimate false negatives is clearly important, but it also must be recognized that this is a difficult issue, because there are in fact no gold standard "true" dataset upon which to answer this question other than the dataset for miR-124. Therefore, we focused our attempts to estimate false negatives on the set of miR124 targets validated by Hannon and colleagues[9]. Unfortunately, this set is only 22 deep. Nonetheless, this would yield a false negative rate of 6/22~25%. For transcripts of even moderate abundance (normalized probes intensity >700), we identified all 10/10 predicted targets (Table 1). Moreover, if we examine raw clusters (i.e. without using in *silico* normalization), we identified 21/22 predicted targets (~0.05% false negative rate).

We also could address sensitivity from this analysis. The 6 potential false negatives in Table 1, transcript levels were low (normalized probe intensity was < 700 for each transcript (excluding Tom1l1, which appears to have an unusual antisense transcript expressed in brain covering the 3' UTR). Of all 12 transcripts in Table 1 with transcript levels at or below 700, 7 were identified, while 5 were missed, suggesting that for the rarest transcripts, we detected 58% of true positives. Classifying our data

in this fashion, we have very low false negative rate estimate (10/10 positives) for over 50% of the transcriptome (> 700; versus median expression level = 771, and average expression level ~1,255). We also compared the recovery of transcripts as a function of abundance versus our ability to identify miR-124 targets. This result (Supplementary Figure 13E) confirms those observed on a smaller scale in Table 1, which is that low abundance transcripts are (not surprisingly) recovered less efficiently.

We can also estimate the sensitivity of the Ago-mRNA map as it relates to miRNA abundance. We readily detected miRNAs present at less than 1% of the total population (Supplemental 2c, and Supplemental Table 3) and the seed signal for these miRNAs showed approximately equal enrichment between the top of this list (e.g. #1, miR-181a, 6% of total Ago-miRNA, Supplemental Fig. 2c, shows 1,190 sites, and a 1.7 fold enrichment in the Ago-mRNA footprint region) and the end of this list (#29, miR-23, <0.36%, which shows 1,036 sites, and a 1.4 fold enrichment). Therefore, the top 20 miRNAs provide a robust set of enriched seed sequences in the Ago-mRNA footprints (Supplemental Table 3, Supplementary Figure 13), and suggests that Ago-HITS-CLIP is likely to work well for additional miRNAs beyond this set.

Another way of estimating the false negative rate is by comparing the number of Ago-mRNA clusters with no predicted seeds in the top 20 Ago-miRNAs, relative to the proportion of the top 20 Ago-miRNAs in the brain. Thus, 27% of Ago-mRNA clusters (BC > 2) have no predicted 6-mer seed sites from among the top 20 Ago-miRNAs families (comprising 88% of all Ago miRNAs). This seed-driven approach yields an estimated false negative rate of ~15% (orphan clusters (27%) – Expected number of clusters from below top 20 (12%) = ~15%). This compares favorably with similar seed driven estimates of false negatives rates of 50% to 70% made by other investigators[4,5,8]. We also recognized that even though this seed-based approach produced reasonable estimates for our false negative rate, currently unknown miRNA binding rules/sites might affect our false negative rate.

Ago miRNA ternary map; predicting miRNA sites in the peak Ago footprint (Fig. 5a).

To generate Ago miRNA ternary maps, we searched for all 6-8 mer seeds of the top 20 miRNAs in the Ago footprint region, defined as -30 to +32 from the peak (defined by cubic spline interpolation analysis). To select the top 20 miRNAs families in the brain, we ranked miRNAs from Ago-miRNA CLIP summing up the frequency of all

miRNAs with same 6-mer seeds (miRNA family ranking). miRNAs whose family ranking was within the top 50 were selected and the top 20 miRNAs with different seeds were ranked and selected based on the frequency of each miRNA. If more than one seed was present, the one closest to the peak was chosen.

The top 20 miRNAs were selected for Ago miRNA ternary map based on following observations. When comparing the cumulative frequency of 6-mer seeds within the Ago footprint region for all the top (N) Ago-miRNAs with the cumulative frequency of 6-mer seeds using the bottom (N) miRNAs, we found the greatest enrichment of miRNA seeds in Ago-mRNA clusters was present within the top 20 miRNAs (Supplementary Figure 13). Beyond the top 20, there still was enrichment, evident even up to the half-way point of the Ago-miRNA data set (Rank #175, P=0.004; Supplemental Figure 13). In addition, we obtained a similar result by using a different method of analysis. In Supplemental Figure 13, we calculated the frequency of conserved 6-mer seed matches observed in Ago-mRNA footprint regions vs. expected conserved 6-mer matches in the P13 brain expressed transcriptome (derived experimentally from Affymetrix exon array data). We again observed that enrichment of Ago-miRNA seed sequences was clearly evident in the top 20 miRNAs and also beyond top 20, suggesting that Ago-HITS-CLIP is likely to work well for additional miRNAs beyond this set.

We selected BC and peak heights for analysis in two ways. First, we developed an empiric approach based on validation experiments. Based on data from Fig. 2, we used BC2 for subsequent analyses. These results were supported by comparison of our data and published data (presented in Table 1), revealing a high sensitivity of our map with BC $\geq$ 2, independent of peak height. This also reflects our experience with Nova HITS-CLIP[10], in which BC was found to be a more robust parameter than peak height. We note that BC 2 in the current Ago context means reproducible results not only with biologic replicates but also with two different antibody reagents. Nonetheless, we are missing 6 hits in the Karginov data (Table 1), five of which were detected by HITS-CLIP but were below our threshold (either for low peak height or low BC); these cases were uniformly low abundance transcripts. This suggests thresholds of transcript abundance will relate to the sensitivity detected at any BC or peak height.

Meta-analysis of microarray studies.

Raw data or normalized data from 5 microarray studies in miR-124 overexpression (Supplementary Fig. 11A) were obtained from GEO database (http://www.ncbi.nlm.nih.gov/geo). Raw data were normalized by the same method used in each study and finally log2 ratio for each transcript, which is the median value in all replicates and probes, was calculated only if its change is statistically significant ($P < 0.05$, t-test). For the meta-analysis of data to compare it with Ago CLIP, only transcripts showing robust change ($P<0.05$) at least 2 times in 5 studies without any significant opposite fold change were selected. Among these transcripts, 1278 brain expressed transcripts were further selected based on exon array result in P13 mouse brain (10743 genes, $p<0.05$ (DABG), normalized probe intensity >100). To compare the cumulative fraction depending on fold change, Kolmogorov-Smirnov test was performed using Scipy (http://www.scipy.org/).

Gene ontology analysis

Gene ontology analysis was done using GoMiner (http://discover.nci.nih.gov/gominer/). Details are given in Supplementary Fig. 12.

Supplementary Figure Legends

Supplementary Table 1. High throughput sequencing results from Ago HITS-CLIP.

Over 80% of RNAs isolated and sequenced from the upper 130 kD RNA-protein crosslinked band correspond to genomic sequences, and between 72 and 88% of RNAs from the lower 110 kD RNA-protein band correspond to miRNAs. Raw tags refer to number of Illumina tags after filtering step which company provided; BLAT refers to number of tags that could be aligned with genomic sequences from mm8, and BLAST from miRBASE, % is number of matched tags from BLAT search, using parameters ($\geq$ 90% identity, Tag start <=3) or BLAST search result, filtered to include only matches with more than 95% identity. Unique tags are tags with different 5' ends after discarding those with identical genomic location miRBASE from May 2008 was used, which included 579 known mouse miRNAs (including passenger strand).

Supplementary Table 2. Correlation of Ago-miRNA and Ago-mRNA CLIP data

A. We correlated all possible 6 to 8-mer motifs present in Ago-mRNA clusters (BC $\geq$ 2) with cluster height ($\log_2$(maximum tag number) per cluster), using the MatrixREDUCE linear regression model:

http://bussemaker.bio.columbia.edu/software/MatrixREDUCE/[11].

The most enriched motifs selected by default parameters (except –max_gap=0) are shown.

B. Since some miRNAs were missing from the analysis in (A), we analyzed the 7mer seeds (position 2-8) present in the top 10 miRNAs identified from Ago-miRNA CLIP. We compared with their observed (obs) versus expected (exp; from P13 brain transcripts) frequency in Ago-mRNA clusters (BC $\geq$ 2). Seed sequences from 8 of the top 10 Ago-CLIP miRNAs were enriched in Ago-mRNA clusters. The miR124 seed match sequence is over-represented in the mouse genome and also in the mouse exons by about 50% (data not shown). This means that there are more miR-124 potential binding sites in the genome (we calculated this as 1.33 fold (51,461 total 6-mers for miR-124, vs. 38,752 for miR-30), and we interpreted this correlation as being consistent with a relative increase in miR-124 bound sites. It is also possible that there is underlying biologic regulation that we do not understand, or that there are additional rules of binding that are not apparent in current analyses.

Supplementary Table 3. Significance of Ago-miRNA seed sequences in the Ago-mRNA footprint region.

A. This table lists the frequency of the top 30 Ago-miRNA seed sequences in the Ago-mRNA (BC $\geq$ 2) footprint region. Ago-miRNA seed sequences are clustered by miRNA family. The cumulative sum of raw 6-mer seeds for each family is outlined in pink; we did not collapse overlapping raw 6-mer sites to allow us to compare observed and expected frequencies. The observed frequency is the number of times each given seed sequence is present in the total sequence space of the Ago-mRNA footprint region (711,963 nt = (62 nt: -30 to + 32 nt) x 11,483 cluster peaks [defined as discernable Ago cluster peaks within the raw set of 11,118 raw cluster peaks; Suppl Fig. 7]). The expected frequency was calculated by measuring the frequency of a given motif in the P13 brain transcriptome (determined by RefSeq; 52,164,693 nt). P values were calculated by Chi square. Results were also recorded from re-analysis using observed or expected seed sequences that were conserved in 4/5 animals (as in Supplementary Methods). From this table, it is possible to estimate the predicted top seed sequences according to confidence judged by enrichment of observed/expected seed (or conserved) seed frequency. As an illustrative example, such analysis is able to identify the octamer sequence CTCAGGGA as the most likely accurate predictor (most specific) and the hexamer CTCAGG as the most sensitive predictor (but with a weaker signal:noise) of miR-125 targets. If the same analysis is restricted to conserved seed sequences, the same 6-mer sequence shows improved specificity.

Supplementary Figure 1. Argonaute HITS-CLIP with 7G1-1 and 2A8*

A. WT and FMRP KO mouse brain lysates were immunoprecipitated with either 7G1-1* (containing monoclonal antibodies to both FMRP and Ago) or 7G1-1 (containing only FMRP monoclonal antibody). Immunoprecipitates were analyzed by immunoblot probed with 7G1-1* (left panel), anti-Ago1 (right panel) and anti-Ago2 antibody (middle panel). Both 7G1-1 and 7G1-1* antibodies immunoprecipitated FMRP, which is absent in FMRP KO. 7G-1* also IPed Ago1 and Ago2 at ~100 kD from both WT and KO recognized by anti-Ago1 and anti-Ago2 antibodies. B. 7G1-1* immunoprecipitate from KO brain was run on SDS PAGE and stained with Coomassie brilliant blue. Protein bands were analyzed with mass spectrometry. Ago proteins 1-4 were identified at ~100 kD. C. 7G1-1* antibody was pre-incubated with 200ug of

FMRP peptide (354-KHLDTKENTHFSQPN-368 ; the epitope of 7G1-1)[2] or an irrelevant control (FLAG peptide). FMRP peptide, but not control peptide, was sufficient to completely block FMRP immunoprecipitation from WT mouse brain and had no effect on Ago immunoprecipitation. D. ^{32}P-labeled RNA crosslinked to Ago and FMRP immunoprecipiated with 7G1-1* from WT or FMRP KO mice as indicated. The antibody pulls down complexes of ~90kD, which correspond to FMRP, as indicated by the disappearance of this band when IPs performed from FMRP KO brains or in the presence of 200 ug of the competing peptide epitope. For HITS-CLIP experiments, two different animals were used, Brain "D" corresponds to CLIP of WT brain done in the presence of peptide competition and isolated the Ago band (D, lane 3), and Brain "E" corresponds to CLIP of FMRP KO brain (D, lane 2). E. Replicate Ago CLIP with 2A8 monoclonal antibody. Overdigestion of Ago RNA cross-linked complex showed the protein size of Ago (~95kD), which is specific in UV-irradiated tissue (+XL) comparing to non-irradiated brain (–XL).

Supplementary Figure 2. Reproducibility of Ago-miRNA CLIP results from 5 experiments

A. Heat map comparing miRNAs by frequency with which they were detected in Ago-miRNA CLIP in five different experiments using two different antibodies. The number of tags in each experiment was normalized by rank quantile normalization and median value for each miRNA from 5 experiments is used as normalized tag frequency to determine a final ranking. A heat map was generated by the Treeview program (http://rana.lbl.gov/EisenSoftware.htm) with $\log_2$ ratio of normalized tag frequency to median. Red tags are ranked above median, green below, white at the median, grey is missing (no tags). % is the number of tags in total tags from 5 experiments. B. Correlation of Ago-miRNA CLIP results using different antibodies. C. The top 30 miRNAs are shown for each of 5 experiments. D. The bottom 30 miRNAs are shown.

Supplementary Figure 3. Reproducibility of Ago-mRNA CLIP results from 5 experiments

A. To define reproducibility of Ago-mRNA clusters, we first identified inter-animal clusters[10] with at least one tag in each of three biologic replicates. Pearson correlation coefficients represent the correlation between the normalized density of tags ($\log_2$ value after *in silico* random CLIP normalization; see Supplementary Figure 4) in each cluster. B. The analysis in (A) was repeated, comparing reproducibility with the two

different Ago monoclonal antibodies. C. A genome-wide graph of data from (B).

Supplementary Figure 4. in silico random CLIP normalization.

A. Assuming that number of CLIP tags from transient nonspecific protein-RNA interactions are correlated with transcript abundance, then each transcript should have different threshold level to differentiate signal:noise. Therefore we estimated the false discovery rate (FDR) of CLIP tag clusters for each brain transcript using a simulation method named *in silico* random CLIP. Transcript abundance was measured by exon arrays (Affymetrix MoEx 1.0 ST, at the same age (P13) and tissue (mouse brain) used for HITS-CLIP). For simulating transcript abundance, a number for each brain transcript (N_n) was assigned using the normalized probe intensity from the microarray (details in Supplementary Methods). RNase treatment was simulated by introducing cleavage at a random site in each population of transcript in given length (Y_n; determined by RefSeq annotation) and this process was repeated until an average length of 50 +/- 2 nt fragments was obtained. 50 nt was determined based on the observation of mean 50 nt size of Ago-mRNA CLIP tags. When normalized to length and abundance, the number of fragments per transcript $F_n = N_n \times Y_n / 50$. The total number of fragments per transcriptome Tot(F) is the sum of F_n for all transcripts.

To simulate immunoprecipitation of nonspecific and transient RNA-protein interactions, we first calculated the expected number of tags for a transcript based on F_n and the total number of unique CLIP tags in that experiment. Thus for a given CLIP experiment, we calculated the expected number of tags in any given transcript (T_n), by multiplying total number of unique tags from actual CLIP experiments (Z) with fraction of fragments for that transcript (F_n/Tot(F)). We then randomly selected T_n fragments for each transcript, and these were aligned with their position in the given transcript. If the selected fragment was more than 36 nt, the 3' end of the fragment was eliminated to leave only 36 nt to simulate an Illumina sequencing read. We calculated the maximum background cluster height (M_n, maximum number of overlapping tags) from the alignment, and repeated this simulation repeated 500 times for every transcript. FDR (P-value) was determined by counting the observed number of maximum clusters (M_n) from each of 500 repeats, and the cluster height giving $p<0.01$ in each transcript was used as threshold for normalization of HITS-CLIP cluster. In this way, the normalization for HITS-CLIP varied for each transcript according to its abundance, length, and its simulated background cluster height.

Schematic of the process and an example are shown. For example, thresholds for gene A were estimated by in silico random CLIP, and this estimate gives a normalization threshold of 3 (to yield an acceptable P value, < 0.01). Hence in the cartoon on bottom left, clusters of peak height < 3 were removed after normalization.

B. Examples of normalized CLIP tags using in *silico* random CLIP. Examples are shown in chromosome 8. High cluster height in abundant brain transcript was removed (red arrow) but low cluster height in low expressed transcript was kept (purple arrow).

C. To estimate the effect of varying (N) on the outcome of *in silico CLIP,* simulations were done with the gene Chad7 that has the minimum abundance used in our experiments and in *in silico* CLIP. The top table shows that Tn does not change as a function of abundance (Nn). The first column values indicate that the array value of every transcript was increased together with Chad7. All simulation has the same final result which selects Ago CLIP clusters of which height is more than 2. The middle table shows the maximum cluster heights over 500 simulations. No clusters are greater than 2 (Mn). The final FDR calculation based on the results of the middle table (e.g. 12/500=0.024) shows that for a cluster height of 3, the FDR does not change, with each FDR < 0.01. In conclusion, this simulation explains that the results are not affected by (N).

Supplementary Figure 5. Biologic complexity normalization.

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 immunoprecipitiation (2A8 and 7G1-1). 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 Brain E, since the animal used was an FMRP KO, to be conservative, we only selected tags that were also present in WT (Brain D), to avoid confounding data that might result from the absence of FMRP.

Supplementary Figure 6. Comparison of Ago-miRNA CLIP data with published profiles of brain expressed miRNAs.

Generally the Ago-miRNA CLIP set of miRNAs corresponded to previously reported results, although there was a higher correlation between biologic replicates in our

experiments than with previously published measures of miRNA frequency in brain, measured by cloning frequency[12] (A,B,C) or bead-based cytometry[13] (D,E,F) (R^2 = 0.2 - 0.32, higher than Ago-miRNA CLIP in brain vs miRNA frequency in liver (0.06-0.12)). These differences might relate to different developmental ages of brain used for isolation; Landgraf et al analyzed P0 mouse brain, while Lu et al analyzed human brain. In addition, regulation of Ago interaction with miRNA could account for differences between crosslinked populations and whole miRNA populations, as the previously reported approaches used whole small RNA populations for analysis. It is also notable that a significant number of miRNAs identified by Ago HITS-CLIP were not found database derived from direct sequencing ((b); 350 miRNAs present in P13 brain Ago HITS-CLIP that were absent from whole sequencing of miRNAs in P0 brain), suggesting the possibility that this method has greater sensitivity than reported approaches.

Supplementary Figure 7. Biologic complexity and peak height analysis of Ago mRNA clusters.

A. Table of Ago mRNA clusters sorted by biologic complexity and number of tags in the cluster peak (Peak height). B. Table of Ago clusters in genes transcribed in P13 brain, as determined by Affymetrix exon array. C. Plot of data in (A), demonstrating that data for samples with different biologic complexity converge at clusters with peak height $\geq$ 20, suggesting that stringent data analysis can be obtained with peak height $\geq$ 20. D. Comparison of distribution of clusters depending on peak height in *in silico* random CLIP vs. Ago-mRNA CLIP, plotted with different biologic complexity. This allows P values to be determined for selecting further stringent set of clusters in given global threshold (peak height) although those clusters were already normalized by *in silico* random CLIP. E. Effect of varying BC and peak height on Ago-HITS-CLIP map predictions. We evaluated enrichment of seeds (estimated by calculating Ago-miRNA seed sequences (observed vs. expected) in Ago-mRNA footprints from P13 mouse brain transcripts) for the #8 ranking (miR-124) and the #18 ranking (miR-19) miRNAs using the indicated thresholds in BC and peak height. We found that even under our lowest stringency conditions (BC > 2 or peak height > 2), our observed/expected ratios were highly significant. The least stringent assumption (6-mer seeds present in Ago-mRNA clusters with BC > 2) gives a maximum false positive rate of ~30%; the best assumption (conserved 8-mers) gives a false positive rate of ~5%, although

clearly sensitivity is lower at this threshold.

Supplementary Figure 8. Ago mRNA crosslinking sites within transcribed genes.

A. Ago mRNA CLIP tags and clusters according to their position in transcribed genes. The region of transcript (5' UTR, coding sequence (CDS), or 3' UTR) was determined by RefSeq annotation for all brain expressed transcripts (P13 mouse). The number of clusters refers to all clusters with BC $\geq$ 2, and the number of tags refers to the number of tags within the cluster. Total length refers to the aggregate length of all transcribed brain sequences for each given transcribed region. Observed/Expected refers to the ratio of number of clusters or tags per expected number based on total length. P value was calculated by Chi square. Ago mRNA crosslinking sites in 3'UTR are ~1.5 fold enriched comparing to the expected number. B. The percentage of clusters in different regions of transcripts is shown according to their Rank (cluster height, defined as the maximum number of tags in each cluster). All clusters were analyzed, but only the top ranking 100 are shown. C. As in (B), with clusters BC = 5.

Supplementary Figure 9. Generalized correlation of Ago-miRNA and Ago-mRNA CLIP.

A. Analysis as in Supplementary Table 2, but correlating all possible 6 to 8-mer motifs present in a narrowly defined Ago-mRNA footprint region (-24 to + 22 nt from peak, a region in which 100% of clusters are represented). All top ranking motifs are seed sequences from top 8 ranked miRNAs in AGO-miRNA CLIP B. For each miRNA in the database, 3 different six-mers (corresponding to positions 1-8) were compared with sequences in Ago-mRNA cluster peaks (-24 to +22) and analyzed by linear regression model according to the peak height. 22 values with $P < 0.05$ were selected.

Supplementary Figure 10. Ago miRNA ternary map in miRNA targets.

This figure shows the same analysis as in Figure 3 on five of additional 3' UTRs. A-C. Although many miRNA binding sites are predicted in those neuronal genes due to high conservation in 3'UTR, only one Ago-miRNA ternary cluster was identified in each transcript by AGO-HITS-CLIP in P13 mouse brain. D-E. Ctdsp1 and Vamp3 are previously validated targets. These 3' UTRs have a combined total of seven predicted conserved miR-124 sites that showed activity when mutated in luciferase assays, and all seven also had overlapping Ago-miR-124 ternary clusters. Moreover,

in Ctdsp1, constructs with mutations in all previously predicted seed sequences only restored activity to 80% of control[9], a discrepancy that may be accounted for by an additional non-conserved Ago-miR-124 ternary cluster near the stop codon (as suggested for Itgb1, Fig. 3). F-G. Itgb1 and Ptbp1 are shown as in Figure 3, along with rna22 prediction of miRNA binding sites, using less stringent parameters than default (M=13), and with target island prediction (light blue graph) also generated by rna22. H-I. Iqgap1[14] and Sox9[15] are previously identified as miR-124 targets. The fragments used for luciferase assay are shown with black bars. The Ago ternary map identified exact sites of validated miRNA binding within the fragments as shown. J-K. Validated miR-9 targets, Fgfr1[16] and FoxG1[17] are shown. Fgfr1 is validated as miR-9 target in Zebrafish[16].

Supplementary Figure 11. Meta-analysis of five microarray studies in miR-124 overexpression.

A. Meta analysis from 5 experiments as indicated. B. We compiled 1278 brain expressed transcripts from a meta-analysis of five microarray experiments (A) which identified transcripts significantly downregulated by miR-124 overexpression in cell lines (see Supplementary Methods) The performance of Ago-miR-124 ternary cluster predictions was assessed by comparing it with miR-124 seed prediction in two sets – decreased transcripts enriched in direct targets (703 true set, negative fold change) and increased transcripts enriched in indirect targets (575 false set, positive fold change). The true positive rate or specificity was calculated as indicated in black box insets (Ago miR-124 ternary cluster prediction; 73% true positive rate and 92.5% specificity, miR-124 seed prediction; 55% true positive rate and 67.5% specificity). C. Comparison of data from Ago HITS-CLIP data (by gene name and color, as in Table 1), with indicated experiments (microarray[14], SILAC[4], "IP microarray1"[9] and "IP microarray2"[18]).

Supplementary Figure 12. Argonaute HITS-CLIP in HeLa cells transfected with control or miR-124 microRNAs.

Argonaute HITS-CLIP in HeLa cells transfected with control or miR-124 microRNAs. A. ^{32}P-labeled RNA crosslinked to Ago immunoprecipiated with 7G1-1* from miR-124 or control miRNA transfected HeLa cells are shown. Two Ago RNA cross-linked complexes (~130kD for Ago-mRNA and ~110kD for Ago-miRNA) are specific in UV-

irradiated tissue (+XL) compared to non-irradiated brain (–XL). B. The same result as A in a CLIP experiment using 2A8. Also shown is a comparison of overdigestion of Ago RNA cross-linked complex (1:100 RNase A). 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-124 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 <=3). Unique tags are tags with different 5' ends or different degenerative 4 nucleotide barcode introduced in 5' fusion linker.

Supplementary Figure 13. Statistical analysis of data in Figure 2C and Supplementary Table 3.

A. Cumulative frequency of top (N) Ago-miRNA 6-mer seed sequences, compared to the frequency of the bottom (N) Ago-miRNA 6-mer seeds, in Ago footprint regions. Data is derived from Supplementary Table 3 (cumulative sum of 6-mer seeds/family, outlined in pink). B. Cumulative frequency as in (A), but comparing the observed vs. expected seed frequencies for the top 30 Ago-miRNA in Ago footprint regions. Both datasets suggest that the correlation between the top 20 miRNAs and Ago-mRNAs is robust as a group, and that correlations likely extend out significantly further. In addition, the shape of the cumulative curve (e.g. (B)) suggests variability among individual miRNA ternary maps. This can be evaluated in more detail in Supplemental Table 3. C-D. Reanalysis of the data in A-B, respectively, but using conservation of conserved seed sequenced in calculation of cumulative frequencies. E. Fraction of total p13 brain transcripts (blue line) or transcripts with Ago-miR-124 cluster (red line) according to their expression indicated as $\log_2$ ratio (normalized probe intensity vs median probe intensity (771)), showing that low abundance transcripts are recovered less efficiently by Ago HITS-CLIP. This analysis is also discussed in Supplementary Methods: *Ago miRNA ternary map; predicting miRNA sites in the peak Ago footprint.*

Supplementary Figure 14. AGO miRNA ternary map.

A. Table representing number of transcripts with given number of Ago-mRNA clusters (BC $\geq$ 2) in 3'UTR, which shows average ~2 clusters per transcript in 3'UTR. The average number of clusters/transcript, Ago-regulated transcript, or 3' UTR of Ago-regulated transcripts are shown. B. Number of transcripts predicted as each top 20 Ago-miRNA targets (collapsed by family; see "pink" lines in Supplementary Table 3), based on Ago miRNA ternary map. C. Ago miRNA ternary map of top 20 Ago-miRNAs in mouse genome.

Supplementary Figure 15. Gene ontology analysis of AGO mRNA ternary map.

This figure shows that target transcripts for each of the Ago top 20 miRNA ternary complexes (see Supplementary Fig. 14b) that were used for gene ontology analysis. A. For each category, enrichment was compared with all P13 brain expressed transcripts using GoMiner (http://discover.nci.nih.gov/gominer/). False discovery rate (FDR) was calculated for each GO category and is represented as a different color as indicated. Hierarchical clustering of miRNAs and GO categories was performed using Cluster program and visualized as heat map and tree by Treeview (http://rana.lbl.gov/EisenSoftware.htm). GO categories were divided into two groups, one with little or no miRNA enrichment, and a second with a large amount of miRNA enrichment. The second group is shown here. B. As in (A), but using targets identified by TargetScan 4.1 rather than Ago HITS-CLIP, are shown for comparison. C-D. GO analysis using DAVID (http://david.abcc.ncifcrf.gov/) illustrating different results of the predicted actin cytoskeleton miRNA regulatory map for miR-124, miR-9 and miR-125 that are obtained with Ago HITS-CLIP and TargetScan 4.1. Taken together, these figures illustrate that both the FDR rate and evidence for a protein network deteriorate substantially when Ago-mRNA tags are not used.

In summary, the Ago-RNA ternary map corresponds remarkably well with the current view of microRNA function in the brain. For example, the current view of miR-124, miR-125[19] and miR-9 biology (Figure 5), including actions to promote neurite outgrowth and differentiation by inhibiting Ago-miR-124 targets such as *Itgb1*[20], *Iqgap1*[14], *Sox9*[15] and *Ptbp1*[21], or Ago-miR-9 targets such as *Fgfr1*[16] and *Foxg1*[17] (Fig. 5c; Supplemental Fig. 10), are closely reflected in the Ago HITS-CLIP results.

Supplementary Figure 16. Simulations estimating the relationship between depth of HITS-CLIP sequencing and number of clusters/tags identified.

A. To estimate whether our sequencing depth was sufficiently saturated to give a global view of Ago HITS-CLIP targets, we undertook a simulation to see the effects of the last 10% of sequencing on our final Ago ternary cluster map. 10% of sequenced tags were removed randomly from each of 5 different AGO mRNA CLIP results, then those tags were processed by in silico normalization, alignment of tags using BLAT and cubic spline interpolation analysis. We then performed the simulation 10 times and compared the number of Ago ternary clusters (BC>=2) between the original set and the simulated one, focusing on the top 20 miRNAs clusters which we used for the Ago maps. The recovery rate of these simulation was ~92% for the Ago top 20 miRNA clusters (14411/15665) and ~95% for miR-124 (1487/1561), which are higher than the expected rate (90%). This increase of recovery rate was statistically significant (top 20; 1.02 = 92/90, $P=2.08 \times 10^{-11}$ and for miR-124; 1.06 = 95/90, $P=2.55 \times 10^{-7}$; Chi square test).

B. The simulation in (A) was repeated as a function of tags in the simulation. This was accomplished by repeating the simulation removing a range of tags, from 10-90%. We estimated the recovery of Ago miR-124 clusters (blue) and all top 20 Ago-miRNA-mRNA clusters as indicated. These results indicate that our sequencing depth is near the saturation point and is sufficient to show global view of miRNA target sites for the top 20 miRNAs. For example, even after removing 50% of tags, we had 88% recovery of the miR-124 clusters.

Supplementary References:

1. Nelson, P. T. et al. A novel monoclonal antibody against human Argonaute proteins reveals unexpected characteristics of miRNAs in human blood cells. *RNA* **13**, 1787-1792 (2007).
2. Brown, V. et al. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in Fragile X Syndrome. *Cell* **107**, 477-87. (2001).
3. Miranda, K. C. et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203-1217 (2006).
4. Baek, D. et al. The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).
5. Selbach, M. et al. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58-63 (2008).
6. Mourelatos, Z. Small RNAs: The seeds of silence. *Nature* **455**, 44-45 (2008).
7. Hammell, M. et al. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat Methods* **5**, 813-819 (2008).
8. Easow, G., Teleman, A. A. & Cohen, S. M. Isolation of microRNA targets by miRNP immunopurification. *RNA* **13**, 1198-1204 (2007).
9. Karginov, F. V. et al. A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci U S A* **104**, 19291-19296 (2007).
10. Licatalosi, D. D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-469 (2008).
11. Foat, B. C., Houshmandi, S. S., Olivas, W. M. & Bussemaker, H. J. Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *Proc Natl Acad Sci U S A* **102**, 17675-17680 (2005).
12. Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401-1414 (2007).
13. Lu, J. et al. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838 (2005).

14. Lim, L. P. et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769-773 (2005).

15. Cheng, L. C., Pastrana, E., Tavazoie, M. & Doetsch, F. miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. *Nat Neurosci* **12**, 399-408 (2009).

16. Leucht, C. et al. MicroRNA-9 directs late organizer activity of the midbrain-hindbrain boundary. *Nat Neurosci* **11**, 641-648 (2008).

17. Shibata, M., Kurokawa, D., Nakao, H., Ohmura, T. & Aizawa, S. MicroRNA-9 modulates Cajal-Retzius cell differentiation by suppressing Foxg1 expression in mouse medial pallium. *J Neurosci* **28**, 10415-10421 (2008).

18. Hendrickson, D. G., Hogan, D. J., Herschlag, D., Ferrell, J. E. & Brown, P. O. Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. *PLoS ONE* **3**, e2126 (2008).

19. Wu, L. & Belasco, J. G. Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. *Mol Cell Biol* **25**, 9198-9208 (2005).

20. Yu, J. Y., Chung, K. H., Deo, M., Thompson, R. C. & Turner, D. L. MicroRNA miR-124 regulates neurite outgrowth during neuronal differentiation. *Exp Cell Res* **314**, 2618-2633 (2008).

21. Makeyev, E. V., Zhang, J., Carrasco, M. A. & Maniatis, T. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol Cell* **27**, 435-448 (2007).