SUPPLEMENT to

Ago HITS-CLIP decodes miRNA-mRNA interaction maps.

Sung Wook Chi, Julie B. Zang, Aldo Mele, Robert B. Darnell

Laboratory of Molecular Neuro-Oncology and Howard Hughes Medical Institute, The Rockefeller University, 1230 York Ave, NY, NY 10021 USA

Supplementary Methods

Ago CLIP

CLIP was performed as described[1-3] with modifications in a detailed method present below and updated on the project website (http://ago.rockefeller.edu), using a monoclonal Ago antibody, 2A8, generously supplied by Z. Mourelatos[4]. 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[5]. 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.

Ago CLIP in miR-124 transfected HeLa cell.

HeLa cells were transfected using Lipofectamine 2000 (InVitrogen) in 100 mm^2 plates with 75 nM RNA duplexes, miR-124 and control miRNA from miRIDIAN miRNA Mimic (Dharmacon). After 24hr, cells were harvested and CLIP

experiments were performed in two biological replicates using either 2A8 or 7G1-1 antibodies. Degenerate barcodes (4 nucleotides tags) were introduced in the 5' fusion linker to increase the complexity in unique tags and to avoid artifacts from preferential PCR duplication.

Exon arrays in P13 mouse brain.

For estimating the level of brain transcripts, total RNAs from three P13 mouse brains were extracted using Trizol (Invitrogen) and RNAeasy kit (Qiagen) 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, median 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 and data have been deposited in GEO database (GSE16338)

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[6] (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 ~3-fold improvement in false positives; 62nt Ago footprint region vs 200nt cluster window).

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$\geq$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[7,8]. They were estimated as ~40%[9] or ~66%(based on the reports that two-thirds of the predicted targets appeared to be nonresponsive to miR-223 loss in neutrophils[7].) This higher number of false positives is also in agreement with findings from two other approaches using purification of miRNA-protein complexes (>50%[10] or ~70%[11]; 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$\geq$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 6 and Supplementary Figure 15).

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[12]. 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 (Figure 5). 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 Figure 5, 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 Figure 5 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 Figure 5, 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[7,8,11]. 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. 6a).

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 Figure 5), revealing a high sensitivity of our map with BC $\geq$ 2, independent of peak height. This also reflects our experience with Nova HITS-CLIP[2], 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 (Figure 5), 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/) or DAVID (http://david.abcc.ncifcrf.gov/). Details are given in Supplementary Fig. 15.

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/[13].

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. 7G1-1* 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)[5] 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[2] 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 the number of CLIP tags from transient nonspecific protein-RNA interactions are correlated with transcript abundance, then each transcript should have different threshold levels to differentiate signal:noise. Therefore we estimated the false discovery rate (FDR) of CLIP tag clusters for each brain transcript using a simulation method termed *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 is $F_n = N_n \times Y_n / 50$. 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 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. *In silico* random CLIP simulations. We used normalized probe intensity values from microarray experiments as the number of transcripts (N) in the simulation as an estimate of the real number of transcripts in CLIP experiments. Although the expected number of tags in any given transcript (T_n) is calculated as ratio and not affected by N, we checked the effect of varying (N) on the outcome of *in silico CLIP*. For this, simulation tests were done with Chad7, a transcript that has the minimum abundance used in our experiments and in *in silico* random CLIP. The top table shows that Tn does not change as a function of abundance (N_n). The first column values indicate that the array value of every transcript was increased together with Chad7. All simulations have 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 (M_n). 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 indicates 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[14] (A,B,C) or bead-based cytometry[15] (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[12], 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[16] and Sox9[17] 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[18] and FoxG1[19] are shown. Fgfr1 is validated as miR-9 target in Zebrafish[18].

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 Figure 5), with indicated experiments (microarray[16], SILAC[7], "IP microarray1"[12] and "IP microarray2"[20]).

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[21] and miR-9 biology (Figure 6), including actions to promote neurite outgrowth and differentiation by inhibiting Ago-miR-124 targets such as *Itgb1*[22], *Iqgap1*[16], *Sox9*[17] and *Ptbp1*[23], or Ago-miR-9 targets such as *Fgfr1*[18] and *Foxg1*[19] (Fig. 6c; 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.

Supplemental Table 1

A **From 130kD**

Sample	Antibody	Age	Raw tags	BLAT (mm8)	% align	Unique tags
Brain A	2A8	P13	5,057,657	4,168,244	**82.4**	237,824
Brain B	2A8	P13	4,291,392	3,504,911	**81.7**	359,405
Brain C	2A8	P13	3,843,297	3,073,034	**80**	328,916
Brain D	7G1-1	P13	6,259,297	5,118,928	**81.8**	420,627
Brain E	7G1-1	P13	6,394,071	5,188,918	**81.2**	423,024

B **From 110kD**

Sample	Antibody	Age	Raw tags	BLAST (miRBASE)	% align
Brain A	2A8	P13	4,918,586	3,717,702	**75.6**
Brain B	2A8	P13	4,160,480	3,004,104	**72.2**
Brain C	2A8	P13	3,490,474	2,576,516	**73.8**
Brain D	7G1-1	P13	4,853,688	4,275,690	**88.1**
Brain E	7G1-1	P13	5,409,367	4,765,957	**88.1**

Supplemental Table 2

A

AGO-mRNA CLIP				AGO-miRNA CLIP			
Top ranking motif	Regression coefficient	Number of tags	P-value	miRNA	Position of motif	Percentage in Ago-miRNA CLIP tags	Rank
TGCCTT	0.521	9153	8.3×10^{-58}	mir-124	2-7	2.97%	8
TACCTC	0.512	6403	1.6×10^{-40}	let-7	2-7	6.24%	5
CCAAAG	0.393	6878	3.1×10^{-32}	mir-9	2-7	9.33%	2
CTGTGA	0.388	4205	5.0×10^{-23}	mir-27	2-7	4.80%	6
TGTTTA	0.436	3206	7.9×10^{-21}	mir-30	3-8	34.01%	1
CTTGAA	0.356	2977	1.5×10^{-14}	mir-26	1-6	6.39%	4

B

AGO-miRNA CLIP					AGO-mRNA CLIP		
Rank	7mer seed motif	miRNA	Number of Ago-miRNA CLIP tags (Total)	%	7mer seed obs/exp	Number of mRNA tags with seed (obs)	Number of clusters with seed
1	TGTTTAC	miR-30	8,417,151	34.29	3.304	1521	172
2	ACCAAAG	miR-9	2,309,955	9.41	2.175	1676	211
3	TGAATGT	miR-181	2,050,154	8.35	1.549	862	132
4	TACTTGA	miR-26	1,581,514	6.44	2.436	687	68
5	CTACCTC	let-7	1,549,064	6.31	4.000	2636	268
6	ACTGTGA	miR-27	1,188,538	4.84	1.190	969	137
7	AGCTCCT	miR-708	1,172,287	4.78	0.202	83	12
8	GTGCCTT	miR-124	735,430	3.00	5.187	3362	294
9	CACTGCC	miR-34	333,089	1.36	0.312	147	30
10	CTCAGGG	miR-125	330,186	1.35	1.187	506	65

Supplemental Table 3

				All seeds				Conserved seeds				
Rank	Frequency in Ago miRNA CLIP	miRNA family	Seed motif	Obs/Exp	P-value	Observed Frequeny in Ago footprint	Expected frequency	Obs/Exp	P-value	Observed Frequency in Ago Footprint	Expected frequency	miRNAs
1	8417151	miR-30	TGTTTACA	7.002	1.99E-175	155	22	13.332	3.78E-297	119	9	miR-30c:miR-30b:miR-30a:miR-30e:miR-30d:miR-384-5p
1	8417151	miR-30	TGTTTAC	4.931	6.21E-167	242	49	10.458	5.25E-307	164	16	miR-30c:miR-30b:miR-30a:miR-30e:miR-30d:miR-384-5p
1	8417151	miR-30	GTTTACA	4.273	3.10E-113	204	48	8.735	2.10E-228	152	17	miR-30c:miR-30b:miR-30a:miR-30e:miR-30d:miR-384-5p
1	8417151	miR-30	GTTTAC	3.085	5.55E-109	349	113	6.889	7.48E-255	231	34	miR-30c:miR-30b:miR-30a:miR-30e:miR-30d:miR-384-5p
1	8417151	miR-30	TGTTTA	2.055	6.92E-52	424	206	4.772	7.35E-161	245	51	miR-30c:miR-30b:miR-30a:miR-30e:miR-30d:miR-384-5p
1	8417151	miR-30	TTTACA	1.905	3.52E-39	399	209	4.237	2.16E-128	235	55	miR-30c:miR-30b:miR-30a:miR-30e:miR-30d:miR-384-5p
1		miR-30	with all 6mers	2.216	4.53E-172	1172	529	5.066	0.00E+00	711	140	
2	2309955	miR-9	ACCAAAGA	6.381	8.87E-170	170	27	11.190	4.66E-234	115	10	miR-9
2	2309955	miR-9	CCAAAGA	4.317	1.58E-252	452	105	8.223	0.00E+00	299	36	miR-9
2	2309955	miR-9	ACCAAAG	4.107	1.23E-166	322	78	7.773	8.79E-264	204	26	miR-9
2	2325768	miR-9	ACCAAA	3.006	2.70E-209	712	237	5.833	0.00E+00	398	68	miR-9:miR-133a:miR-133b
2	2309955	miR-9	CCAAAG	2.847	8.84E-243	924	325	5.604	0.00E+00	548	98	miR-9:miR-1897-5p
2	2309955	miR-9	CAAAGA	2.013	5.44E-85	749	372	4.167	7.66E-233	441	106	miR-9
2		miR-9	with all 6mers	2.555	0.00E+00	2385	934	5.102	0.00E+00	1387	272	
3	2050154	miR-181	TGAATGTT	3.044	3.84E-15	45	15	8.177	4.00E-36	25	3	miR-181d:miR-181a:miR-181c:miR-181b
3	2050281	miR-181	GAATGTT	1.879	9.63E-10	91	48	4.105	3.26E-22	40	10	miR-181d:miR-181a:miR-181c:miR-181b:miR-543
3	2050154	miR-181	TGAATGT	2.966	1.86E-65	224	76	6.383	1.82E-121	121	19	miR-181d:miR-181a:miR-181c:miR-181b
3	2050281	miR-181	GAATGT	1.903	7.36E-41	418	220	3.688	3.21E-87	200	54	miR-181d:miR-181a:miR-181c:miR-181b:miR-543
3	2050281	miR-181	AATGTT	1.541	1.62E-15	334	217	2.803	3.56E-36	136	49	miR-181d:miR-181a:miR-181c:miR-181b:miR-543
3	2050154	miR-181	TGAATG	1.783	1.28E-34	438	246	3.388	1.79E-85	228	67	miR-181d:miR-181a:miR-181c:miR-181b
3		miR-181	with all 6mers	1.745	2.77E-84	1190	682	3.317	1.59E-200	564	170	
4	1581514	miR-26	TACTTGAA	6.130	2.42E-81	85	14	9.901	5.21E-89	50	5	miR-26a:miR-26b
4	1581514	miR-26	TACTTGA	3.600	8.61E-54	127	35	6.661	6.26E-69	64	10	miR-26a:miR-26b
4	1581514	miR-26	ACTTGAA	2.754	6.80E-46	181	66	6.260	8.00E-106	108	17	miR-26a:miR-26b
4	1581514	miR-26	ACTTGA	1.920	4.13E-35	347	181	4.351	4.96E-94	164	38	miR-26a:miR-26b
4	1581514	miR-26	CTTGAA	1.811	3.12E-34	410	226	4.152	3.65E-108	204	49	miR-26a:miR-26b
4	1581514	miR-26	TACTTG	1.959	1.85E-28	261	133	3.876	5.40E-46	95	25	miR-26a:miR-26b
4		miR-26	with all 6mers	1.884	9.23E-94	1018	540	4.158	1.61E-243	463	111	
5	1530176	let-7	CTACCTCA	9.250	5.66E-306	190	21	14.866	0.00E+00	140	9	miR-98:let-7c:let-7b:let-7a:let-7g:let-7f:let-7e:let-7i
5	1549064	let-7	CTACCTC	6.118	0.00E+00	364	59	11.358	0.00E+00	255	22	miR-98:let-7c:let-7b:let-7a:let-7g:let-7f:let-7e:let-7d:let-7i
5	1530176	let-7	TACCTCA	6.627	0.00E+00	370	56	12.152	0.00E+00	273	22	miR-98:let-7c:let-7b:let-7a:let-7g:let-7f:let-7e:let-7i
5	1551121	let-7	CTACCT	2.501	5.76E-108	541	216	4.987	9.10E-222	317	64	miR-98:let-7c:let-7b:let-7a:let-7g:let-7f:let-7e:let-7d:let-7i:miR-196a:miR-196b
5	1549232	let-7	TACCTC	4.479	0.00E+00	706	158	9.295	0.00E+00	467	50	miR-98:let-7c:let-7b:let-7a:let-7g:let-7f:let-7e:let-7d:miR-202-3p:let-7i
5	1533600	let-7	ACCTCA	2.755	6.59E-191	777	282	5.674	0.00E+00	453	80	miR-98:let-7c:let-7b:let-7a:let-7g:let-7f:let-7e:let-7i:miR-672
5		let-7	with all 6mers	3.085	0.00E+00	2024	656	6.388	0.00E+00	1237	194	
6	1188538	miR-27	ACTGTGAA	2.990	2.69E-26	85	28	6.255	5.76E-49	49	8	miR-27b:miR-27a
6	1211720	miR-27	ACTGTGA	2.525	1.31E-48	233	92	4.613	9.24E-81	128	28	miR-128:miR-27b:miR-27a
6	1188538	miR-27	CTGTGAA	1.753	2.78E-15	193	110	3.622	4.22E-41	95	26	miR-27b:miR-27a
6	1219457	miR-27	CTGTGA	1.559	1.21E-28	615	394	2.850	1.98E-77	289	101	miR-673-5p:miR-128:miR-27b:miR-27a
6	1211720	miR-27	ACTGTG	1.879	8.57E-61	658	350	3.865	1.94E-136	291	75	miR-128:miR-27b:miR-27a
6	1189726	miR-27	TGTGAA	1.480	1.74E-18	495	335	2.553	2.14E-50	236	92	miR-27b:miR-27a:miR-350
6		miR-27	with all 6mers	1.638	1.31E-97	1768	1079	3.032	1.35E-243	816	269	
7	1172287	miR-708	AGCTCCTT	0.839	4.68E-01	17	20	1.391	5.66E-01	3	2	miR-708:miR-28
7	1172287	miR-708	GCTCCTT	0.718	1.99E-02	49	68	1.287	3.80E-01	12	9	miR-708:miR-28
7	1172287	miR-708	AGCTCCT	0.562	1.44E-05	55	98	0.713	2.82E-01	10	14	miR-708:miR-28
7	1172287	miR-708	CTCCTT	0.934	2.55E-01	275	295	1.497	7.46E-04	69	46	miR-708:miR-28
7	1172287	miR-708	AGCTCC	0.582	2.19E-13	179	308	0.925	5.85E-01	49	53	miR-708:miR-28
7	1172287	miR-708	GCTCCT	0.574	1.52E-14	187	326	0.813	1.68E-01	44	54	miR-708:miR-28
7		miR-708	with all 6mers	0.691	4.55E-21	641	928	1.057	4.78E-01	162	153	
8	735430	miR-124	GTGCCTTA	12.685	1.48E-294	125	10	20.229	0.00E+00	82	4	miR-124
8	735430	miR-124	TGCCTTA	7.277	0.00E+00	324	45	13.758	0.00E+00	212	15	miR-124
8	735430	miR-124	GTGCCTT	6.299	0.00E+00	415	66	12.725	0.00E+00	281	22	miR-124
8	735430	miR-124	GCCTTA	4.358	2.02E-302	534	123	9.744	0.00E+00	335	34	miR-124
8	735430	miR-124	TGCCTT	3.155	0.00E+00	1020	323	7.489	0.00E+00	605	81	miR-124
8	735430	miR-124	GTGCCT	2.877	1.32E-198	738	256	6.866	0.00E+00	453	66	miR-124
8		miR-124	with all 6mers	3.263	0.00E+00	2292	702	7.690	0.00E+00	1393	181	
9	270706	miR-34	CACTGCCA	0.615	2.81E-02	20	33	1.344	4.01E-01	8	6	miR-34a:miR-449a
9	333089	miR-34	CACTGCC	0.879	2.23E-01	89	101	1.807	1.23E-03	29	16	miR-699:miR-34c:miR-34a:miR-34b-5p:miR-449b:miR-449c:miR-449a
9	270706	miR-34	ACTGCCA	0.948	6.36E-01	77	81	1.931	9.35E-05	34	18	miR-34a:miR-449a
9	333089	miR-34	ACTGCC	1.024	7.03E-01	259	253	2.062	3.50E-14	105	51	miR-699:miR-34c:miR-34a:miR-34b-5p:miR-449b:miR-449c:miR-449a
9	333089	miR-34	CACTGC	0.857	1.34E-02	257	300	1.595	3.13E-05	78	49	miR-699:miR-34c:miR-34a:miR-34b-5p:miR-449b:miR-449c:miR-449a
9	270706	miR-34	CTGCCA	0.930	1.58E-01	379	408	1.602	1.00E-08	145	90	miR-34a:miR-449a
9		miR-34	with all 6mers	0.932	3.51E-02	895	960	1.724	1.83E-23	328	190	
10	330186	miR-125	CTCAGGGA	2.365	2.79E-12	62	26	7.292	1.60E-51	42	6	miR-125b-5p:miR-351:miR-125a-5p
10	332795	miR-125	TCAGGGA	1.459	7.81E-05	108	74	4.251	1.75E-36	64	15	miR-125b-5p:miR-351:miR-125a-5p:miR-670
10	330186	miR-125	CTCAGGG	1.415	1.72E-04	116	82	4.476	4.74E-40	65	15	miR-125b-5p:miR-351:miR-125a-5p
10	333749	miR-125	CAGGGA	0.738	2.06E-06	242	328	2.025	9.18E-14	107	53	miR-125b-5p:miR-351:miR-339-5p:miR-125a-5p:miR-670
10	332795	miR-125	TCAGGG	0.951	4.61E-01	219	230	2.807	1.21E-28	106	38	miR-125b-5p:miR-351:miR-125a-5p:miR-670
10	330186	miR-125	CTCAGG	1.055	3.17E-01	351	333	2.745	3.38E-42	167	61	miR-125b-5p:miR-351:miR-125a-5p
10		miR-125	with all 6mers	0.911	8.20E-03	812	891	2.509	5.17E-77	380	151	

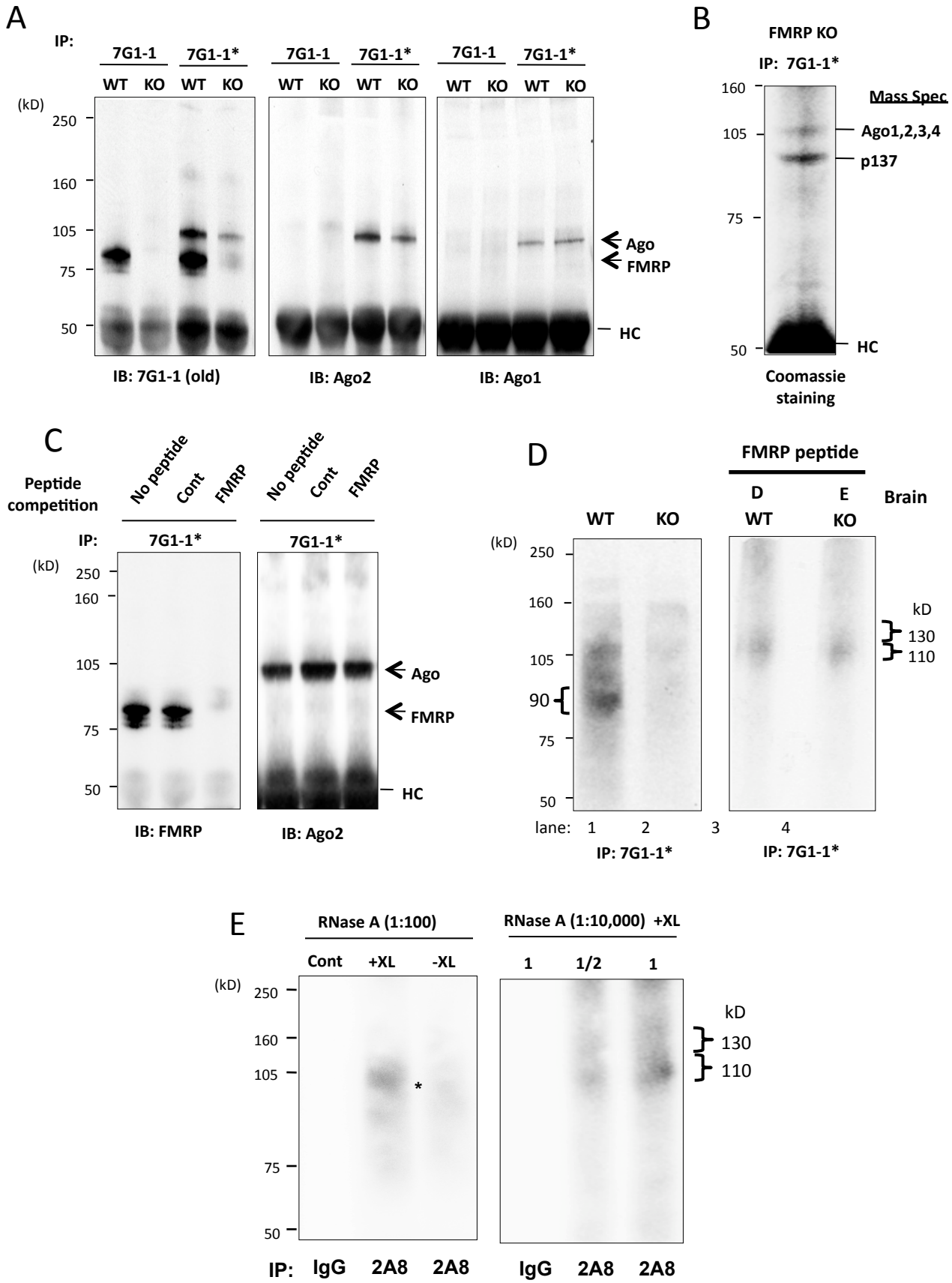
Supplemental Table 3 (continued)

			All seeds					Conserved seeds				
Rank	Frequency in Ago miRNA CLIP	miRNA family	Seed motif	Obs/Exp	P-value	Observed Frequeny in Ago footprint	Expected frequency	Obs/Exp	P-value	Observed Frequency in Ago Footprint	Expected frequency	miRNAs
11	307064	miR-193	GGCCAGTT	0.894	6.99E-01	12	13	1.242	7.07E-01	3	2	miR-193:miR-193b
11	307064	miR-193	GGCCAGT	0.669	6.62E-03	45	67	1.561	5.04E-02	19	12	miR-193:miR-193b
11	307064	miR-193	GCCAGTT	0.906	5.12E-01	44	49	0.830	5.98E-01	8	10	miR-193:miR-193b
11	330471	miR-193	GGCCAG	0.468	4.85E-25	177	378	0.819	9.66E-02	69	84	miR-328:miR-193:miR-193b
11	307064	miR-193	GCCAGT	0.814	3.64E-03	198	243	1.298	3.67E-02	64	49	miR-193:miR-193b
11	307064	miR-193	CCAGTT	0.893	1.06E-01	203	227	1.114	4.27E-01	54	48	miR-193:miR-193b
11		miR-193	with all 6mers	0.681	1.51E-20	578	849	1.027	7.13E-01	187	182	
12	138054	miR-17	GCACTTTG	1.311	2.14E-01	21	16	2.760	4.42E-04	11	4	miR-17:miR-20b:miR-93:miR-106a
12	297690	miR-17	GCACTTT	1.815	4.51E-09	94	52	3.806	1.22E-30	64	17	miR-17:miR-292-3p:miR-291b-3p:miR-295:miR-294:miR-20b:miR-20a:miR-93 :miR-106a:miR-106b:miR-290-3p:miR-291a-3p
12	138054	miR-17	CACTTTG	1.001	9.91E-01	66	66	1.845	2.30E-03	24	13	miR-17:miR-20b:miR-93:miR-106a
12	306438	miR-17	GCACTT	1.306	7.00E-05	221	169	2.784	3.44E-30	114	41	miR-17:miR-292-3p:miR-291b-3p:miR-295:miR-294:miR-467b:miR-467c:miR-20b:miR-20a:miR-93:miR-105:miR-302b:miR-302c:miR-302a:miR-302d:miR-467d:miR-106a:miR-106b:miR-467a:miR-290-3p:miR-291a-3p
12	297690	miR-17	CACTTT	1.233	6.16E-04	267	217	2.834	3.62E-35	129	46	miR-17:miR-292-3p:miR-291b-3p:miR-295:miR-294:miR-20b:miR-20a:miR-93 :miR-106a:miR-106b:miR-290-3p:miR-291a-3p
12	138054	miR-17	ACTTTG	1.022	7.14E-01	282	276	1.628	7.01E-08	120	74	miR-17:miR-20b:miR-93:miR-106a
12		miR-17	with all 6mers	1.164	2.59E-05	770	662	2.266	8.46E-58	363	160	
13	277918	miR-344	CTAGATCA	1.522	2.63E-01	7	5	1.928	5.04E-01	1	1	miR-344
13	277918	miR-344	CTAGATC	1.312	2.02E-01	22	17	2.873	2.71E-02	4	1	miR-344
13	277918	miR-344	TAGATCA	1.452	4.72E-02	28	19	2.201	3.22E-02	7	3	miR-344
13	278382	miR-344	CTAGAT	1.323	2.83E-03	113	85	2.714	3.46E-07	24	9	miR-878-5p:miR-344
13	277918	miR-344	TAGATC	1.308	1.36E-02	84	64	2.456	5.33E-06	24	10	miR-344
13	277918	miR-344	AGATCA	0.856	2.99E-02	194	227	1.041	7.41E-01	69	66	miR-344
13		miR-344	with all 6mers	1.039	4.49E-01	391	376	1.378	4.99E-04	117	85	
14	273527	miR-138	CACCAGCT	0.745	1.48E-01	24	32	1.954	1.07E-02	14	7	miR-138
14	273527	miR-138	CACCAGC	0.807	3.51E-02	96	119	1.837	3.84E-06	56	30	miR-138
14	273527	miR-138	ACCAGCT	0.821	7.92E-02	79	96	1.439	2.41E-02	38	26	miR-138
14	274185	miR-138	CCAGCT	0.660	6.00E-13	295	447	1.022	8.17E-01	111	109	miR-763:miR-138
14	273527	miR-138	ACCAGC	0.870	1.91E-02	282	324	1.621	1.41E-09	154	95	miR-138
14	273527	miR-138	CACCAG	0.699	1.66E-08	245	351	1.166	1.27E-01	99	85	miR-138
14		miR-138	with all 6mers	0.732	3.21E-19	822	1122	1.262	8.82E-06	364	289	
15	213825	miR-15	TGCTGCTA	3.649	4.53E-34	77	21	8.520	1.10E-59	40	5	miR-15b:miR-15a:miR-16:miR-195
15	250252	miR-15	TGCTGCT	1.724	1.77E-24	343	199	2.862	2.38E-40	146	51	miR-15b:miR-15a:miR-16:miR-322:miR-195:miR-497:miR-1907:miR-103:miR-107
15	214894	miR-15	GCTGCTA	2.072	4.41E-16	119	57	4.666	1.37E-37	57	12	miR-15b:miR-15a:miR-16:miR-195:miR-503
15	251321	miR-15	GCTGCT	1.080	6.14E-02	597	553	1.807	2.19E-22	263	146	miR-15b:miR-15a:miR-16:miR-322:miR-195:miR-497:miR-1907:miR-503:miR-103:miR-107
15	250938	miR-15	TGCTGC	1.275	6.23E-10	644	505	2.060	2.15E-36	291	141	miR-15b:miR-15a:miR-16:miR-322:miR-195:miR-497:miR-1907:miR-1906:miR-761:miR-103:miR-107
15	214894	miR-15	CTGCTA	1.485	3.03E-11	279	188	2.939	1.17E-29	100	34	miR-15b:miR-15a:miR-16:miR-195:miR-503
15		miR-15	with all 6mers	1.220	8.23E-15	1520	1246	2.038	3.33E-77	654	321	
16	179104	miR-101	GTACTGTA	3.500	6.15E-13	29	8	8.574	7.12E-34	22	3	miR-101a:miR-101b
16	195599	miR-101	TACTGTA	2.869	1.40E-32	116	40	6.778	1.86E-69	63	9	miR-101a:miR-101b:miR-144
16	179104	miR-101	GTACTGT	1.948	6.54E-09	73	37	4.637	1.64E-23	35	8	miR-101a:miR-101b
16	208225	miR-101	ACTGTA	1.901	3.54E-33	337	177	4.463	2.77E-80	134	30	miR-101a:miR-101b:miR-139-5p:miR-144:miR-582-5p
16	198109	miR-101	TACTGT	2.136	2.67E-53	391	183	4.269	4.21E-89	160	37	miR-101a:miR-101b:miR-144:miR-199b:miR-199a-3p
16	179104	miR-101	GTACTG	1.363	1.57E-05	193	142	2.809	5.25E-20	72	26	miR-101a:miR-101b
16		miR-101	with all 6mers	1.835	4.42E-78	921	502	3.930	7.24E-176	366	93	
17	207067	miR-153	CTATGCAA	3.011	1.72E-09	27	9	6.289	4.02E-21	20	3	miR-153
17	208220	miR-153	TATGCAA	2.421	7.07E-18	89	37	4.775	1.57E-39	58	12	miR-153:miR-448
17	207067	miR-153	CTATGCA	1.789	4.60E-07	73	41	5.233	2.23E-38	49	9	miR-153
17	208220	miR-153	TATGCA	1.604	4.23E-13	231	144	3.146	2.05E-40	121	38	miR-153:miR-448:miR-669j:miR-669k:miR-669i:miR-669h-3p
17	208220	miR-153	ATGCAA	1.521	1.99E-11	252	166	3.043	5.62E-41	131	43	miR-153:miR-448
17	207067	miR-153	CTATGC	1.407	2.01E-06	192	136	3.171	3.08E-32	94	30	miR-153
17		miR-153	with all 6mers	1.513	2.41E-27	675	446	3.113	6.39E-110	346	111	
18	153999	miR-19	TTTGCACA	1.969	2.67E-05	37	19	4.159	2.81E-15	26	6	miR-19b:miR-19a
18	153999	miR-19	TTTGCAC	1.977	1.63E-11	94	48	4.052	7.39E-33	62	15	miR-19b:miR-19a
18	153999	miR-19	TTGCACA	1.542	1.43E-04	76	49	3.869	1.52E-23	47	12	miR-19b:miR-19a
18	201896	miR-19	TTGCAC	1.669	2.02E-14	218	131	3.937	5.46E-62	126	32	miR-301b:miR-301a:miR-19b:miR-19a:miR-130a:miR-130b:miR-721
18	153999	miR-19	TTTGCA	1.161	1.25E-02	278	239	2.051	2.86E-18	141	69	miR-19b:miR-19a
18	155631	miR-19	TGCACA	0.929	2.52E-01	242	260	1.801	6.66E-10	107	59	miR-466h:miR-466j:miR-19b:miR-19a
18		miR-19	with all 6mers	1.171	1.84E-05	738	630	2.335	4.62E-64	374	160	
19	194366	miR-21	ATAAGCTA	1.956	2.34E-02	11	6	3.096	3.91E-02	3	1	miR-21
19	194811	miR-21	ATAAGCT	1.429	2.51E-02	39	27	3.144	1.31E-05	13	4	miR-590-5p:miR-21
19	194366	miR-21	TAAGCTA	1.559	1.79E-02	28	18	4.211	7.50E-07	10	2	miR-21
19	194811	miR-21	ATAAGC	1.291	5.62E-03	117	91	2.907	2.86E-12	39	13	miR-590-5p:miR-21
19	194811	miR-21	TAAGCT	1.349	3.78E-04	140	104	2.695	1.95E-10	38	14	miR-590-5p:miR-21
19	194366	miR-21	AAGCTA	1.136	9.14E-02	175	154	2.485	1.09E-11	52	21	miR-21
19		miR-21	with all 6mers	1.240	7.65E-06	432	348	2.663	5.49E-31	129	48	
20	155769	miR-221	ATGTAGCT	1.317	3.02E-01	14	11	3.856	3.63E-03	4	1	miR-221:miR-222
20	155769	miR-221	TGTAGCT	1.257	8.33E-02	57	45	2.913	5.32E-05	13	4	miR-221:miR-222
20	155769	miR-221	ATGTAGC	1.153	3.92E-01	36	31	2.729	1.61E-04	13	5	miR-221:miR-222
20	155769	miR-221	ATGTAG	1.182	4.42E-02	145	123	2.583	1.44E-11	47	18	miR-221:miR-222
20	155769	miR-221	TGTAGC	1.143	9.79E-02	153	134	2.424	2.02E-09	43	18	miR-221:miR-222
20	155769	miR-221	GTAGCT	1.000	9.99E-01	116	116	2.311	1.58E-06	31	13	miR-221:miR-222
20		miR-221	with all 6mers	1.111	3.19E-02	414	373	2.452	2.01E-24	121	49	

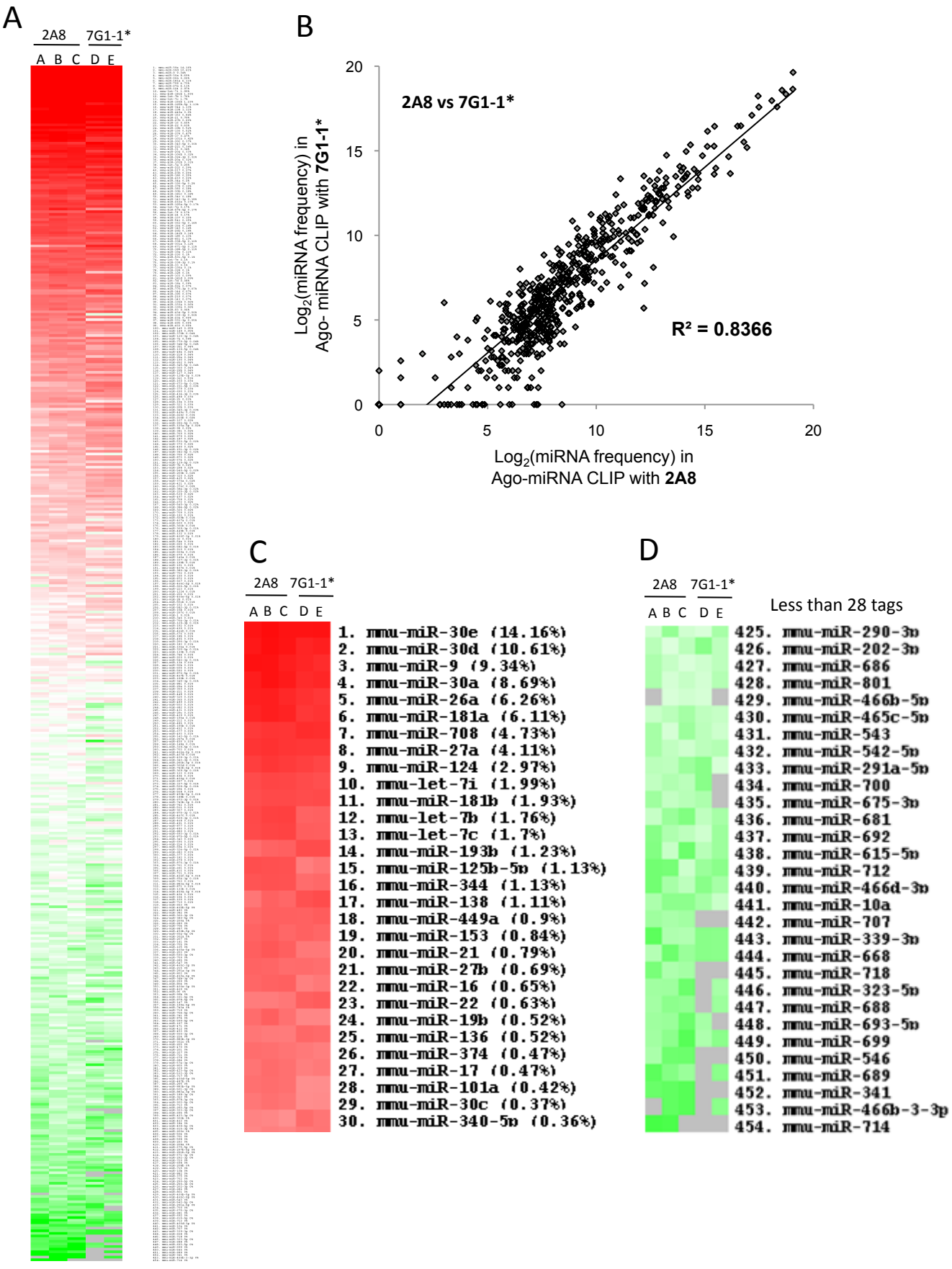
Supplemental Table 3 (continued)

				All seeds				Conserved seeds				
Rank	Frequency in Ago miRNA CLIP	miRNA family	Seed motif	Obs/Exp	P-value	Observed Frequeny in Ago footprint	Expected frequency	Obs/Exp	P-value	Observed Frequency in Ago Footprint	Expected frequency	miRNAs
21	116912	miR-374	TATTATAT	1.071	8.04E-01	13	12	2.220	6.71E-02	5	2	miR-374
21	116912	miR-374	TATTATA	1.195	2.48E-01	42	35	3.489	1.65E-10	23	7	miR-374
21	116912	miR-374	ATTATAT	0.830	2.75E-01	34	41	1.518	1.63E-01	11	7	miR-374
21	171240	miR-374	TTATAT	1.058	4.53E-01	179	169	1.780	1.01E-06	70	39	miR-410:miR-374
21	116912	miR-374	ATTATA	1.216	1.61E-02	151	124	2.223	4.70E-09	51	23	miR-374
21	119877	miR-374	TATTAT	1.135	1.03E-01	165	145	1.928	3.57E-07	58	30	miR-369-3p:miR-374
21		miR-374	with all 6mers	1.128	7.26E-03	495	439	1.938	1.96E-19	179	92	
22	154156	miR-22	GGCAGCTT	0.764	2.65E-01	17	22	1.018	9.72E-01	4	4	miR-22
22	154156	miR-22	GCAGCTT	0.800	7.21E-02	65	81	1.590	2.49E-02	23	14	miR-22
22	154156	miR-22	GGCAGCT	0.626	5.50E-05	73	117	0.957	8.24E-01	26	27	miR-22
22	157986	miR-22	CAGCTT	0.884	3.83E-02	283	320	1.652	1.55E-07	107	65	miR-320:miR-22
22	154156	miR-22	GCAGCT	0.623	1.49E-15	279	448	0.906	2.98E-01	111	123	miR-22
22	154156	miR-22	GGCAGC	0.452	1.73E-29	191	423	0.652	3.48E-04	69	106	miR-22
22		miR-22	with all 6mers	0.632	7.09E-37	753	1191	0.979	7.22E-01	287	293	
23	156668	miR-20	GCACTTTA	3.574	1.10E-18	42	12	6.309	5.66E-32	31	5	miR-20a:miR-106b
23	156668	miR-20	CACTTTA	2.116	2.01E-12	84	40	4.810	4.60E-30	43	9	miR-20a:miR-106b
23	297690	miR-20 or miR-17	GCACTTT	1.815	4.51E-09	94	52	3.806	1.22E-30	64	17	miR-17:miR-292-3p:miR-291b-3p:miR-295:miR-294:miR-20b:miR-20a:miR-93 :miR-106a:miR-106b:miR-290-3p:miR-291a-3p
23	157636	miR-20	ACTTTA	1.332	2.41E-05	215	161	2.897	3.92E-26	90	31	miR-20a:miR-142-5p:miR-106b
23	306438	miR-20 or miR-17	GCACTT	1.306	7.00E-05	221	169	2.784	3.44E-30	114	41	miR-17:miR-292-3p:miR-291b-3p:miR-295:miR-294:miR-467b:miR-467c:miR-20b:miR-20a:miR-93:miR-105:miR-302b:miR-302c:miR-302a:miR-302d:miR-467d:miR-106a:miR-106b:miR-467a:miR-290-3p:miR-291a-3p
23	297690	miR-20 or miR-17	CACTTT	1.233	6.16E-04	267	217	2.834	3.62E-35	129	46	miR-17:miR-292-3p:miR-291b-3p:miR-295:miR-294:miR-20b:miR-20a:miR-93 :miR-106a:miR-106b:miR-290-3p:miR-291a-3p
23		miR-20	with all 6mers	1.285	2.75E-11	703	547	2.833	6.58E-88	333	118	
24	127192	miR-136	AATGGAGT	1.228	3.96E-01	17	14	2.290	5.67E-02	5	2	miR-136
24	127946	miR-136	ATGGAGT	0.803	1.29E-01	48	60	0.991	9.72E-01	15	15	miR-880:miR-136
24	127192	miR-136	AATGGAG	0.880	2.90E-01	69	78	1.341	1.58E-01	23	17	miR-136
24	128805	miR-136	TGGAGT	0.690	2.26E-07	193	280	0.851	2.11E-01	60	71	miR-880:miR-136:miR-509-5p
24	127946	miR-136	ATGGAG	0.685	2.34E-09	247	360	0.851	1.08E-01	99	116	miR-880:miR-136
24	127192	miR-136	AATGGA	1.127	3.96E-02	297	264	1.679	3.52E-09	127	76	miR-136
24		miR-136	with all 6mers	0.816	3.03E-08	737	904	1.090	1.46E-01	286	262	
25	82032	miR-204	AAAGGGAA	0.534	1.35E-02	15	28	1.026	9.54E-01	5	5	miR-204:miR-211
25	82032	miR-204	AAAGGGA	0.562	1.53E-04	42	75	1.659	3.00E-02	18	11	miR-204:miR-211
25	82032	miR-204	AAGGGAA	0.683	3.73E-03	57	84	1.204	4.72E-01	15	12	miR-204:miR-211
25	109152	miR-204	AAGGGA	0.664	1.47E-07	163	245	1.387	2.03E-02	50	36	miR-204:miR-343:miR-211:miR-188-5p
25	82032	miR-204	AGGGAA	0.807	1.31E-03	224	278	1.923	5.22E-09	77	40	miR-204:miR-211
25	82032	miR-204	AAAGGG	0.610	9.55E-09	132	216	1.395	2.84E-02	43	31	miR-204:miR-211
25		miR-204	with all 6mers	0.702	5.43E-16	519	739	1.590	1.07E-09	170	107	
26	87290	miR-340	CTTTATAA	1.778	4.26E-03	24	13	2.914	3.01E-03	7	2	miR-340-5p
26	87290	miR-340	CTTTATA	1.358	2.16E-02	56	41	2.903	3.08E-07	21	7	miR-340-5p
26	87290	miR-340	TTTATAA	1.191	1.03E-01	87	73	1.765	4.61E-04	37	21	miR-340-5p
26	88258	miR-340	CTTTAT	1.329	7.51E-06	246	185	2.673	9.79E-23	92	34	miR-340-5p:miR-142-5p
26	87290	miR-340	TTATAA	1.184	1.93E-02	192	162	1.999	1.21E-09	74	37	miR-340-5p
26	87290	miR-340	TTTATA	1.100	1.29E-01	252	229	1.889	1.35E-13	131	69	miR-340-5p:miR-466l
26		miR-340	with all 6mers	1.197	2.18E-06	690	576	2.109	1.41E-39	297	141	
27	82568	miR-31	TCTTGCCT	1.838	3.11E-04	34	18	4.579	2.24E-06	8	2	miR-31
27	82568	miR-31	TCTTGCC	1.202	1.51E-01	61	51	3.179	4.63E-08	20	6	miR-31
27	82568	miR-31	CTTGCCT	1.506	2.00E-05	107	71	3.075	1.39E-08	23	7	miR-31
27	82568	miR-31	TTGCCT	2.030	6.53E-56	475	234	4.430	2.72E-108	184	42	miR-31
27	82568	miR-31	TCTTGC	1.177	2.06E-02	202	172	2.550	5.12E-15	65	25	miR-31
27	82568	miR-31	CTTGCC	1.129	6.19E-02	236	209	2.481	8.89E-15	68	27	miR-31
27		miR-31	with all 6mers	1.485	2.42E-33	913	615	3.357	4.30E-116	317	94	
28	78086	miR-324-3p	GGCAGTGG	0.202	2.58E-06	7	35	0.416	2.01E-01	2	5	miR-324-3p
28	78086	miR-324-3p	GGCAGTG	0.373	1.36E-10	39	105	0.897	6.85E-01	14	16	miR-324-3p
28	78086	miR-324-3p	GCAGTGG	0.392	2.93E-10	42	107	0.455	1.54E-02	9	20	miR-324-3p
28	78086	miR-324-3p	GCAGTG	0.618	1.33E-12	213	345	1.048	6.96E-01	69	66	miR-324-3p
28	78086	miR-324-3p	CAGTGG	0.673	1.16E-10	262	389	1.040	7.06E-01	91	87	miR-324-3p
28	78086	miR-324-3p	GGCAGT	0.535	1.67E-12	123	230	1.005	9.75E-01	38	38	miR-324-3p
28		miR-324-3p	with all 6mers	0.620	4.73E-32	598	964	1.036	6.18E-01	198	191	
29	74880	miR-23	AATGTGAT	1.541	2.94E-02	25	16	2.516	6.87E-03	8	3	miR-23b:miR-23a
29	74880	miR-23	AATGTGA	1.667	5.02E-09	128	77	3.300	4.12E-24	64	19	miR-23b:miR-23a
29	74880	miR-23	ATGTGAT	1.106	4.35E-01	60	54	1.440	1.55E-01	15	10	miR-23b:miR-23a
29	75593	miR-23	TGTGAT	1.284	1.76E-05	293	228	2.240	6.30E-19	115	51	miR-23b:miR-23a:miR-377
29	75148	miR-23	AATGTG	1.571	1.85E-20	414	264	2.873	1.62E-46	168	58	miR-323-3p:miR-23b:miR-23a
29	74880	miR-23	ATGTGA	1.380	4.61E-09	329	238	2.482	1.27E-27	134	54	miR-23b:miR-23a
29		miR-23	with all 6mers	1.419	1.06E-29	1036	730	2.546	4.05E-87	417	164	
30	65491	miR-217	ATGCAGTA	1.324	2.93E-01	14	11	2.648	1.31E-02	6	2	miR-217
30	65491	miR-217	TGCAGTA	1.140	3.40E-01	53	46	1.827	7.66E-03	19	10	miR-217
30	65491	miR-217	ATGCAGT	1.024	8.54E-01	60	59	1.452	9.33E-02	20	14	miR-217
30	66220	miR-217	TGCAGT	0.890	6.56E-02	247	278	1.407	1.29E-03	88	63	miR-883a-3p:miR-883b-3p:miR-217
30	65491	miR-217	GCAGTA	0.962	6.42E-01	141	147	1.911	3.35E-07	60	31	miR-217
30	65491	miR-217	ATGCAG	0.865	2.08E-02	252	291	1.085	4.29E-01	95	88	miR-217
30		miR-217	with all 6mers	0.894	4.62E-03	640	716	1.338	5.10E-06	243	182	

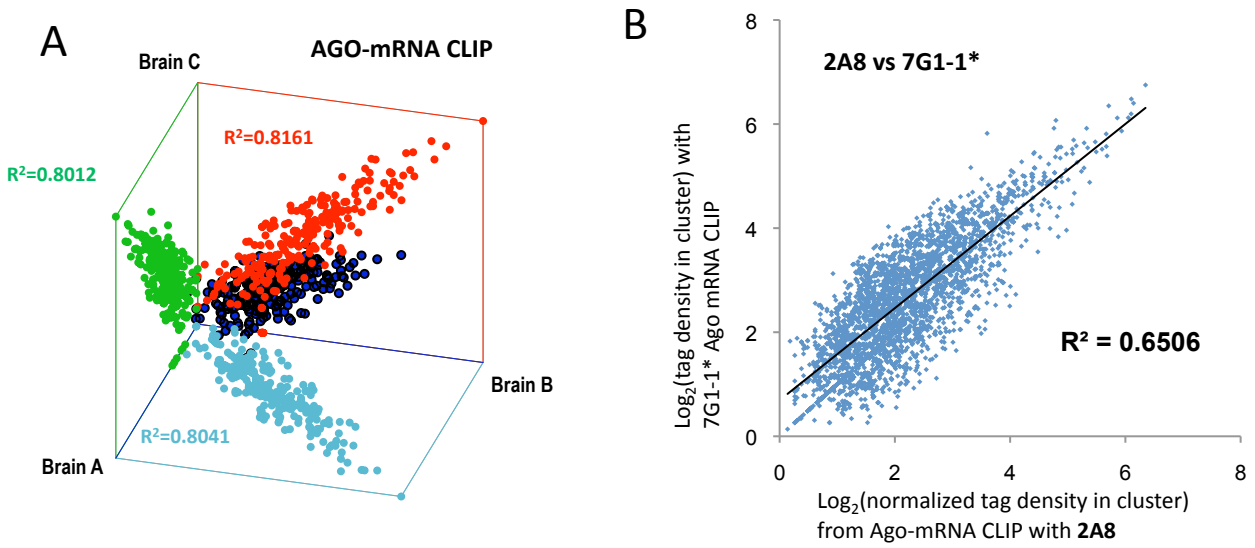
Supplemental Figure 1



Supplemental Figure 2

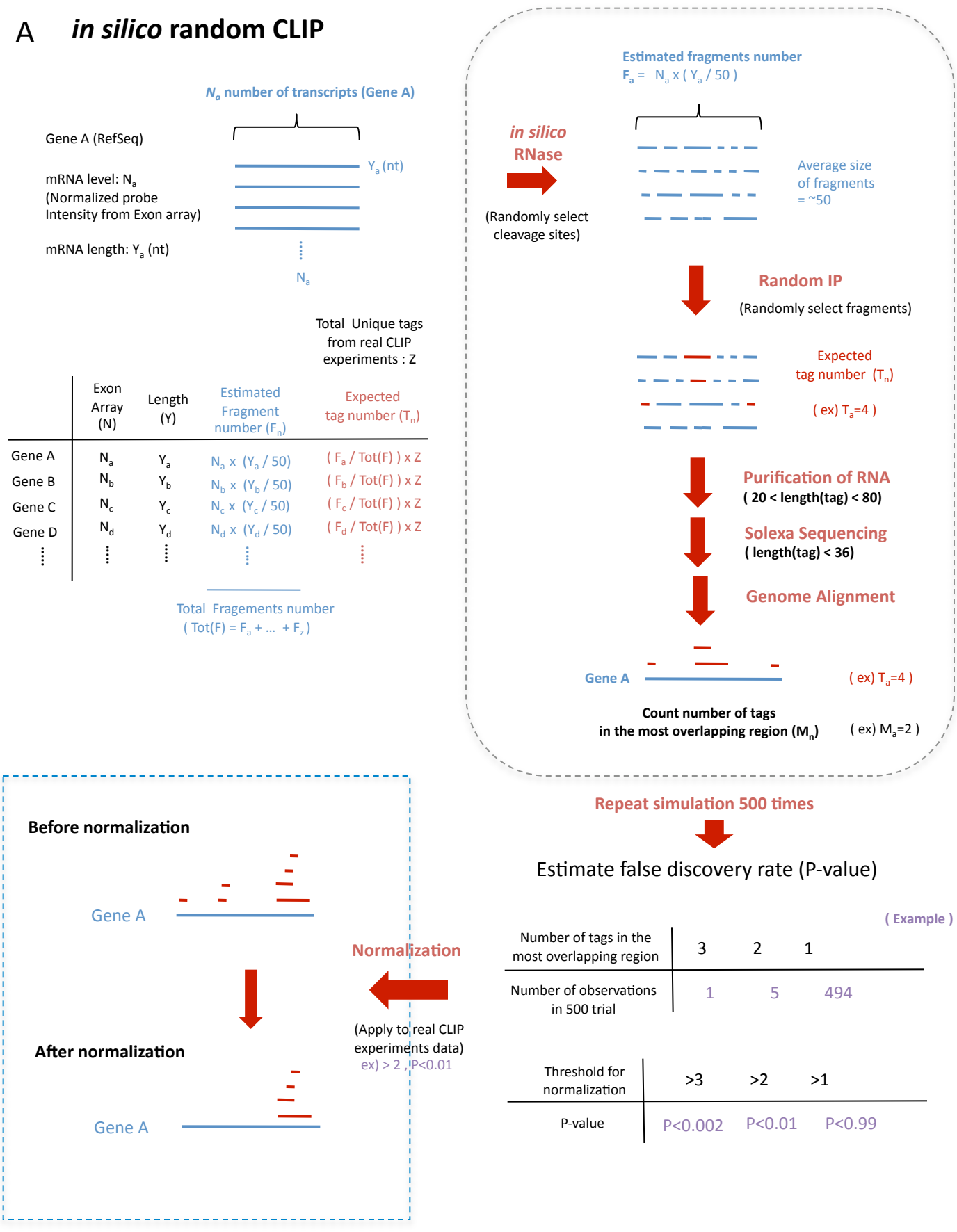


Supplemental Figure 3



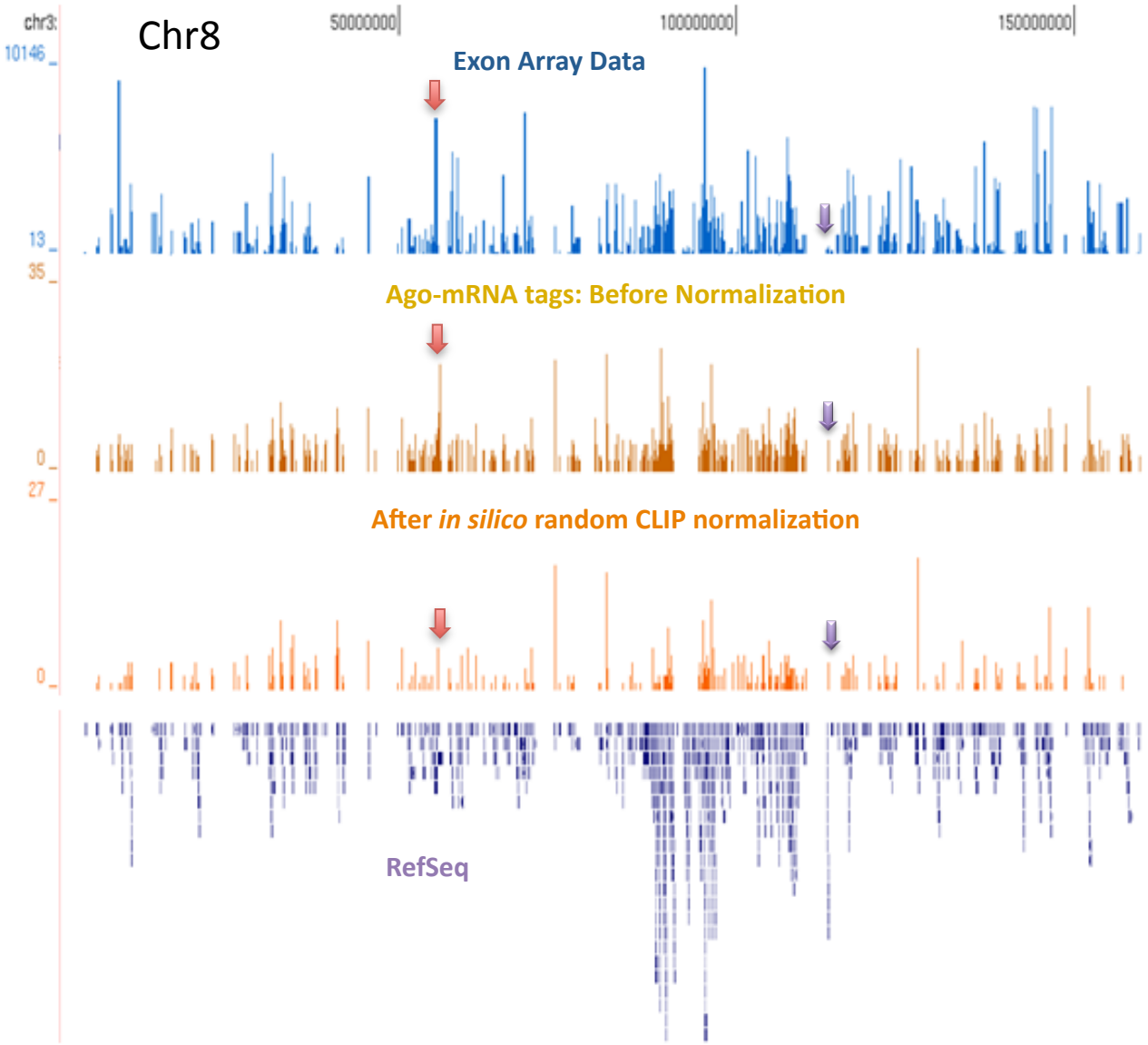
C

Supplemental Figure 4



Supplemental Figure 4

B



Supplemental Figure 4

C

Chad7 (NM 001081417)	(N_n) Exon array value	Length (Y_n)	Expected number of tags (T_n)
1x	100	9444	5
2x	200	9444	5
4x	400	9444	5
8x	800	9444	5
16x	1600	9444	5
32x	3200	9444	5
72x	7200	9444	5
100x	10000	9444	5
500x	50000	9444	5

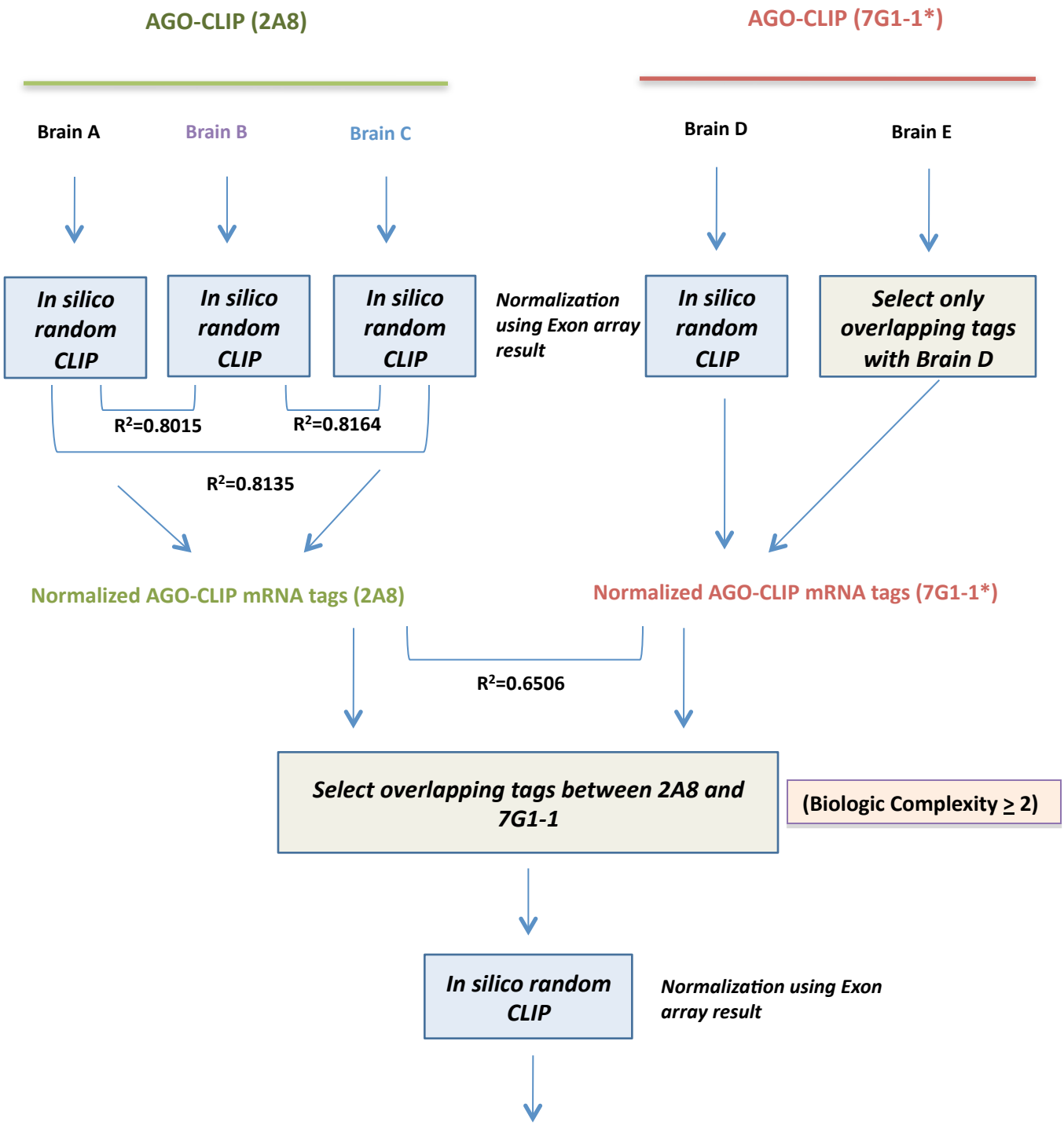
Number of tags in the most overlapping region (Mn)

Simulation	3	2	1
1x	0	12	488
2x	0	52	448
4x	0	21	479
8x	0	57	443
16x	0	43	457
32x	0	20	480
72x	0	41	459
100x	0	61	439
500x	0	35	465

False discovery rate base on *in silico* CLIP simulation result

Simulation	3	2	1
1x	<0.002	0.024	0.976
2x	<0.002	0.104	0.896
4x	<0.002	0.042	0.958
8x	<0.002	0.114	0.886
16x	<0.002	0.086	0.914
32x	<0.002	0.04	0.96
72x	<0.002	0.082	0.918
100x	<0.002	0.122	0.878
500x	<0.002	0.07	0.93

Supplemental Figure 5

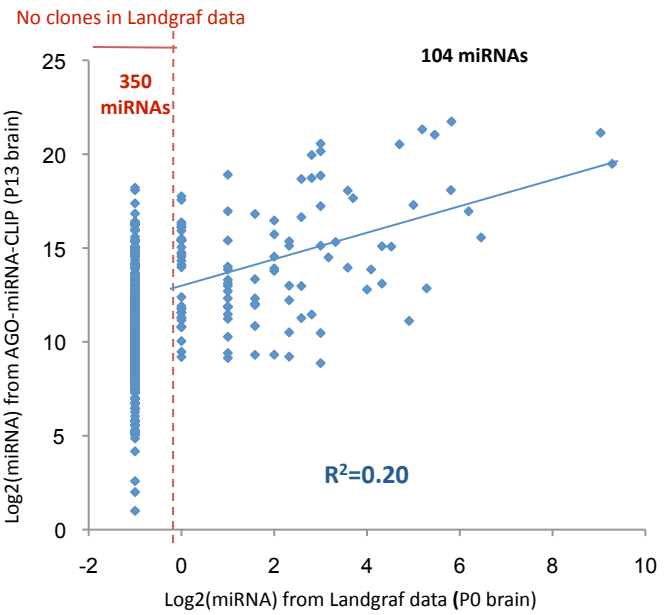


Supplemental Figure 6

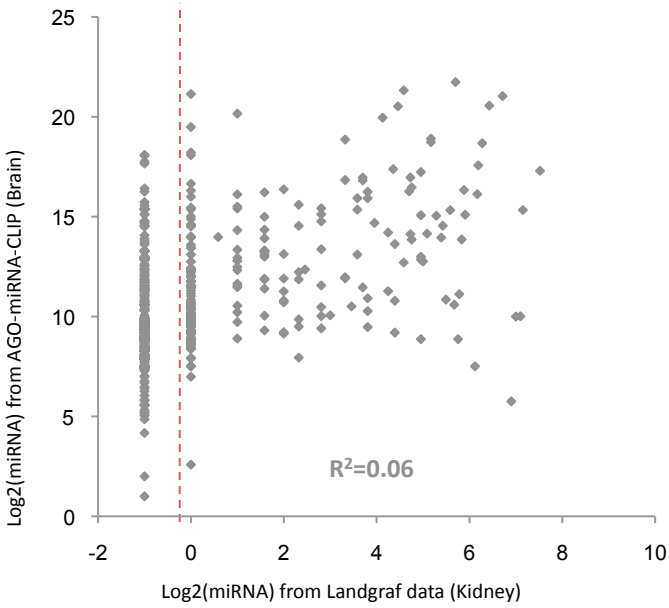
A

miRNA	TOTAL CLONES NODE (Landgraf et al)	Total clones in AGO-miRNA-CLIP	Rank in AGO-miRNA CLIP
mmu-miR-124	740	735430	9
mmu-miR-9	547	2309955	3
mmu-miR-126-3p	131	10238	103
mmu-miR-16	105	160843	22
mmu-miR-140	96	34537	62
mmu-miR-125b-5p	91	278206	15
mmu-miR-27a	84	1017631	8
mmu-miR-30e	82.5	3504225	1
mmu-miR-136	75	127192	25
mmu-miR-143	71	15878	89
mmu-miR-872	64	2224	195
mmu-miR-30a	60	2150543	4
mmu-miR-30d	47.5	2624768	2
mmu-miR-15a	41	35050	61
mmu-miR-379	39	7411	123
mmu-let-7c	35	419429	13
mmu-miR-29b	32	8154	116
mmu-miR-24	32	41024	57
mmu-miR-22	32	154156	23
mmu-miR-130a	28	14884	92

B



C

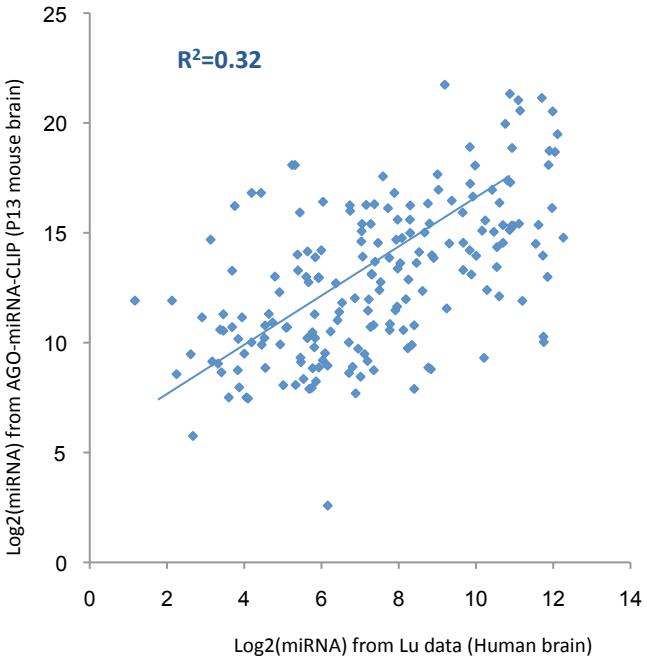


Supplemental Figure 6

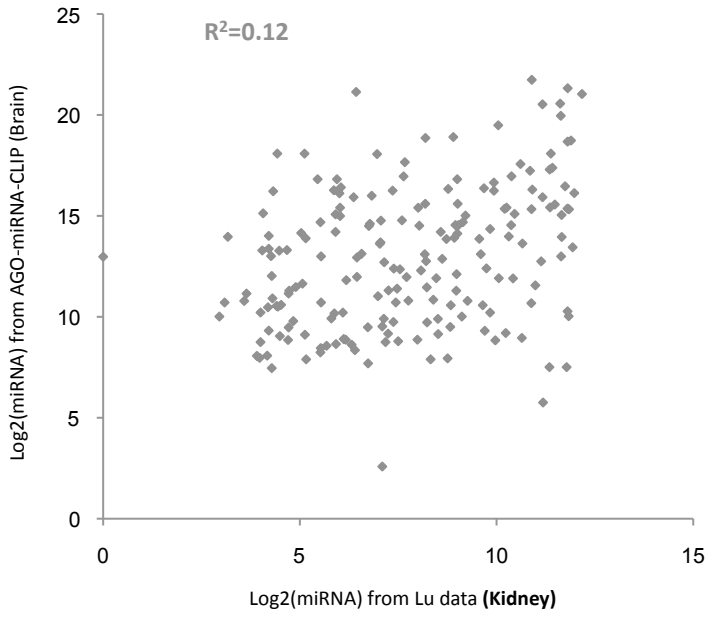
D

miRNA	Lu et al: Expression level	AGO miRNA CLIP: Total clones in AGO-miRNA CLIP	AGO miRNA CLIP: Rank in AGO-miRNA CLIP
miR-338-5p	4913	28123	67
miR-124	4425	735430	9
let-7c	4229	419429	13
miR-181a	4054	1510285	6
let-7a	4009	71661	38
let-7b	3824	435438	12
miR-125b-5p	3767	278206	15
miR-29b	3701	8154	116
miR-29c	3455	1045	246
miR-29a	3439	1238	236
miR-219	3403	15966	88
miR-9	3339	2309955	3
miR-125a-5p	3148	42068	53
miR-128	2993	23182	79
miR-320	2353	3830	156
miR-26a	2262	1547706	5
miR-181c	2217	43566	49
miR-30a	2190	2150543	4
miR-24	2005	41024	57
miR-181b	1958	475794	11
let-7f	1949	41306	56
miR-16	1897	160843	22
miR-30d	1882	2624768	2
miR-137	1874	35776	58
miR-27b	1803	170907	21
miR-27a	1733	1017631	8

E



F



Supplemental Figure 7

A

Number of Ago-mRNA clusters

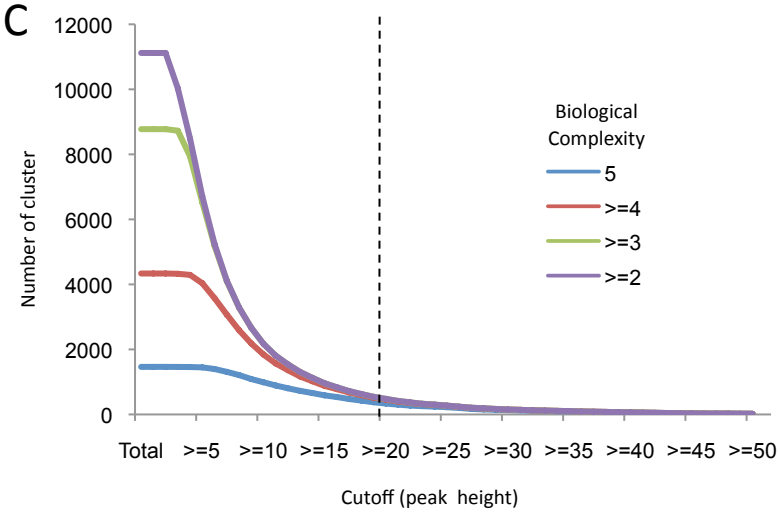
Cutoff (peak height)	Biological Complexity 5	>=4	>=3	>=2
Total	1463	4335	8775	11118
>=5	1448	4036	6521	6710
>=10	990	1844	2172	2173
>=15	587	878	959	959
>=20	336	456	479	479
>=25	223	265	272	272
>=30	134	151	151	151
>=35	94	102	102	102
>=40	60	63	63	63
>=45	31	32	32	32
>=50	20	20	20	20

B

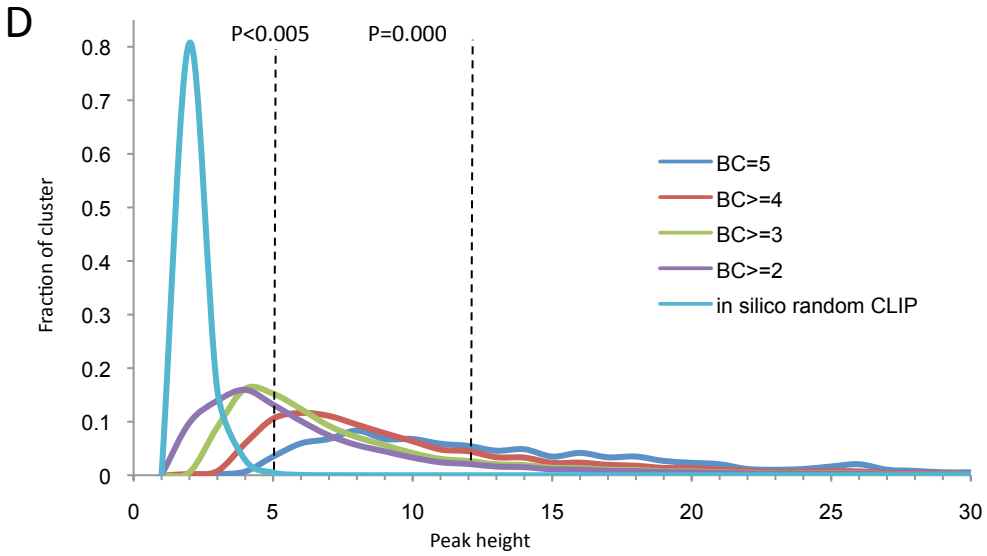
complexity	>=5	>=4	>=3	>=2
Number of cluster	1463	4335	8775	11118
Number of genes	829	2081	3563	4280
Clusters / genes	1.77	2.08	2.46	2.60

P13 brain expressed gene: 10743 genes
(DABG : p<0.05 , Value > 100)

C



D



Supplemental Figure 7

E

miR-124 6mers

BC	Obs/Exp	p-value	True positive rate(%)	Obs	Exp
>=2	3.2649573	0.00 x 100	76.5439348	2292	702.3587
>=3	3.7799929	0.00 x 100	79.07946684	2032	537.5671
>=4	4.071868	0.00 x 100	80.28339873	1180	289.7933
=5	5.0520888	0.00 x 100	83.47677917	453	89.66588

miR-124 6mers

peak	Obs/Exp	p-value	True positive Rate(%)	Obs	Exp
>=2	3.2649573	0.00 x 100	76.55310621	2292	702
>=10	5.3136153	0.00 x 100	84.16121417	832	156.5789
>=20	6.4373708	1.32 x 10-290	86.55438834	289	44.8941
>=30	6.6297288	1.08 x 10-124	86.89337425	118	17.79862
>=40	5.5610835	1.4E-42	84.7586149	50	8.991054
>=50	5.0924938	2.68E-15	83.58636006	19	3.730981

miR-19 6mers

BC	Obs/Exp	p-value	True positive rate (%)	Obs	Exp
>=2	1.1714286	1.69E-05	53.94736842	738	630
>=3	1.2227305	9.95E-07	55.01029036	590	482.5266
>=4	1.1302395	0.035681	53.05692112	294	260.1219
=5	1.428835	0.000119	58.82799722	115	80.48515

miR-19 6mers

peak	Obs/Exp	p-value	True positive rate(%)	Obs	Exp
>=2	1.1714286	1.69E-05	53.94736842	738	630
>=10	1.1384087	0.100824	53.23625524	160	140.5471
>=20	1.0174334	0.91188	50.43206914	41	40.29748
>=30	1.0640796	0.797852	51.55225689	17	15.97625
>=40	1.2390843	0.497008	55.33888625	10	8.070476
>=50	0.597198	0.46104	37.39035571	2	3.348973

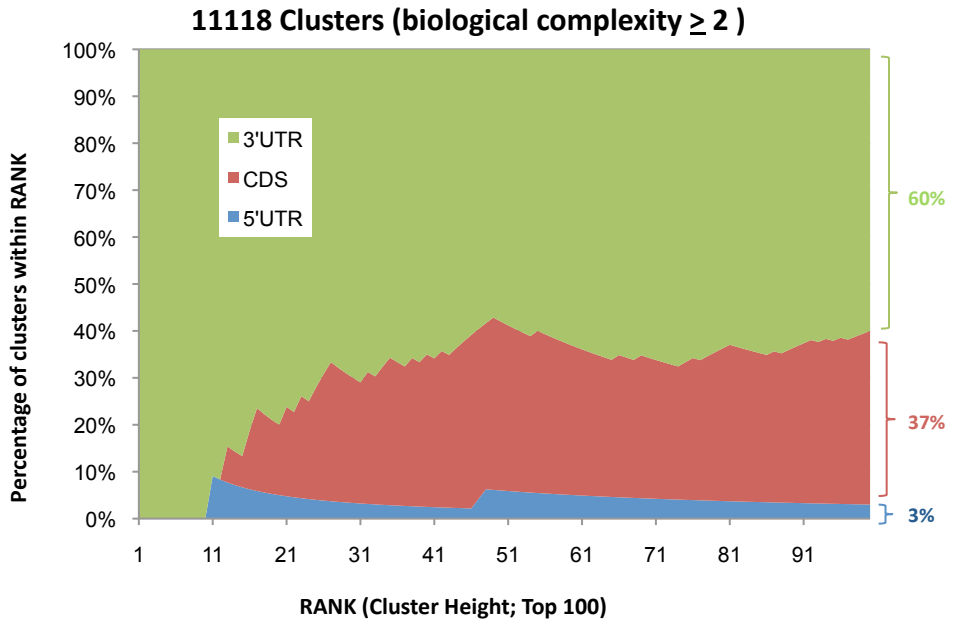
Supplemental Figure 8

A

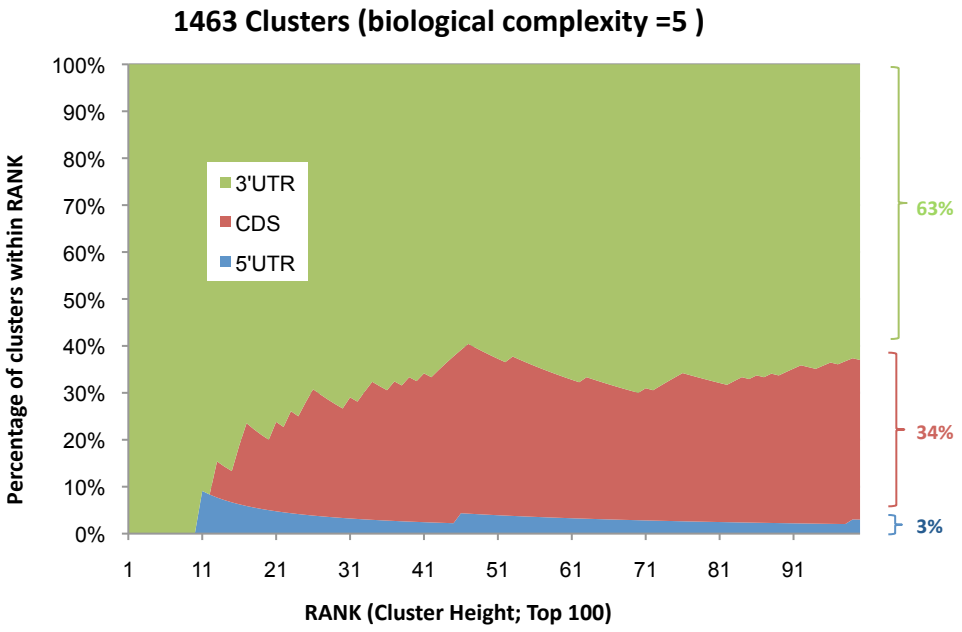
Enrichment of AGO-CLIP mRNA clustered tags in 3' UTR

	Number of clusters (BC ≥ 2)	% Clusters	Number of tags	% tags	Total length (brain expressed transcripts)	% Total Length	Obs/Exp (number of clusters/ total length)	P-value	Obs/Exp (number of tags/ total length)	P-value
5'UTR	211	1.9	1,742	1.7	3,686,859	6.8	0.28	1.92×10^{-87}	0.25	0.00×10^{0}
CDS	4992	44.9	42,553	41.8	31,287,950	57.4	0.78	8.77×10^{-68}	0.73	0.00×10^{0}
3'UTR	5915	53.2	57,589	56.5	19,571,210	35.9	**1.48**	1.26×10^{-203}	**1.58**	0.00×10^{0}
Total	11,118	100	101,884	100	54,546,019	100				

B



C



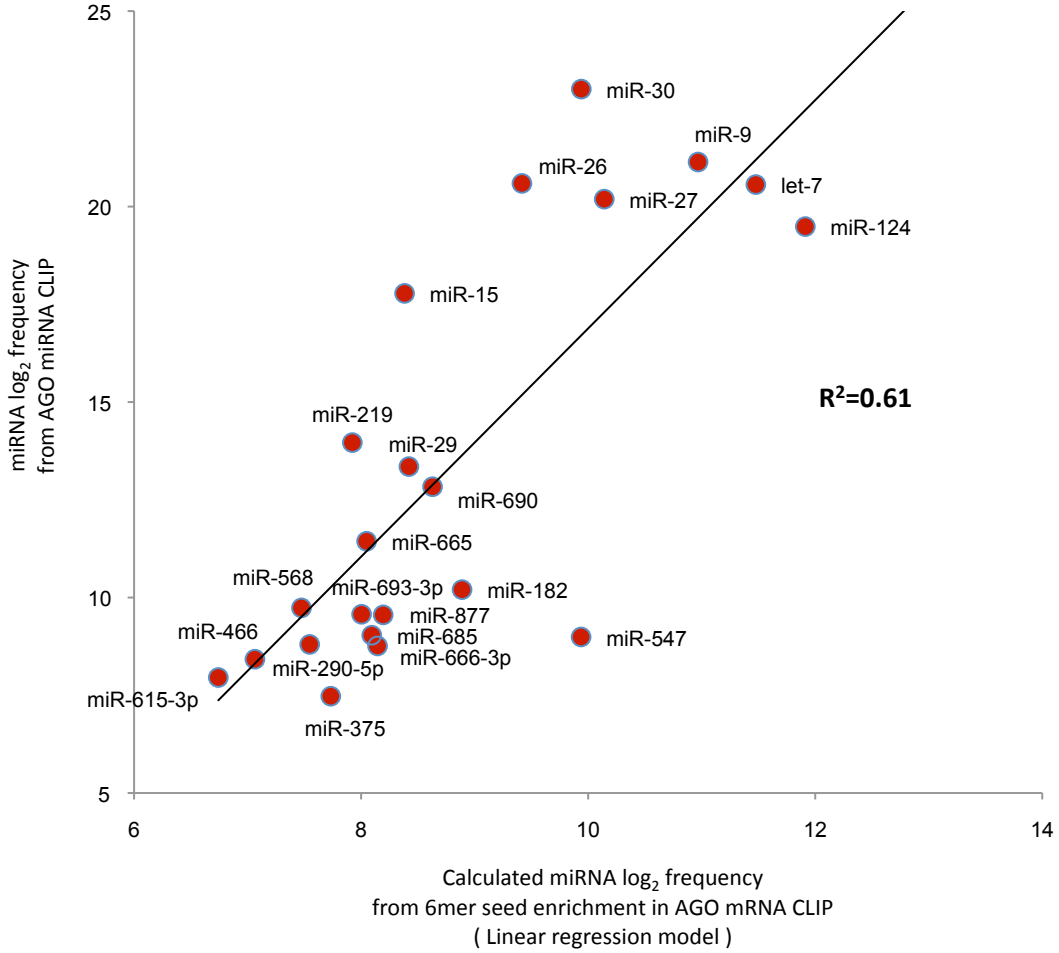
Supplemental Figure 9

A

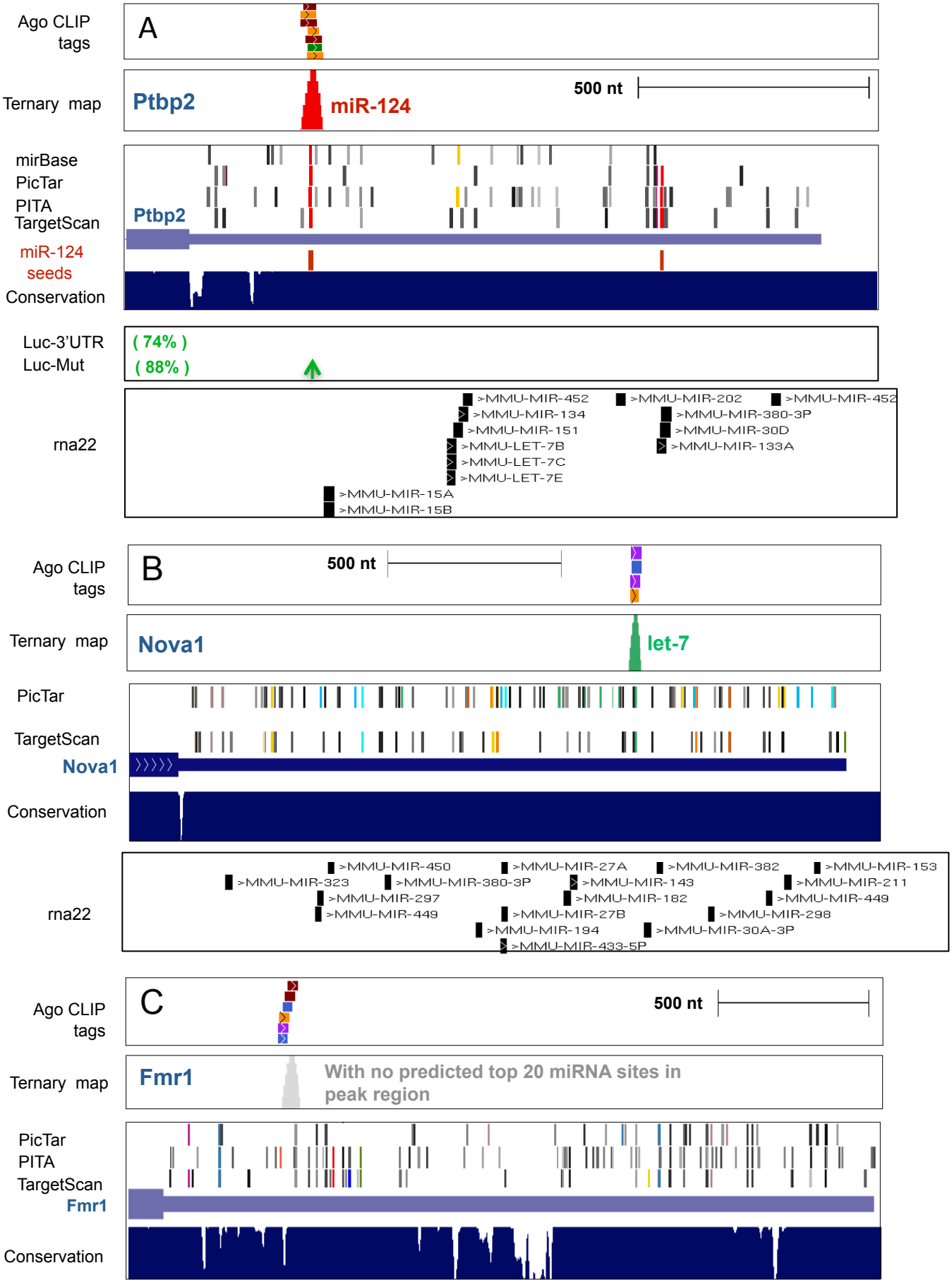
AGO-mRNA CLIP				AGO-miRNA CLIP			
Top ranking motif	Regression coefficient	Number of unique tags	P-value	miRNA	Matched seed position	Fraction in AGO-miRNA CLIP	Rank in AGO-miRNA CLIP
GTGCCTT	0.655	4581	9.01×10^{-31}	mir-124	2-8	2.97%	8
TGCCTT	0.421	9153	3.74×10^{-30}		2-7		
GTGCCT	0.247	6915	8.45×10^{-9}		3-8		
TACCTC	0.452	6403	2.13×10^{-25}	let-7	2-7	6.24%	5
TACCTCA	0.529	3558	1.82×10^{-18}		2-8		
CCAAAG	0.287	6878	9.22×10^{-14}	mir-9	2-7	9.33%	2
CCAAAGA	0.376	3549	2.06×10^{-12}		1-7		
TGTTTAC	0.404	2136	3.29×10^{-8}	mir-30	2-8	34.01%	1
GTTTAC	0.331	2964	7.19×10^{-8}		2-7		
CTGTGA	0.274	4205	1.47×10^{-9}	mir-27	2-7	4.80%	6
ACTGTGA	0.394	1665	3.01×10^{-7}		2-8		
TACTTGAA	0.676	877	3.89×10^{-8}	mir-26	1-8	6.39%	4
ACTTGA	0.290	2353	3.48×10^{-6}		2-7		

BC>=2 , -24 ~ +22 region

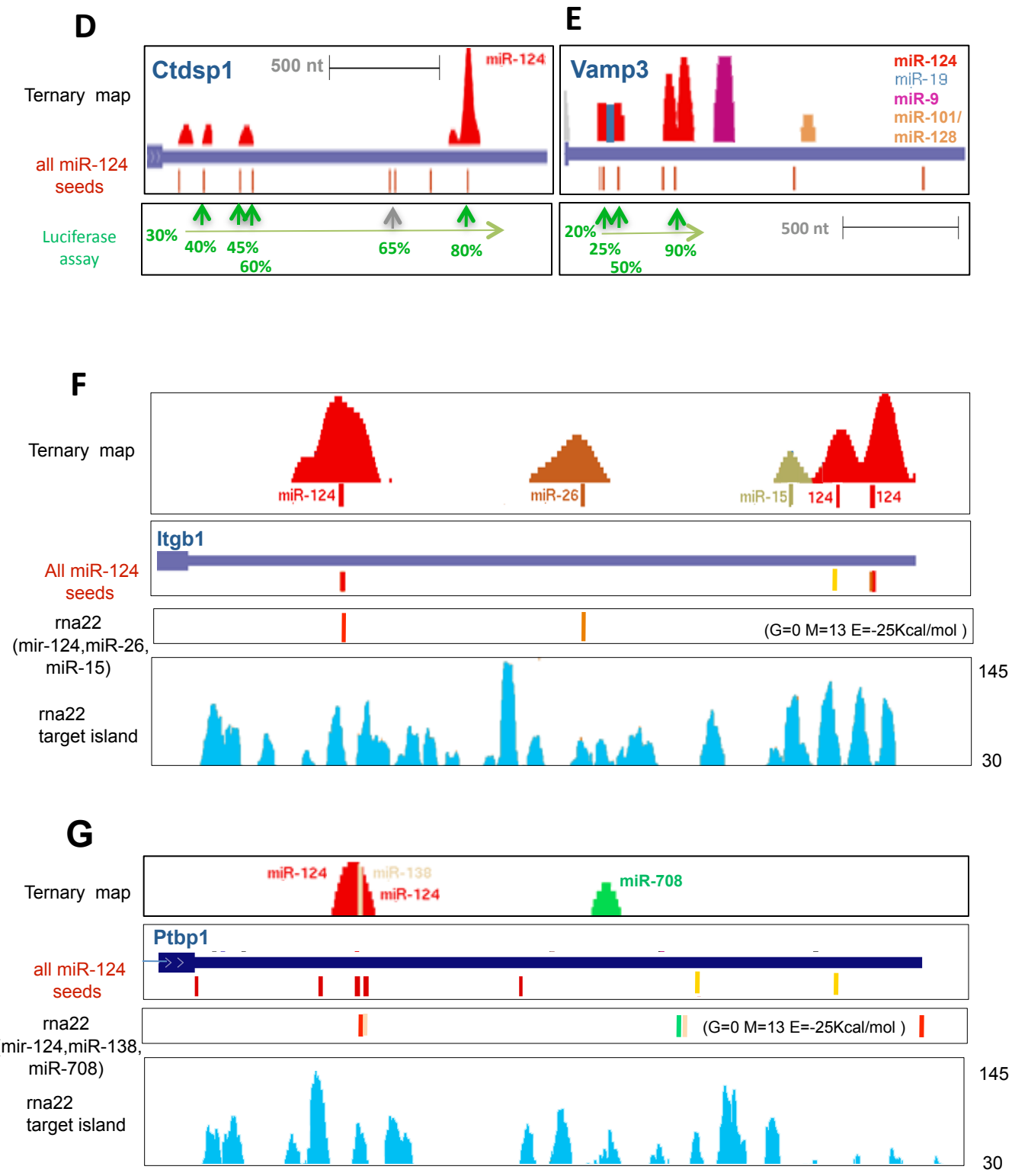
B



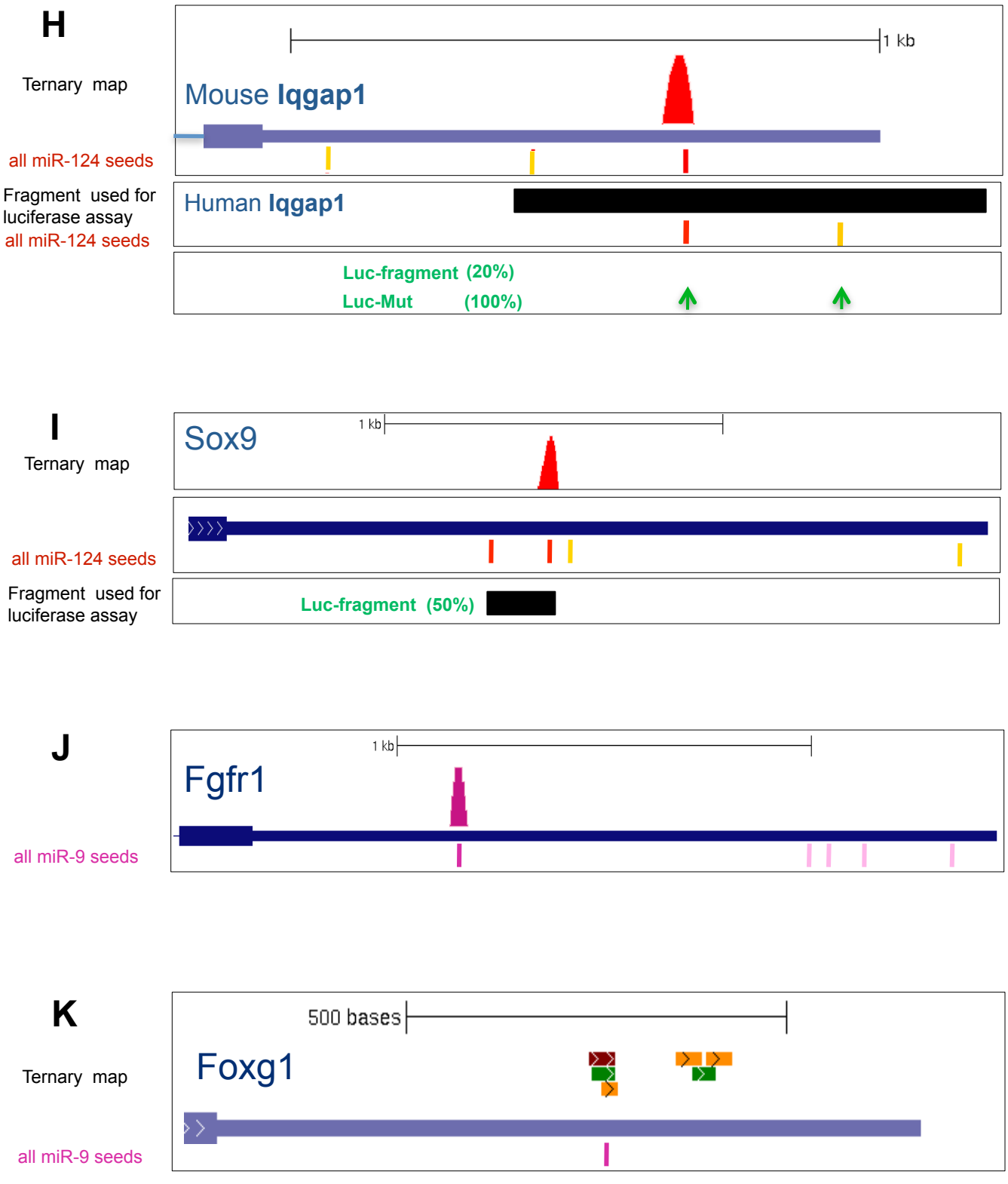
Supplemental Figure 10



Supplemental Figure 10



Supplemental Figure 10



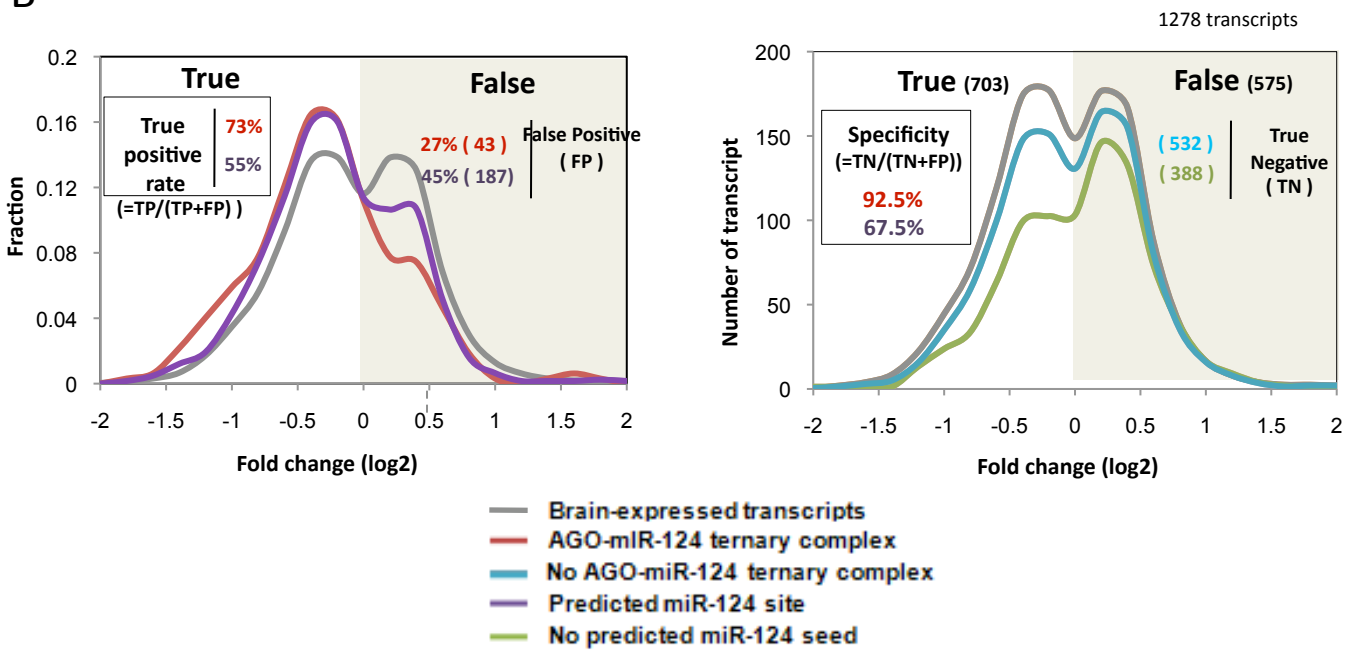
Supplemental Figure 11

A

GEO	experiments	Cell line	species	measure	Reference
GSE8498	overexpression	neuroblastoma CAD	mouse	microarray	Makeyev EV, Zhang J, Carrasco MA, Maniatis T. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol Cell 2007 Aug 3;27(3): 435-48.
GSE6207	overexpression	HepG2	human	microarray	Wang X, Wang X. Systematic identification of microRNA functions by combining target prediction and expression profiling. Nucleic Acids Res 2006;34(5):1646-52
GSE2075	overexpression	Hela	human	microarray	Lim LP, Lau NC, Garrett-Engele P, Grimson A et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 2005 Feb 17;433(7027):769-73
GSE11968	overexpression	Hela	human	SILAC-MS microarray	Baek D, Villén J, Shin C, Camargo FD et al. The impact of microRNAs on protein output. Nature 2008 Sep 4;455(7209):64-71
GSE11080	overexpression	HEK293T	human	microarray	David G. Hendrickson, Daniel J. Hogan, Daniel Herschlag, James E. Ferrell and Patrick O. Brown, Systematic Identification of mRNAs Recruited to Argonaute 2 by Specific microRNAs and Corresponding Changes in Transcript Abundance PLoS ONE, 2008 May 7;3(5):e2126

5 microarray (meta-analysis) : 1278 genes (more than two exp : p<0.05, brain expressed)

B

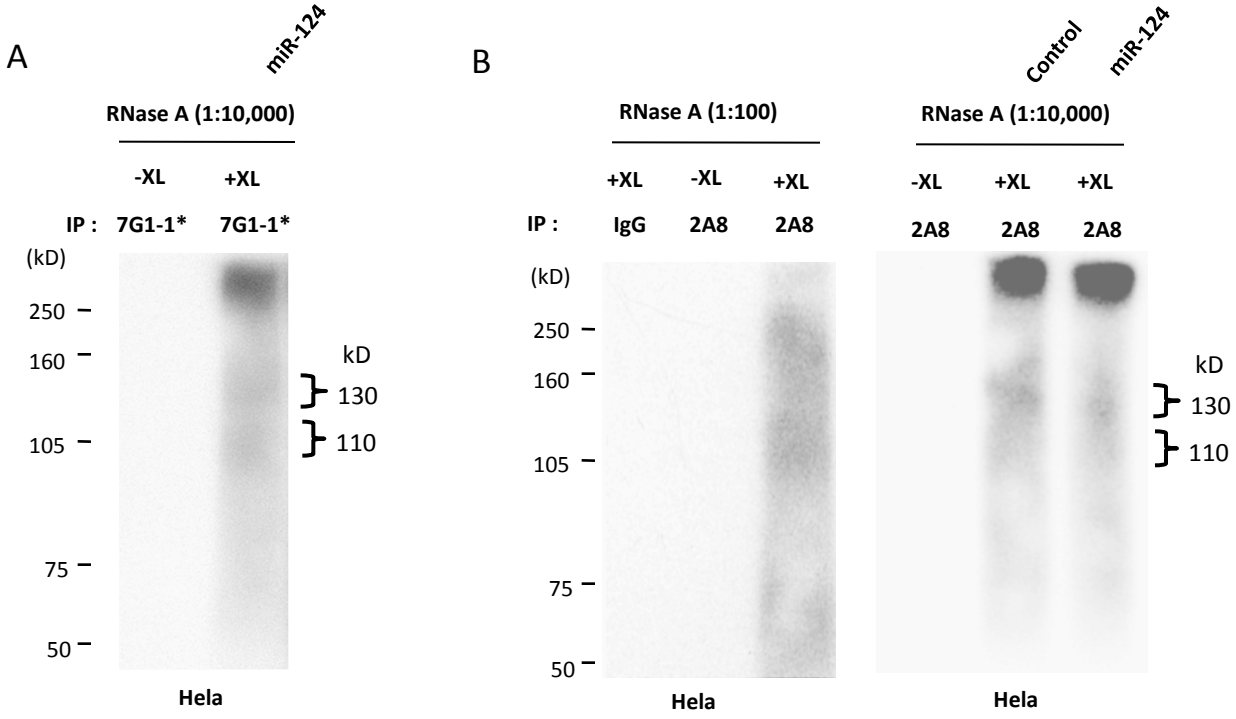


Supplemental Figure 11

C

miR-124 dependent changes (log_2)					
Gene Name	Micro-array	SILAC	IP-micro-array[1] ($P<10^{-4}$)	IP-micro-array[2] (<1% Local FDR)	Tx level P13 Brain
Itgb1	-4.90				2959
Ctdsp1	-3.60			1.73	1035
Cd164	-3.40			2.63	2690
Plod3	-3.20	-3.48	1.64		694
Lamc1	-3.15				881
Sypl	-3.05			1.31	2259
Hadhb	-2.90	-1.20	2.43		absent
Lass2	-2.75				3194
Ptbp1	-2.50	-0.97			469
C6orf72	-2.35				436
Arfip1	-2.30				101
Cebpa	-2.20			1.61	370
Epim	-2.20			1.72	544
Fcho2	-2.15				633
B4galt1	-2.10		3.20		340
E2ig4	-2.05				absent
Tom1l1	-2.00				1275
Ctdsp2	-1.95			1.50	623
Nek9	-1.90				691
Mapk14	-1.85			1.43	2061
Vamp3	-1.60		2.70	2.30	1138
Ptpn12	-1.50			1.61	750

Supplemental Figure 12



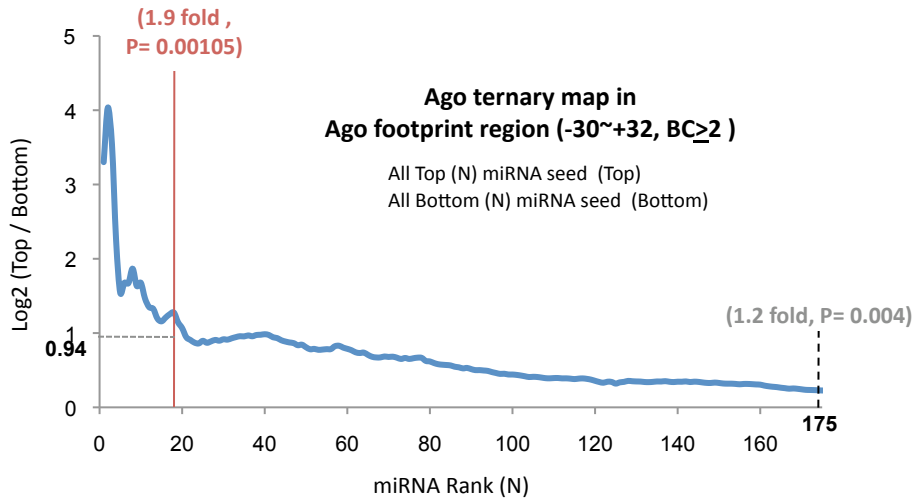
C

Sample	Transfected miRNA	Antibody	Raw tags (prefiltered)	Aligned tags (hg18)	% align	Unique tags
HeLa 1	miR-124	2A8	11,563,568	6,082,765	52.6	1,271,461
HeLa 2	miR-124	2A8	12,681,689	5,715,019	45.1	1,435,945
HeLa 3	miR-124	7G1-1	12,484,953	5,816,073	46.6	1,459,849
HeLa 4	miR-124	7G1-1	12,441,955	5,419,481	43.6	1,603,967
HeLa 5	control miRNA	2A8	10,802,378	5,951,833	55.1	1,184,079
HeLa 6	control miRNA	2A8	11,104,169	5,813,633	52.4	1,247,774
HeLa 7	control miRNA	7G1-1	12,204,108	6,916,662	56.7	1,272,437
HeLa 8	control miRNA	7G1-1	12,957,786	3,396,298	26.2	1,127,326

Supplemental Figure 13

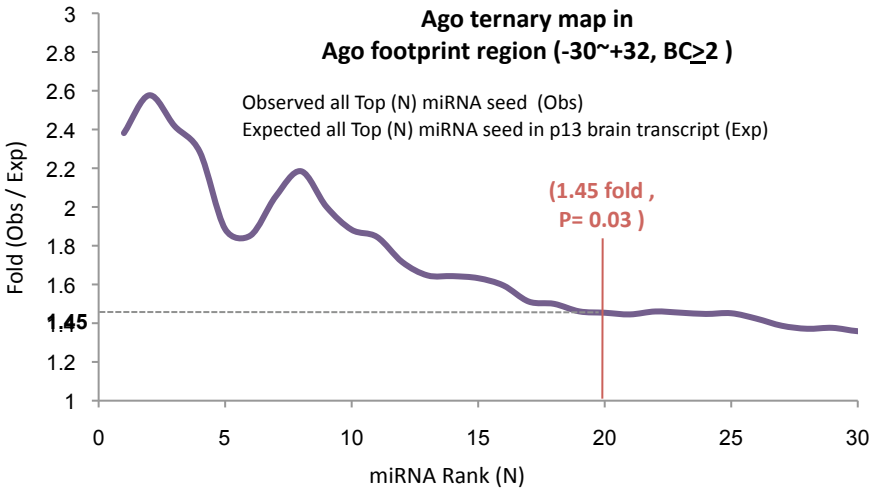
Without conservation

A



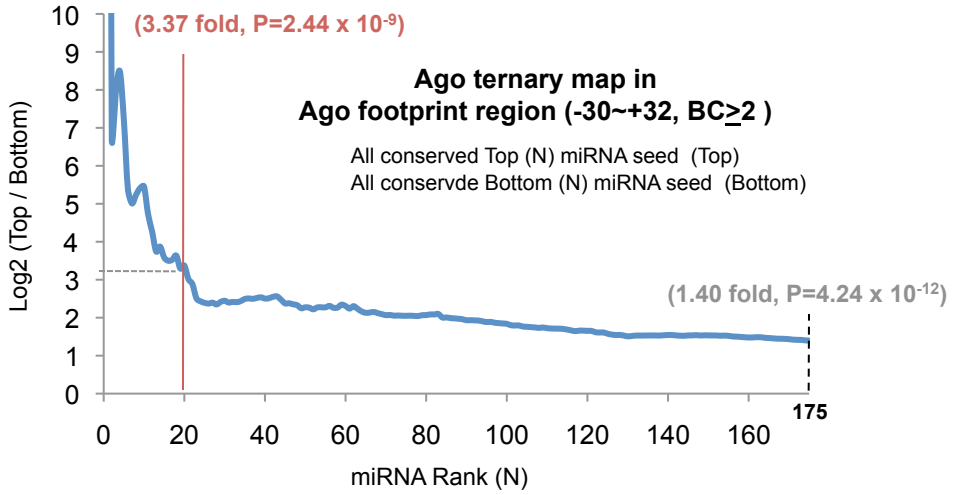
Without conservation

B

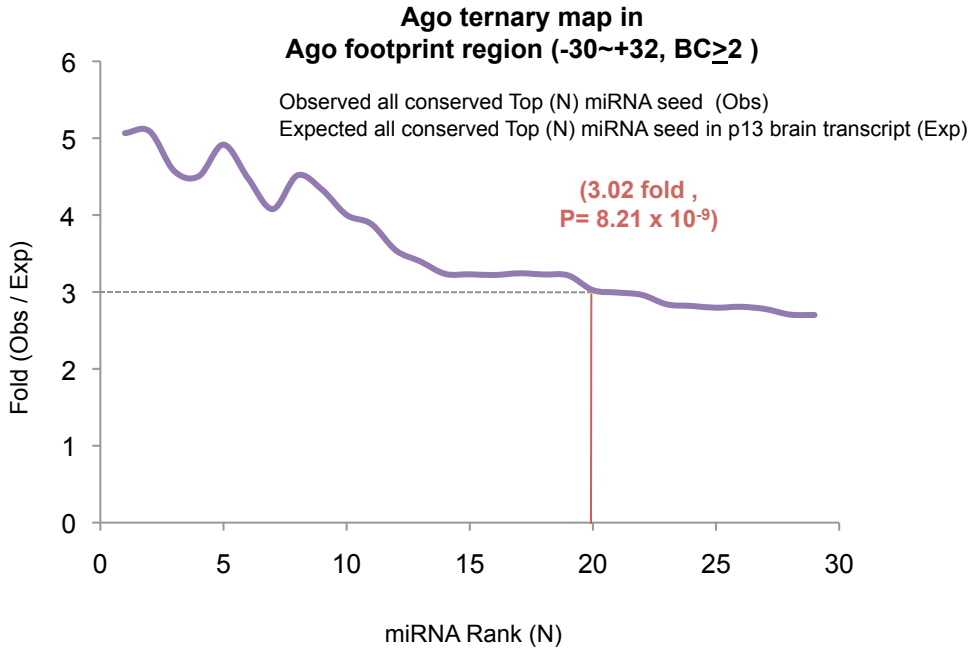


Supplemental Figure 13

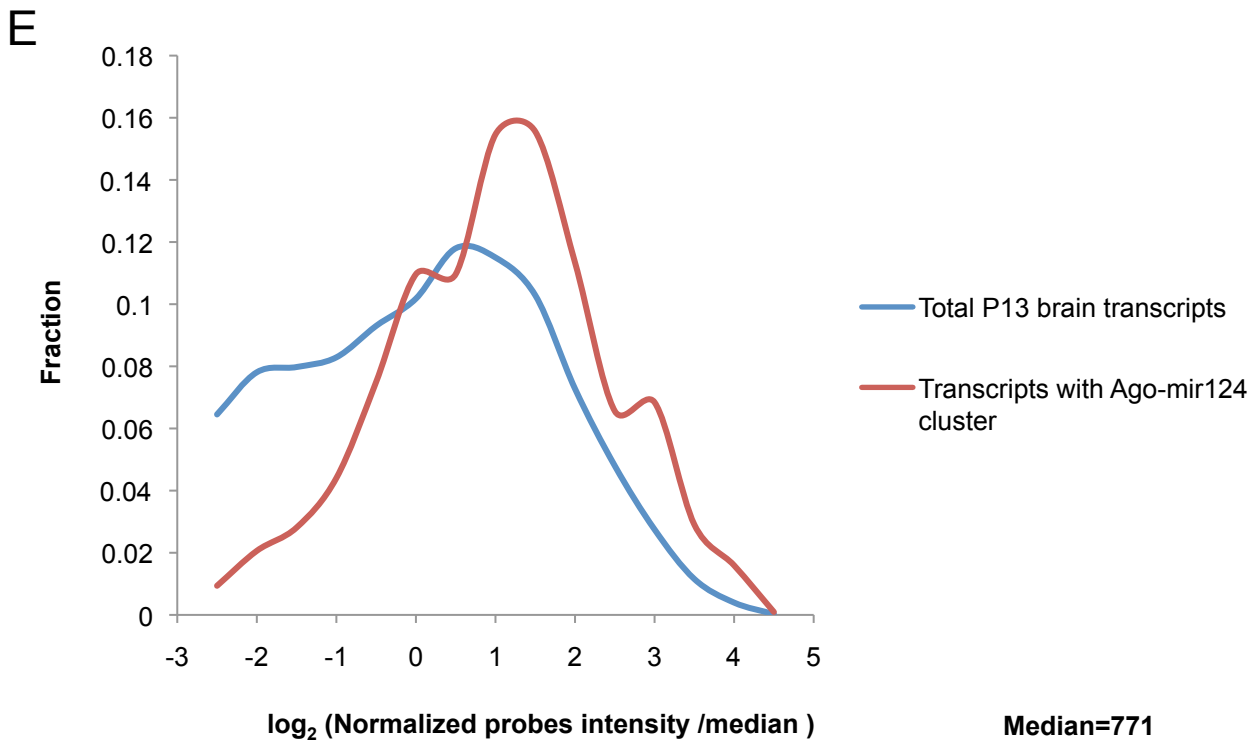
C



D



Supplemental Figure 13



Supplemental Figure 14

A

BC>=2

Number of Ago-mRNA clusters in 3'UTR	Number of 3'UTRs
1	1665
2	584
3	262
4	130
5	87
6	49
7	34
8	27
9	17
10	9
>10	43

Averages:
1.03 clusters/transcript in brain
2.60 clusters/Ago regulated transcript
2.12 clusters /3'UTR of Ago regulated transcript

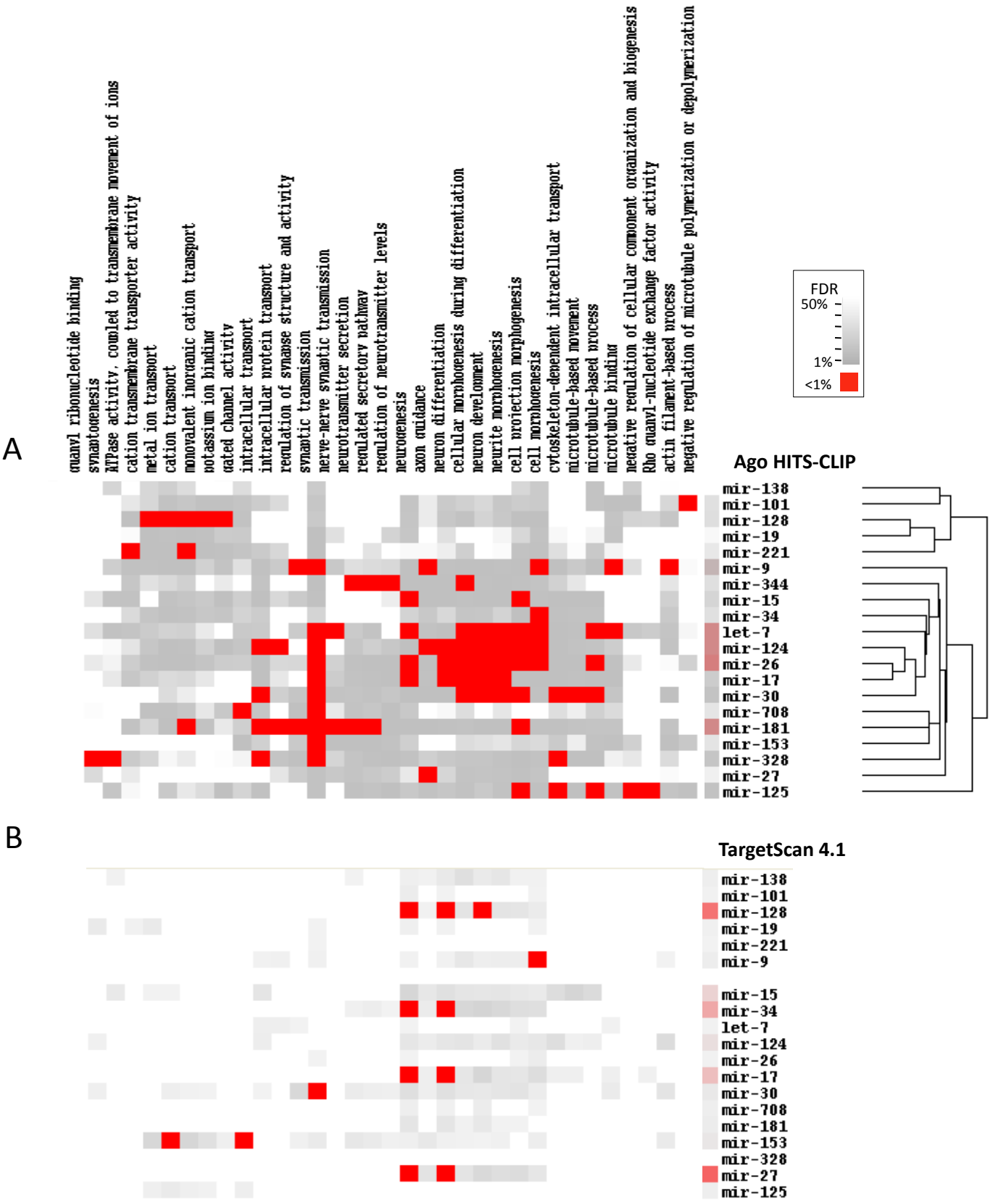
B

miRNA family	Number of target transcripts
miR-124	1254
miR-9	1207
miR-27	1080
let-7	1060
miR-15	787
miR-181	740
miR-101	631
miR-34	626
miR-30	610
miR-26	602
miR-138	573
miR-17	536
miR-125	517
miR-708	498
miR-19	486
miR-153	455
miR-193	448
miR-21	358
miR-344	320
miR-221	315

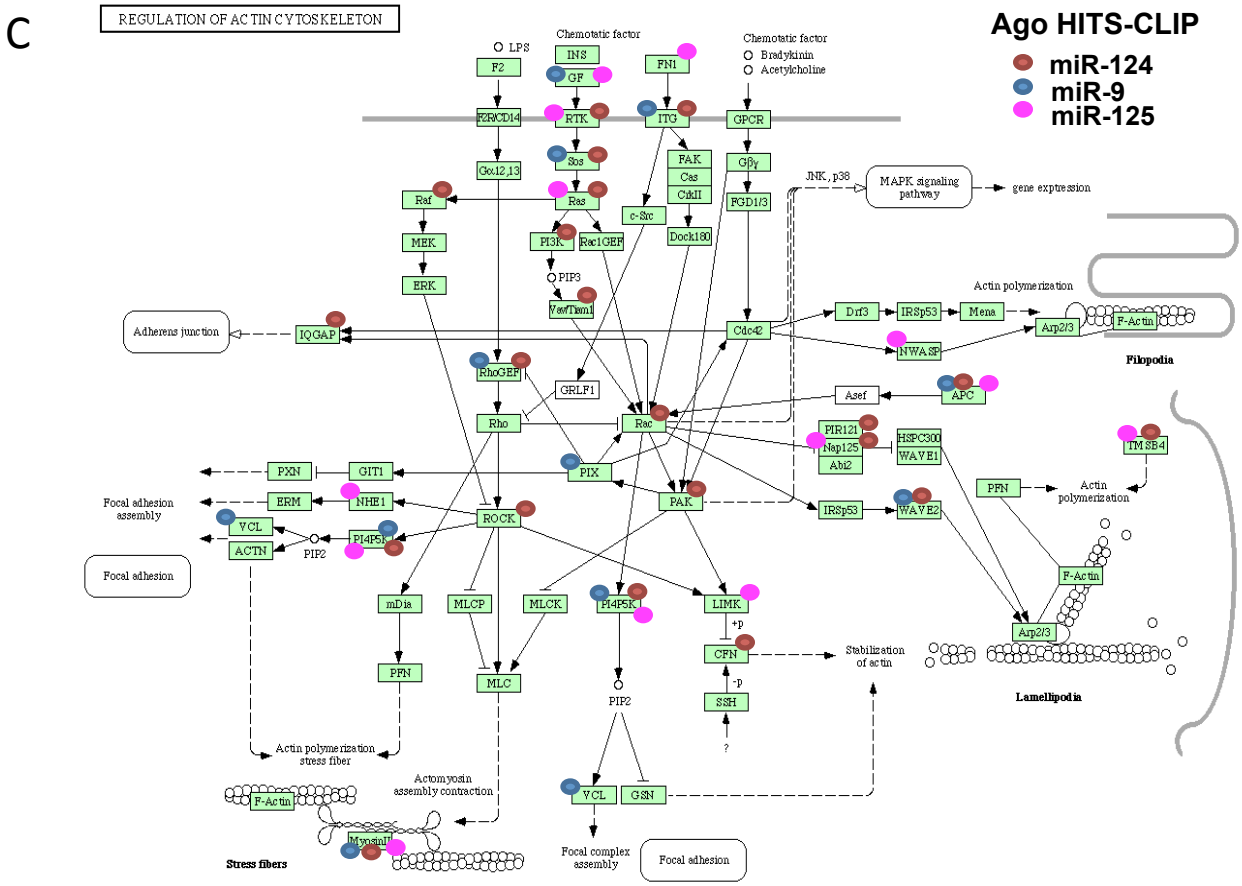
Average 655 targets / miRNA

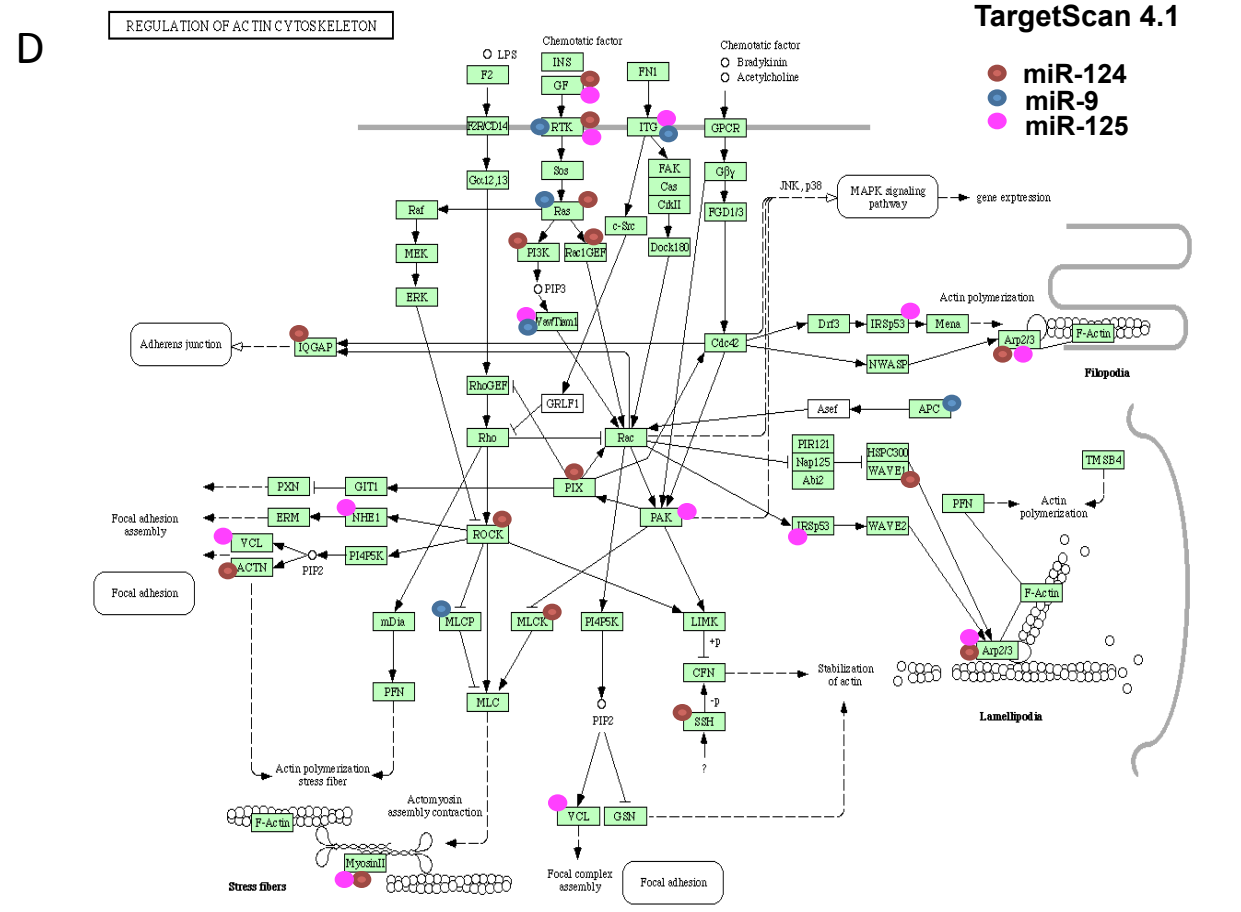
C

Supplemental Figure 15



Supplemental Figure 15



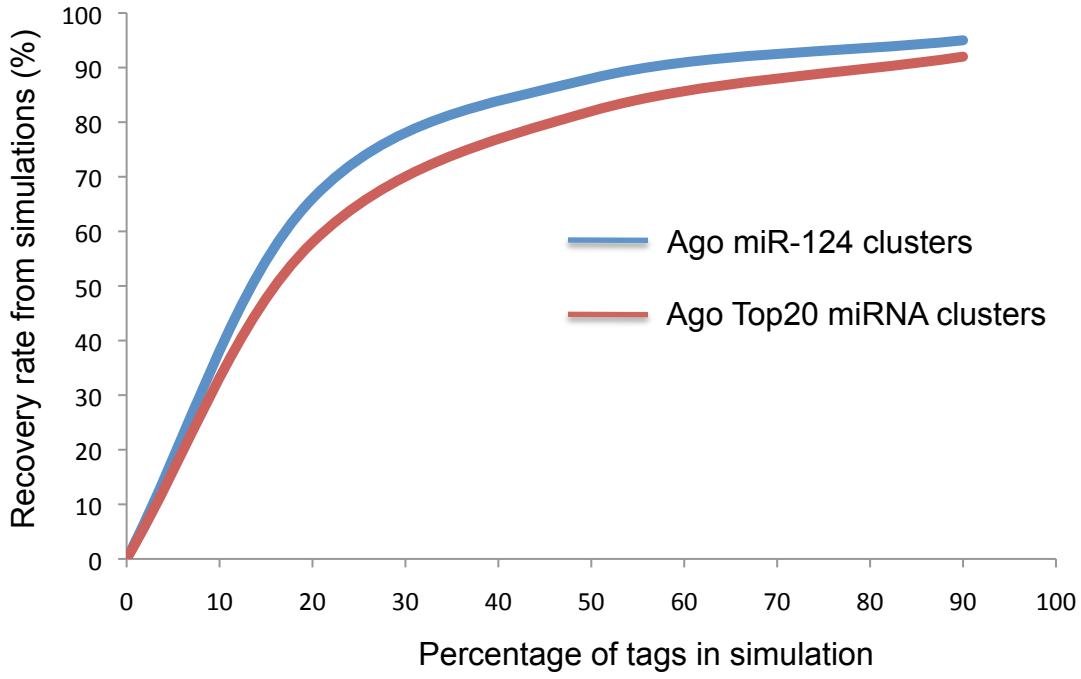


Supplemental Figure 16

A

	Number of Ago mir-124 clusters	Number of Ago Top20 miRNA clusters
Ago-CLIP	1561	15665
Average # of clusters in 10 simulations	1487	14410.7
Standard deviation	2.05	16.57
simulation #1	1484	14401
simulation #2	1490	14417
simulation #3	1489	14401
simulation #4	1488	14440
simulation #5	1484	14413
simulation #6	1487	14408
simulation #7	1487	14388
simulation #8	1488	14412
simulation #9	1488	14393
simulation #10	1485	14434
recovery rate(%)	95	92

B



Supplementary References:

1. Jensen, K. B. & Darnell, R. B. CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol Biol* **488**, 85-98 (2008).
2. Licatalosi, D. D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-469 (2008).
3. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **37**, 376-386 (2005).
4. 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).
5. 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).
6. 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).
7. Baek, D. et al. The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).
8. Selbach, M. et al. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58-63 (2008).
9. Mourelatos, Z. Small RNAs: The seeds of silence. *Nature* **455**, 44-45 (2008).
10. Hammell, M. et al. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat Methods* **5**, 813-819 (2008).
11. Easow, G., Teleman, A. A. & Cohen, S. M. Isolation of microRNA targets by miRNP immunopurification. *RNA* **13**, 1198-1204 (2007).
12. Karginov, F. V. et al. A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci U S A* **104**, 19291-19296 (2007).
13. 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).

14. Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401-1414 (2007).

15. Lu, J. et al. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838 (2005).

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

17. 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).

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

19. 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).

20. 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).

21. 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).

22. 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).

23. 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).

Argonaute HITS-CLIP detailed method

(HITS-CLIP: HIgh-Throughput Sequencing of RNAs from *in vivo* Cross-Linking and Immuno-Precipitation)

Sung Wook Chi, Julie B. Zang, Aldo Mele & Robert B. Darnell (June, 2009)

Day 1

I. General method for UV cross-linking of tissue/cell lines

For mouse tissue:
Harvest necortex from P13 mice and let tissue sit in ice cold HBSS until harvest is complete.

[In our case, half of P13 neocortex tissue suffices for high quality data using anti-Ago antibodies. If cell culture is used, HBSS can be used instead of PBS.]

Add 10 cell volumes of HBSS and triturate tissue first using a 5 or 10 ml pipette, and then again by adding a 100μl micropipette tip at the end of the pipette.

[Because UV light can penetrate a few cell layers, stringent trituration to the single-cell is not necessary.]

Irradiate suspension (using 10mls per 10 cm tissue culture plate) three times for 400mJ/cm^2 in Stratalinker (we use model 2400 from Stratagene). Mix suspension between each irradiation.

[The length of crosslinking should be optimized for each protein separately, as each RNA binding domain crosslinks with different efficiency, depending on the availability of aromatic amino acids. For a preliminary experiment, try 100, 200 and 400mJ/cm^2, and then use the shortest condition that gives >70% of the maximum signal.]

For cell culture:
Grow cells in a 100 or 150mm plates, rinse once with PBS, and place in Stratalinker with the cover off. Irradiate one time for 400mJ/cm^2 and additional 200 mJ/cm^2 in Stratalinker.

[Because cells grown in a single layer are all equally exposed to UV light, a single UV and additional half energy UV irradiations are sufficient.]

Collect suspension, pellet cells at 2500rpm for 5 min at 4°, re-suspend pellet in (~3x dry volume) of PBS and distribute 1ml of suspension to each eppie; quick spin at 4°, remove supernatant and freeze pellets at -80° until use (each tube is about 200 μl of cells).

To prepare **1x HBSS**:
50 ml 10x Hank's Balanced salt solution, Ca-Mg-free (Gibco, #14186-012)
5 ml 1M HEPES, pH 7.3
445 ml ddH_2O

Day 2

II. Immunoprecipitation

a. Solutions (to be made fresh)

1X PXL (Wash Buffer)
1X PBS (tissue culture grade; no Mg++, no Ca++)
0.1% SDS
0.5% deoxycholate
0.5% NP-40

5X PXL (High-salt Wash Buffer)
5X PBS (tissue culture grade; no Mg++, no Ca++)
0.1% SDS
0.5% deoxycholate
0.5% NP-40

1X PNK Buffer
50 mM Tris-Cl pH 7.4
10 mM $MgCl_2$
0.5% NP-40

1X PNK+EGTA Buffer
50 mM Tris-Cl pH 7.4
20 mM EGTA
0.5% NP-40

b. Bead preparation:

For each Eppie of crosslinked lysate use 400 μl of protein A Dynabeads (Dynal, 100.02). Wash beads 3x with 0.1 M Na-phosphate, pH 8.0.

[We use high pH here because this increases antibody capture by Protein A. However, you may need to adapt all the immunoprecipitation buffers to the antibody that you use. For instance, we found that some antibodies don't work with the deoxycholate in the wash buffer.]

Resuspend beads in 350 μl 0.1 M Na-phosphate pH 8.0 and add 50 μl bridging Ab (2.4mg/ml; rabbit anti-mouse IgG; Jackson Immunoresearch: 315-005-008)

[Ago antibodies(2A8 or 7G1-1*) we use are mouse monoclonal. So we pre-incubate the beads with rabbit anti-mouse IgG to increase the avidity.]

Rotate tubes at room temperature for ~45 min.; wash 3x with 0.1 M Na-phosphate, pH 8.0.

Resuspend beads in 400 μl 0.1 M Na-phosphate pH 8.1 and add 3 μl of 2A8 or 12 μl of 7G1-1* ascite fluid (5mg/ml).

[This protocol is optimized for Ago HITS-CLIP using half of neocortex from P13 mouse. If you are using less tissue, you can scale down the amount of bead and antibody used here.]

Rotate tubes at 4°C for ~4 hrs.; wash 3x with 1X PXL; if you are not yet ready to add crosslinked lysate, leave beads in last wash step.

c. 5' end labeling of RL3 (-P) linker (no phosphate)

[Prepare labeled 5'end linker before or during IP. We use pre-labeled 3'linker for AGO HITS-CLIP which gives more specific signal in autoradiogram than labeling isolated RNA after IP by PNK reaction]

2.4 μl RL3(-P) linker (50pmol/μl ; no phosphate in 5'end)
5 μl 10X PNK Buffer (NEB)
25 μl ^{32}P-γ-ATP
8 μl T4 PNK enzyme (NEB, M0201L)
3 μl RNAsin (Promega)
6.6 μl water
50 μl total

Incubate in Thermomixer R (Eppendorf) at 37° for 30 minutes
Add 0.2μl of 10mM ATP, and let the reaction go for an additional 5 minutes.

Resuspend the resin in the G-25 column (GE healthcare:27-5325-01) by vortexing.
Pre-spin the column for 1min at 735 x g, apply the sample to resin, and spin the column for 2min at 735 x g; if you are not yet ready to use, store it at -20°C.

[This protocol is for 10 samples; 5 µl for each sample. If you are using less than 10 samples, you can scale down the amount of materials used here.]

d. Partial RNA digestion and ultracentrifugation

Resuspend each tube of crosslinked lysate using ~700 μl of 1X PXL (with protease inhibitors; ~1 ml total volume); add 15 μl RNAsin (Promega) and let sit on ice for 10 min.

Add 30 μl of RQ1 DNAse (Promega, M6101) to each tube; incubate at 37° for 5 min, 1000 rpm.

[DNAse 1 is not supposed to work in the absence of Ca and Mg (ideally, it requires 25 mM $MgCl_2$ and 5 mM $CaCl_2$). However, at this step after cell lysis, the contaminating Ca++ and Mg++ from the cell lysate and high concentration of the DNAse are sufficient to digest the DNA.]

Make a dilution of RNAse A (USB 70194Y) at 1:100 in 1X PXL (high-RNAse).
Make a dilution of RNAse A (USB 70194Y) at 1:10,000 in 1X PXL (low-RNAse).

[Each experiment should be done in duplicate with two RNAse concentrations – the dilution

depends also on the batch of RNAse, so in the first experiment, several dilutions should be tested. The over-digest experiment is used as a control to confirm that the size of the radioactive band on SDS-PAGE gel changes in response to different RNAse concentrations (which confirms that the band corresponds to a protein-RNA complex). In addition, this experiments helps determine the size of the immunoprecipitated RNA-binding proteins, because these proteins will be bound to short RNAs and will thus migrate as less diffuse bands only ~7kDa above the expected MW.]

Add 10μl of each RNAse dilution to the two duplicate tubes; incubate at 37° for 5 minutes.

Spin lysates in pre-chilled ultra-microcentrifuge (polycarbonate tubes in TLA 120.2 rotor), 30K for 20' at 4°C.

[You should determine the optimal speed for your protein. Higher speed (90K) is beneficial to clear the lysate of all high-molecular weight material, like ribosomes or very large RNPs, but in that case you need to make sure your protein stays in supernatant.]

Carefully remove the supernatant, and save 10μl for immunoblot analysis.

e. Immunoprecipitation

Add the rest of supernatant to one prepared tube of beads. Rotate beads/lysate mix for 2-4 hours at 4°.

Remove the supernatant and save 10μl for immunoblot analysis (in order to test the relative depletion of the antigen).

Wash beads with ice-cold buffer:
2x 1X PXL (Wash Buffer)
2x 5X PXL (High-salt Wash Buffer)
2x 1X PNK Buffer

[The first time CLIP is done for a particular protein, you can skip steps III-IV, and go straight to step V. Continue until the exposure of membrane to X-ray film and check the following:
1. Do I get radioactive band ~7kDa above the MW of my protein in high RNAse experiment?
2. Does my band disappear in the control experiments? Possible controls are no UV crosslink, irrelevant antibody pulldown, knockout organism, RNAi knockdown, or no transfection of a tagged construct.
3. Does this band move up and become more diffuse in low-RNAse experiment? If yes, you are dealing with an RNA-binding protein, and you can in your following experiments proceed to step III.
[Continue to step III with the sample that contained low RNAse (1:300,000). Skip steps III and IV with the over-digest sample – leave it at 4° until step V.]

III. CIP Treatment (On-Bead)

8 μl 10x dephosphorylation buffer (Roche, 712023)
3 μl alkaline phosphatase (Roche, 712023)
2 μl RNAsin (Promega)
67 μl water

80 µl total (per 400 µl of beads)

Incubate in Thermomixer R (Eppendorf) at 37° for 20 min (1000 rpm every 2 minutes for 15 seconds).

Wash 1X with 1X PNK Buffer
Wash 1X with 1X PNK+EGTA Buffer
Wash 2X with 1X PNK Buffer

IV. 3' RNA Linker Ligation (On-Bead)

8 µl	10X T4 RNA ligase buffer (Fermentas)
8 µl	BSA (0.2 µg/µl)
8 µl	ATP (10 mM)
2 µl	T4 RNA ligase (Fermentas)
2 µl	RNAsin (Promega)
5 µl	Labeled hot RL3 (12pmol, Prepared in IIc)
47 µl	water
80 µl	total

Add 80 µl of ligase mix to each tube of beads.

Incubate at 16°C for 1hr in Thermomixer R (Eppendorf) (1000 rpm every 2 minutes for 15 seconds).

Add 4µl of RL3(20pmol/µl; with Phosphate on 5'end) , and let the reaction go for overnight.

Day 3

Wash 1x with 1X PXL buffer
Wash 1x with 5X PXL buffer
Wash 3x with 1X PNK buffer

V. PNK Treatment (On-Bead)

8 µl	10X PNK Buffer (NEB)
1 µl	cold ATP (10mM)
4µl	T4 PNK enzyme (NEB, M0201L)
2 µl	RNAsin (Promega)
65 µl	water
80 µl	total

Add 80 µl of PNK mix to each sample and incubate in Thermomixer R (Eppendorf) at 37° for 20 minutes (1000 rpm every 4 minutes for 15 seconds).

Wash 1x with 1X PXL buffer
Wash 1x with 5X PXL buffer
Wash 3x with 1X PNK buffer

VI. SDS-PAGE & nitrocellulose transfer

[Do western blot for input and post-IP; load equal volumes.]

Re-suspend the beads in 30 μl of 1X PNK+ and 30 μl of Novex loading buffer (without reducing agent).

[Antibody bands may interfere in the way your protein runs on the gel; if the MW of your protein is less than 50kDa, DO NOT add any reducing agent.]

[The Novex NuPage gels are critical. A pour-your-own SDS-PAGE gel (Laemmli) has a pH during the run that can get to ~9.5 and can lead to alkaline hydrolysis of the RNA. The Novex NuPAGE buffer system is close to pH 7.]

Load 1 tube per 2 wells of a 10 well Novex NuPAGE 10% Bis-Tris gel. Run the gel at 175V in the cold room.

[Some pre-stained molecular weight markers may run differently on Novex NuPage gels. We use rainbow marker (Amersham, RPN800), which runs at the expected molecular weights.]

[The gel itself will be a lot less hot than the amount of radioactivity that you loaded. Most of the signal will be in the lower buffer, which results from free ATP and small free RNAs that run out of the gel. Free RNA of up to 100 nucleotide length will migrate on the gel below 30 kDa.]

After gel run, transfer gel to S&S BA-85 nitrocellulose using the Novex wet transfer apparatus. Transfer 1 h at 30V in NuPAGE Transfer Buffer with 10% methanol.

[This pure nitrocellulose is a little fragile, but it works better for the RNA/protein extraction step.]

After transfer, rinse the nitrocellulose filter in 1X PBS, and gently blot on Kimwipes; wrap membrane in plastic wrap and expose to autoradiogram.

[Most free RNA will pass through the membrane, so due to the loss of free RNA in the transfer step, the membrane will be less hot than the gel.]

[Use a luminescent sticker, so that you can later align the filter back to the autorad.]

Day 4

[In case of Ago, we usually see a signal on the autoradiogram after ~2 hours exposure if half of neocortex from p13 brain tissue was used as starting material. If much less material is used, an overnight exposure will sometimes be necessary to see a decent signal.]

We don't clone RNAs from the over-digested sample, but use it to determine the specificity of the RNA-protein complexes by the following steps:

1. ***Look at the over-digested sample, and see if you have a band ~7kDa above the expected MW of your protein (~100kDa for Ago).***
2. ***Calculate the distance from the closest contaminating band, and estimate the relative signal strength of your protein relative to other bands.***
3. ***You should have no problem purifying RNA specific for your protein if your protein band***

migrates more than 10kDa away from any other protein band.

The RNA-protein complexes that were digested by low RNAse will appear as a diffuse radioactivity with a modal size of ~15-20kDa above the expected MW of your protein.

[In case of Ago, we usually see two different modal sizes. One is ~110kDa from Ago-miRNAs complex and the other is ~130kDa from Ago-mRNA complex]

4. Average MW of 50 nucleotides long RNA is ~16 kDa. As the tags contain 20 nucleotides long linker (L3), the position of protein-RNA complex that will generate CLIP tags longer than 50 nucleotides is 20 kDa above the expected MW of the protein.
5. Because the RNAse digestion is random, the tag sizes will vary from ~50-150 nucleotides, and thus the complex will migrate more diffuse than the low RNAse complex (where the tag sizes are ~20-60 nucleotides).
6. In order to be able to later separate CLIP tags specific to different proteins, you need to cut a band as thin as possible (~3kDa wide band) approximately 20 kDa above the expected MW of your protein.
7. In addition, cut also two bands below and above this band, with ~15 kDa distance. The lower band should not contain any RNA longer than 15 nucleotides specific to your protein (but it will show RNA bound to any smaller proteins).

Cut out thin bands (~4-8kDa wide) using a clean scalpel blade, and put the nitrocellulose pieces into separate microtubes. The smaller the pieces, the better. Count radioactivity in a scintillation counter.

VII. RNA Isolation and Purification

1X PK Buffer:
100 mM Tris-Cl pH 7.5
50 mM NaCl
10 mM EDTA

1X PK Buffer/7M urea (this buffer must be fresh):
100 mM Tris-Cl pH 7.5
50mM NaCl
10 mM EDTA
7 M urea

Make a 4mg/ml proteinase K (Roche, 1373196) solution in 1X PK Buffer; pre-incubate this stock at 37° for 20 min to kill any RNAses.

Add 200 μl of proteinase K solution to each tube of isolated nitrocellulose pieces; incubate 20 min at 37° at 1000 rpm.

Add 200 μl 1x PK/7M urea solution; incubate another 20 min at 37° at 1000 rpm.

Add 400 μl RNA phenol (Ambion, 9710) and 130 μl of $CHCl_3$ to solution; 37° for 20' at 1000 rpm.

[RNA phenol can also be prepared by equilibrating pure phenol with 0.15 M NaOAc pH 5.2; $CHCl_3$

is chloroform 49:1 with isoamyl alcohol.]

Spin tubes at full speed in microcentrifuge; take aqueous phase Add 50 μl 3M NaOAc pH 5.2, 0.75 μl of glycogen (Ambion, 9510) and 1ml of 1:1 EtOH:isopropanol.

[0.5 µl of glycogen is necessary to precipitate small quantity of RNA, but don't add more, otherwise the RNA ligase may be inhibited.]

Precipitate overnight at -20°.

Day 5

VIII. 5' RNA Linker Ligation

Spin down RNA (10' at max speed in microcentrifuge). Check to see if you got decent precipitation of counts. Wash and dry pellet.

[It is important to wash well, as residual salt can decrease ligation efficiency. We recommend two washes with 150 µl cold 75% ethanol. One can vortex the second wash and spin down the RNA again for 10' at max speed in microcentrifuge. Don't over-dry – 2' in speedvac is usually enough.]

Count RNA in scintillation counter. Resuspend in 5.9 μl H_2O.

RNA ligation:
1 μl 10X T4 RNA ligase buffer (Fermentas) 1 μl BSA (0.2 μg/μl)
1 μl ATP (10 mM)
0.1 μl T4 RNA ligase (3U, Fermentas)
1 μl RL5 or RL5D RNA linker @ 20 pmol/μl
4.1 μl

Add 5.9 μl RNA resuspended in H_2O
10 μl total

[The linker itself cannot circularize (it has 5'-OH and 3'-OH), and the CLIP tag-3' linker product cannot circularize (It has 5'-P and 3'-Pmn).]

Incubate at 16° for 1-5 hours (or overnight).

Add to the reaction:
79 μl H2O
11 μl 10X DNAse I Buffer
5 μl RNAsin
5 μl RQ1 DNAse

Incubate 37° for 20 minutes.

Add:
300 μl H_2O
300 μl "RNA phenol"
100 μl $CHCl_3$

Vortex, spin and take aqueous layer.

Precipitate by adding: 50 μl 3M NaOAc pH 5.2
0.5μl of glycogen (Ambion, 9510)
1 ml 1:1 EtOH:isopropanol

Precipitate overnight (or 1h) @ -20°.

Day 6

IX. RT-PCR.

Spin down the RNA. Wash and dry the pellet. Count RNA in scintillation counter to see if you got decent precipitation of counts. Re-suspend in 8 μl of H_2O.

RT reaction:
Mix 8 μl of the ligated RNA and 2 μl of DP3 (5 pmol/μl – i.e. 1 pmol/μl final concentration) and 3 µl 3 mM dNTPs.

Heat 65° for 5 min; chill and quick spin.

Add:
1 μl 0.1 M DTT
4 μl 5X SuperScript RT Buffer
1μl RNAsin
1 μl SuperScript III (Invitrogen, 18080-044)
20 μl total

Incubate at 50° for 45 min., 55° for 15min., 90° for 5 min., leave at 4°.

PCR reaction:
27 μl Accuprime Pfx Supermix (Invitrogen, 12344-040)
0.75 μl DP5 primer, 20 pmol/μl
0.75 μl DP3 primer, 20 pmol/μl
2 μl of the RT reaction
30.5 μl total

Settings: 95° 2' Cycle 25-35x (depending on how much RNA you started with): 95° 20'' / 58° 30'' / 68° 20'' 68° 5'.

[Do each tube in duplicate. Run one sample and keep the second one as backup.]

Day 7

Pour a 10% denaturing poly-acrylamide gel. We use Owl vertical electrophoresis system (P9DS-2).
In this case, we prepare 20ml of the following for each gel:
8.4 g urea (Fisher, U15-3)
5 ml 40% 19:1 Acrylamide:Bis-acrylamide (Fisher)
4 ml 5x TBE
water to 20 ml total

immediately before pouring the gel, add:
200 μl 10% APS
7.5 μl TEMED

We use the following 2x loading buffer for polyacrylamide gels:
95% formamide, deionized (Sigma F-9037)
5% 100mM EDTA, pH 8
Bromophenol blue + Xylene cyanol (Sigma B-3269)

Run the entire PCR reaction; use low molecular weight markers (we use 3 μl of Amplisize Molecular Ruler, Biorad – don't heat it), visualize DNA by immersing the gel for 10 minutes in 10.000-fold dilution of SYBR Gold (Molecular Probes) in TBE.

[If 3kDa wide band was cut from the membrane with the SDS-PAGE gel separated complexes, the resulting DNA tags will vary about 10-20nt in size. In case of the band that was cut ~20kDa above the expected MW of your protein, the DNA amplified from the RNA tags specific for your protein will appear as a band migrating between ~80-100 nucleotides (average RNA insert size ~50 nucleotides + 36 nucleotides for the linkers)]

[In case of Ago, we usually see two different bands here. One is ~56 nucleotides (20+36) and the other is ~80-100 nucleotides (50+36). Based on the sequence results, we confirmed that majority of ~56 nucleotides are from miRNAs and ~80-100 nucleotides are from target RNAs. In case of band that was cut ~110kD, we usually see ~56 nucleotides but from ~130kD, we usually see both ~56 and ~80-100 nucleotides bands.]

[When analyzing the PCR gel, you need to go back and look at the SDS-PAGE gel and find out how many bands migrated faster than your protein-RNA complex.]

Cut out the DNA of 80-100nt, extract DNA with QIAquick Gel Extraction Kit (follow "user-developed" protocol for extracting DNA fragments from polyacrylamide gels).

X. Re-PCR with Solexa Fusion Primers

PCR reaction:

27 µl Accuprime Pfx Supermix (Invitrogen, 12344-040)
0.5 µl DSFP5 primer, 20 pmol/µl
0.5 µl DSFP3 primer, 20 pmol/µl
3µl of the RT reaction
31 µl total

Settings: 95° 2'
Cycle 6-14x (depending on how much DNA you started with): 95° 20'' / 58° 30'' / 68° 40''
68° 5'

Pour a 2% agarose (Nusieve or MetaPhor) / Ethidium Bromide gel.
Run entire PCR product; use low molecular weight markers (we use 5 µl of Amplisize Molecular Ruler). Visualize DNA on UV box. Cut out DNA of 150-170 nt, extract DNA with QIAquick Gel Extraction Kit following protocol for extracting DNA fragments from agarose gels).

XI. Quantitation of DNA

We use the Quant-it DNA Assay Kit, High Sensitivity (Invitrogen, Q-33120) to determine DNA concentration (follow manufacturer's instructions).
Make 10-30 µl of 10 nM DNA per sample; Submit for Solexa Sequencing.

XII. Linker and primer sequences

[The RNA linkers need to be gel purified. Run 50 µl of 500 µM stock of deprotected RNA on 20% polyacrylamide gel, visualize the RNA by UV shading, cut out the band and purify as described above.]

RNA linkers (from Dharmacon):

RL5: 5'-OH AGG GAG GAC GAU GCG G 3'-OH
RL5D: 5'-OH AGG GAG GAC GAU GCG Gr(N)r(N) r(N)r(N)G 3'-OH
RL3: 5'-P GUG UCA GUC ACU UCC AGC GG 3'-puromycin
RL3(-P): 5'-OH GUG UCA GUC ACU UCC AGC GG 3'-puromycin

DNA primers (from Operon):

DP5: 5'-AGG GAG GAC GAT GCG G-3'
DP3: 5'-CCG CTG GAA GTG ACT GAC AC-3'

Solexa Fusion Primers (from Operon):

DSFP5: 5'-AATGATACGGCGACCACCGACTATGGATACTTAGTCAGGGAGGACGATGCGG-3'
DSFP3: 5'-CAAGCAGAAGACGGCATACGACCGCTGGAAGTGACTGACAC- 3'

Solexa Sequencing Primer (from Operon):

SSP1: 5'-CTA TGG ATA CTT AGT CAG GGA GGA CGA TGC GG-3'