Molecular Cell,Volume *48*

Supplemental Information

Transcriptome-wide miR-155 Binding Map Reveals Widespread Noncanonical MicroRNA Targeting

Gabriel B. Loeb, Aly A. Khan, David Canner, Joseph B. Hiatt, Jay Shendure, Robert B. Darnell, Christina S. Leslie, and Alexander Y. Rudensky

Supplemental Experimental Procedures

mRNA expression

Processing mRNA microarray data

RNA from 6 biological replicates of CD4+ T cells from WT and 155KO mice (3 samples each) were labeled and hybridized to Mouse Genome 430A 2.0 Arrays (430A2). All 430A2 samples were normalized with robust multi-array analysis (RMA) normalization using the "affy" Bioconductor package in R. Gene level expression was summarized from mean probe set level intensities using Mouse430A_2 Annotations, Release 31 (08/11/10). *log(expression change)* was calculated using the WT expression level as reference. We used processed mRNA expression levels as provided by the authors of previously published WT and miR-17~92KO datasets (Mu et al., 2009). All log gene expression changes were centered with mean *log(expression change)* and normalized to have unit variance. This normalization resulted in a simple Z-transformation of *log(expression change)*.

KS statistics for mRNA expression changes

In order to statistically evaluate mRNA expression changes between two gene sets, we compared their distributions of Z-transformed *log(expression change)* using a one-sided Kolmogorov-Smirnov (KS) test statistic. The KS statistic assesses whether the distribution of expression changes for one set (predicted RNA target genes) is significantly shifted upwards (induced) compared to the distribution for the other set (background set of all genes). The KS statistic computes the maximum difference in value of the empirical cumulative distribution functions (cdfs):

$$\sup_x \left(F_1(x) - F_2(x)\right),$$

where $F_j(x) = \frac{1}{n_j}\sum_{i=1}^{n_j} I_{X_i \le x}$ is the empirical cdf for gene set j = 1, 2, based on n_j (Z-transformed) *log(expression change)* values and $I_{X_i \le x}$ is the indicator function, equal to 1 if the ith observation for gene set j is less than or equal to x, 0 otherwise. We used the Matlab function kstest2 to calculate the KS test statistic and asymptotic P-value.

Processing CLIP sequence reads

Multiplex index and degenerate barcode tag

CLIP libraries were designed with an 11nt index read. Positions 1-5 of the index read are a 5 nucleotide multiplexing index, and positions 6-11 are a 6 nucleotide degenerate barcode. The reads were demultiplexed and assigned into specific sample libraries (along with the corresponding degenerate barcode for each CLIP read, which was appended to the 5' end of the CLIP read for convenience). To account for possible sequencing errors, we allowed up to 1 mismatch for assignment to a sample library. Because each index sequence is designed to differ by at least 3nt, allowing 1 mismatch safely avoids cross-library misassignment. Unassigned CLIP sequence reads were discarded.

Design of the multiplex indexes

To perform sample multiplexing on the Illumina sequencer, we designed many unique index sequences. These multiplex indexes were designed using a greedy lexicographic code construction method (Conway and Sloane, 1986). First, all 4^5 DNA tags of length 5 are generated using the alphabet {A, C, G, T}. Second, all sequences are ordered by a strict lexicographic ordering: A < C < G < T. Third, all low-complexity sequences containing fewer than 3 different nucleotides are removed. Initially, a random sequence is selected from the ordering and placed in a set denoted *L*. Next, starting at the beginning of the ordering, each sequence with a Hamming distance >2 from every sequence in *L* is also placed in to *L*. We repeat this last step until we reach the end of the lexicographic ordering. As a result, we constructed 24 unique index sequences that differ by at least 3nt. Since these indexes can distinguish reads from individual libraries, this strategy allows multiple libraries to be sequenced in the same lane of an Illumina GA2 or HiSeq2000 sequencer.

CLIP sequence reads

Once the pool of CLIP sequences were deconvolved and assigned to sample-specific libraries, reads were processed uniformly in the following manner. 1) The 3' ends of reads were trimmed until the base quality score was >20. 2) Linker sequences were removed (3' linker1 = GTGTCTTTACACAGCTACGGCGTCG; 3' linker2 = GTGTCACTGATAGCAACCCGGTGCT; 5' end of 5' linker = CGACCAGCAT). 3) Reads were consolidated; that is, we merged identical reads with identical barcodes. To implement these steps, we used the fastx-toolkit (Hannon laboratory, CSHL; http://hannonlab.cshl.edu/fastx-toolkit/) and custom Perl scripts.

Mapping CLIP sequence reads

We mapped CLIP reads to genes using a two step alignment process. First, we aligned reads that mapped uniquely to the mm9 Mouse genome with up to one mismatch. To implement this, we used Bowtie (Langmead et al., 2009) with parameters: (mm9 -5 6 -v 1 --tryhard -a -m 1 --best --strata) and where mm9 is the reference mouse genome (we used -5 6, which trims the first six nucleotides of the alignment, because we had appended the degenerate barcode as the 5' 6 nucleotides of the read) . Second, we realigned the uniquely mapped reads to a curated 3'UTR database. We used Bowtie with parameters: (mouse.utr -5 6 -v 1 --norc --tryhard -a

-m 1 --best --strata) and where mouse.utr is a custom non-redundant 3'UTR database.

3'UTR database generation

We constructed our non-redundant 3'UTR sequence database using sequences from RefSeq and Ensembl (Aug 15 2011). These sequences were processed with the following steps.

1) We retrieved all 3'UTR annotations from RefSeq and Ensembl.

2) We selected the longest 3'UTR sequence based on gene name. Gene annotation information was parsed from raw data files using custom scripts written with BioPython libraries (Version 4; http://biopython.org).

Removal of PCR-amplification bias

In order to reduce PCR bias, a 6nt degenerate barcode (random 6mer) was included in the RT primer during library construction. As mentioned above, this barcode is sequenced as part of an index read. We counted the number of unique barcodes associated with reads that mapped to the same starting position. This integer count of unique barcodes is used to estimate relative AGO binding and is defined as the "read count". This results in an unbiased lower bound estimate of AGO binding and is the value used for calculating log-scale changes in miRNA-mRNA interactions between WT and 155KO libraries.

Mapping putative crosslink deletion sites

In order to precisely map AGO-mRNA interaction sites, we followed methods developed by (Zhang and Darnell, 2011)and aligned CLIP sequence reads from WT and 155KO libraries with < 3 substitutions, insertions or deletions. We used novoalign (http://www.novocraft.com) with parameters: mouse.utr -t 85 -l 25 -s 1. We identified putative crosslink induced deletion sites by tracking deletions between the first and last 2 nucleotides of the CLIP read alignments. To remove duplicates resulting from PCR amplification, CLIP read alignments were collapsed if they had the same barcode and the same genomic starting position.
We scaled the number of deletions in WT and 155KO libraries by 1) normalizing the number of WT and 155KO deletions with the total number of deletions present in their respective libraries and 2) multiplying the average number of deletions across both libraries to the normalized value.

Conservation analysis

Conservation of regions surrounding AGO bound canonical and noncanonical miR-155 sites were assessed by evaluating a 200bp window centered around canonical 6mer (nt 2-8) or non-canonical mismatched miR-155 seeds (nt 1-7 or 2-8). 23-way multiple species alignments were used to calculate the median number of species with a nucleotide identical to the mouse nucleotide at a given position relative to the seed. Alignments were derived from TargetScan (http://www.targetscan.org/).

AGO binding sites

Peak calling

Regions with a high density of mapped reads are indicative of AGO binding sites. Often, the strongest AGO binding sites can have coverage from hundreds of overlapping uniquely mapped CLIP reads. We developed a computational approach for identifying transcriptome-wide AGO binding sites using principles from edge detection in computer vision. For example, sharp changes in brightness often occur at the edges of an object in an image; we consider sharp changes in counts of reads aligned along a transcript as a way to non-parametrically identify boundaries of AGO binding sites.

For each sample library, we constructed a 1D signal profile of normalized read counts mapped to positions along the 3'UTR for every *i*th gene. We define normalized read count as the sum of reads mapped to a position, divided by the total

number of reads mapped in the sample library, per million (or reads per million). Specifically, for all positions *x* in gene *i*, $W_i[x]$ (or $K_i[x]$) reflects the cumulative normalized read count from all WT (or 155KO) libraries. We also construct a signal profile $D_i[x]$ by calculating the absolute difference between $W_i[x]$ and $K_i[x]$ in order to detect "valleys" of AGO binding. All signal profiles are convolved with a discrete Gaussian kernel g_D in order to smooth the counts and dampen low artifactual peaks. We set the discrete Gaussian filter mask $*m*$ to the average length of mapped reads (21nt), and the standard deviation σ was set to 15.

Next, we approximate the gradient of each profile by calculating a first order finite difference. The resulting gradient captures changes in the counts of reads aligned along a transcript. Using a fixed window of 36nt, we use gradient ascent (and descent) to identify positions with local maximal (and minimal) magnitudes. We pair consecutive maximal and minimal positions to identify the boundaries of AGO binding sites (peak). Because our method can also easily incorporate reproducibility threshold requirements before calling a peak, we require at least one read from 7/12 biological replicates be present in a binding site from WT or 155KO libraries. As a final step, we merge boundaries of AGO binding sites if the boundaries overlap in $W_i[x]$, $K_i[x]$ and $D_i[x]$ by using an interval tree.

Once we have identified regions of AGO binding, we quantify differential changes between WT and 155KO libraries with the following steps. 1) For each replicate *r*, we calculate $m_{p,r}$, the median normalized read count over all positions in a binding site *p* (or peak). 2) We then calculate the median again over all replicates, $m_p = median_{\forall r}(m_{p,r})$. Thus, for each peak, log ($m_{p^{WT}}$) - log ( ) reflects the change in AGO binding between WT and 155KO in log scale.

Peak significance

We developed a way of assessing the significance of changes in AGO binding between WT and 155KO libraries. We used a negative binomial noise model for read counts mapping to binding site events and subsequently assessed significance of differential AGO binding. The negative binomial model suggests that the frequency of read counts in an AGO binding site can be estimated by a continuous mixture of Poisson distributions, where the mixing distribution of the Poisson rate λ is sampled from a gamma distribution (Robinson et al., 2010).

We used a generalized linear model approach to assess the significance of differential binding. We constructed a design matrix to describe the systematic part of the read count variation by including phenotype {*WT or 155KO*} and linker type {*3' linker 1 or 3' linker 2*} as factors in our model for testing differential binding. The design matrix represents a simple additive model where the estimated mean read count at each CLIP peak is described as a linear combination of phenotype and linker type, and noise about the mean is described using a negative binomial dispersion model.

Second, we proceeded to fit both the full model to the data and a reduced model without phenotype. We used the "deviance", or the difference between the log-likelihood of the fitted model and the maximum possible log-likelihood for both models to conduct a likelihood-ratio test (LRT) (Robinson et al., 2010). If there is no true effect for the phenotype term, the difference in deviances between the nested models should be small compared with a chi-square distribution having one degree of freedom.

We note that we do not model differential transcript abundance between conditions in our statistical approach; since we expect targets of miR-155 to be induced in KO relative to WT, our approach will therefore tend to be conservative in calling peaks with reduced AGO binding.

We used methods available in the "edgeR" package (Robinson et al., 2010) and implemented our solution with the following additional steps: 1) we rounded normalized read counts (i.e. $m_{p,r}$) to the nearest integer; 2) we added a pseudo-count to all zero valued normalized read counts by using the geometric mean from all libraries; 3) we estimated tag wise dispersion.

Finally, we define miR-155 dependent binding sites as the set of binding sites with an adjusted *P*-value < .01 and log ($m_{p^{WT}}$) - log ($m_{p^{155KO}}$) > 0.

Sequence analysis of AGO binding sites

Supervised modeling of binding sites

We developed a method for inferring positions of non-canonical and canonical sites in miR-155 dependent binding sites. We define a positive training set of sequences spanning miR-155 dependent binding sites as T. We define a random negative training set, of equal size to T, from AGO binding sites devoid of any significance (an arbitrarily high p-value > .5) as F. We hypothesized that the sequences in the positive training set are candidates for real miR-155 targeting, and that sequences in the negative set are not targets of miR-155 as evidenced by the lack of any associated differential AGO binding.

We used a large-margin based approach to learn ungapped alignment parameters $\vec{w}$ (e.g. cost of mismatches and matches) that can discriminate between the sequences in T and F based on the alignment score with miR-155. The alignment score is calculated as a linear function:

$$score_{\vec{w}}(s = sequence, m = miR-155, a) = \vec{w} * \Phi(s, m, a)$$

where, given an ungapped alignment a, Φ is a function that describes the features of the alignment (e.g.\ mismatches or matches at every position). We perform our ungapped alignments using a variation of the longest common substring algorithm.

While we can easily solve $a = \text{argmax}_a(\vec{w} * \Phi(s, m, a))$ using dynamic programming, simultaneously learning the parameters $\vec{w}$ such that $score_{\vec{w}}(s_f m, a) < score_{\vec{w}}(s_t, m, a), \forall t \in T, \forall f \in F$ is intractable. We implemented a solution by iteratively optimizing a and $\vec{w}$. Our basic steps involve starting with some initial parameters for $\vec{w}$ and solving a. Next, using the alignment features from a, we update $\vec{w}$ using a large-margin based approach. We repeat these steps until we reach some predefined stopping criterion (for example, 100 iterations or updates to $\vec{w}$ between consecutive iterations is less than some value). A similar method was previously developed independently for discrimination of protein sequences between homologous protein sequences and decoy sequences (Yu et al., 2008). Moreover, a sequence alignment based method was one of the first miRNA target prediction methods, but it relied on hand tuning $\vec{w}$ (John et al., 2004).

Argonaute dCLIP protocol

Protocol changes

The following protocol is based on the previously published Argonaute HITS-CLIP protocol (Chi et al., 2009). Major changes to that protocol are:

1) Different antibody
These experiments were completed using a previously published rabbit polyclonal antibody generated against the N-terminal peptide of Argonaute 2 (MYSGAGPVLASPAPTTSPIPGYAFKC) (O'Carroll D, 2007).

2) Different RNase
RNase I was used in place of RNase A because it is suggested to have less nucleotide bias.

3) Different 3' linker sequences
Two separate linkers were designed so that the protocol could be performed on wild type and knockout samples without cross-contamination (the RT primers are linker specific). There is linker ligation bias that is linker sequence dependent. This was accounted for by 1) flipping which linker is used for which genotype (equivalent to a dye swap in a two color microarray experiment or a label swap in a SILAC experiment) and 2) modeling differential binding with a generalized linear model that incorporated linker type as a factor.

4) Cutting the membrane
miRNA and mRNA are cut together on the membrane (from ~100-150kD) rather than making two separate libraries. This increases the amount of material going into a CLIP library (which should act as a carrier) and reduces the number of samples to carry through subsequent steps.

5) RT primers
Contain a multiplexing sequence for combining libraries on the sequencer and a degenerate barcode for collapsing reads that are PCR amplifications of the same cloned fragment. Sequences of all primers and linkers can be found at the end of this protocol.

The structure of the libraries are:

> 5' Flow cell binding site-Read 1 Seq Primer (custom)-Insert-GUGUC- Linker 1 or Linker 2 primer site- Index Sequencing Primer-5 nt Multiplex Index- 6 nt Degenerate Barcode-3' Flow cell binding site

6) Replace two rounds of PCR and gel extraction with a single round of each.
This protocol uses one round of PCR and gel extraction.

7) The entire RT reaction contributes to the final library.

Amplification is performed on a real time thermocycler to prevent over-amplification. This allows the entire RT reaction to be used for the PCR that generates the final library. This improves library complexity and simplifies the protocol.

Protocol Specifics

This protocol is for differential CLIP with two different linkers to prevent cross-contamination following the ligation step. The protocol is written with volumes for completing 6 replicates in a single experiment and use of both linkers (with three samples using linker 1 and three samples using linker 2).

For washes: Completely resuspend beads in wash buffer (by repeatedly flicking tube after adding wash buffer). Then place on magnet to remove supernatant. Extended washes are not necessary; although samples should be left on ice in wash buffer while reaction mixes are prepared.

Argonaute is ~100kD. Western blots to check the IP should be done with a secondary that will not recognize the denatured antibody used for IP.

IP conditions should be optimized for each antibody.

Lysate preparation and reactions done on beads are incubated in a Thermomixer (Eppendorf), which can shake and maintain a temperature.

Cells and crosslinking

1. Purify CD4+ cells from dissected lymph nodes and spleens of mice using **Dynabeads Mouse CD4** (Invitrogen) according to the recommended protocol.
    a. Keep cells from each mouse separate to serve as biological replicates.
2. Make a solution of **1μg/mL anti-CD3 and CD28** antibodies in **PBS** and add **~400 μL/well** to a 24 well plate. Incubate at 37°C for 1-2 hours.
3. Remove antibody solution from plate just before adding cells (do not allow plate to dry).
4. Add **$8.5*10^5$ CD4+ cells/well** to coated plate at **1mL RPMI** with **10% Fetal Calf Serum** and **20U/mL human Il-2**.
5. On day 3, add **1 mL of media with Il-2** to each well.
6. On day 4, harvest cells from plates and count.
7. Resuspend **$65*10^6$ cells** from each mouse in **1mL PBS** and place in a well of a 6 well plate (non tissue culture treated).
8. Crosslink by placing plate (lid off) on a platter of ice in a Stratalinker UV crosslinker (Stratagene).
9. Crosslink with 400 mJ [ ].
10. Swirl plate.
11. Crosslink with 200 mJ [ ].
12. Swirl plate.
13. Crosslink with 200 mJ [ ].
14. Remove crosslinked cells from plate and place in a 1.5mL tube.
15. Wash plate in 0.5 mL and add to samples.
16. Centrifuge 800xg in a microcentrifuge for 7 minutes.
17. Completely remove supernatant and freeze cell pellet dry ice.
18. Place frozen cell pellets at -80°C.

CLIP Buffers
Filter sterilize. Tris, $MgCl_2$, and EDTA are from Ambion Buffer Kit (RNAse free).

1xPNK buffer
50 mM Tris-Cl pH=7.4, 10 mM MgCl2, 0.5% Ipegal CA-630

Volume (mL)	Component
450	**Nuclease-Free water** (Ambion)
19.5	**1M Tris pH=7.0** (Ambion)
5.5	**1M Tris pH=8.0** (Ambion)
5	**1M $MgCl_2$** (Ambion)
25	**10% Ipegal CA-630** (for molecular biology, Sigma)

1xPNK + EGTA buffer
50 mM Tris-Cl pH=7.4, 20mM EGTA, 0.5% Ipegal CA-630

Volume (mL)	Component
42.5	**Nuclease-Free water** (Ambion)
1.95	**1M Tris pH=7.0** (Ambion)
0.55	**1M Tris pH=8.0** (Ambion)
2.5	**400mM EGTA**
2.5	**10% Ipegal CA-630** (for molecular biology, Sigma)

1x PXL
1x PBS, 0.1% Sodium dodecyl sulfate (SDS), 0.5% Sodium deoxycholate, 0.5% Ipegal CA-630

Volume (mL)	Component
395	**Nuclease-Free water** (Ambion)
50	**10x PBS** (Ambion)
5	**10% Sodium dodecyl sulfate (SDS)**
25	**10% Sodium deoxycholate**
25	**10% Ipegal CA-630** (for molecular biology, Sigma)

5x PXL
5x PBS, 0.1% Sodium dodecyl sulfate (SDS), 0.5% Sodium Deoxycholate, 0.5% Ipegal CA-630

Volume (mL)	Component
195	**Nuclease-Free water** (Ambion)
250	**10x PBS** (Ambion)
5	**10% Sodium dodecyl sulfate (SDS)**
25	**10% Sodium deoxycholate**
25	**10% Ipegal CA-630** (for molecular biology, Sigma)

PK buffer (For RNA release step)
100mM Tris pH 7.5, 50mM NaCl, 10mM EDTA

Volume (mL)	Component
13	**Nuclease-Free water** (Ambion)
1.05	**1M Tris pH=7.0** (Ambion)
0.45	**1M Tris pH=8.0** (Ambion)
0.15	**5M NaCl** (Ambion)
0.3	**0.5M EDTA** (Ambion)

1. Prepare Protein G Dynabeads

1. Add 6*90 µL = **540 µL Dynabeads Protein G** (Invitrogen) to a 1.5 mL tube.
2. [ ] [ ] [ ] Wash beads 3x with **1mL PBS w/ 0.02% Tween 20**.
3. Resuspend washed beads in **540 µL PBS w/ 0.02% Tween 20**.
4. Add 6*12 = **72 µL Ago 2 serum** (Custom rabbit polyclonal generated against a peptide from the Ago2 n-terminus: MYSGAGPVLASPAPTTSPIPGYAFKC).
5. Rotate 4°C for ~2 h.
6. After ~1 hour of incubation, proceed to lysate preparation.
7. [ ] [ ] [ ] Wash beads 3x with **1mL 1xPXL**. In last wash, split evenly into 6 tubes.

2. Lysate Preparation

1. Dissolve **1 Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablet** (Roche) in **10 mL 1xPXL**.
2. Pre-chill microcentrifuge to 4°C for step 11.
3. Add 6*7.5*1.15= **52 µL RNasin Plus RNase Inhibitor** (40U/µL, Promega) to 6*370* 1.15= **2.6 mL PXL w/protease inhibitor**.
4. Resuspend each tube of crosslinked lysate in **375 µL 1X PXL+protease inhibitors+RNasin Plus**.
5. Incubate on ice for 10 minutes. Agitate every couple minutes.
6. Add **20 µL RQ1 RNase-Free DNAse** (1U/µL, Promega) to each tube.
7. Incubate in a Thermomixer at 37°C for 5 minutes, 1000 rpm.
8. Add additional **20 µL RQ1 RNase-Free DNAse** to each tube.
9. Dilute **2 µL RNAse1** (100U/µL, Ambion) in **500 µL 1x PXL**.
10. Add **9.75 µL diluted RNAse1** to each sample.
    1. Incubate at 37°C for 5 minutes, 1000 rpm.
    2. Run lysates 5x through each needle size (below) to further break up DNA.
        1. 18 gauge (can be skipped if it goes through the 23 gauge easily)
        2. 23 gauge
        3. 25 gauge
        4. 27 gauge
11. Spin lysates in pre-chilled microcentrifuge, maximum speed, for 20 minutes at 4°C.
12. While spinning, prepare tubes for IP and aliquot beads to each tube.

3. Immunoprecipitation

1. **15 µL of the supernatant** from centrifuged samples can be removed to new tubes as "input." This can be used for Western Blot to check the IP efficiency.
2. Place beads on magnet, remove liquid, and immediately replace with lysate supernatant from step 12 (above).
3. Rotate beads/lysate mix for 1.5h at 4°C.

4. 5' end labeling of 3' linkers (Complete during the IP)

Note: To reduce upfront cost, I ordered only phosphorylated 3' linkers, which require dephosphorylation before adding a radioactive phosphate (protocol below). To simplify, 3' linkers can be ordered both 5' phosphorylated and non-phosphorylated. If you use this approach skip to step 9 and use the non-phophosphorylated linker for that step.

1. Make Dephosphorylation Master Mix (Use components from KinaseMax Kit (Ambion)):

Volume (µL)	Component
27.2	**RT-PCR water** (Ambion)
4	**10x Dephosphorylation Buffer** (Kinase Max, Ambion)
4	**Calf Alkaline Phosphatase** (1U/µL, Kinase Max, Ambion)

2. Add **17.6 µL of Dephosphorylation Master Mix** to each of two tubes (one for each linker).
3. Add **2.4 µL of 3' linker 1** (20pmol/µL) to one aliquoted Dephosphorylation Master Mix and **2.4 µL of 3' linker 2** to the other aliquot.
4. Incubate 30 minutes at 37°C.
5. Add **20 µL of phosphatase removal reagent** (KinaseMax, Ambion) to each tube.
6. Flick tube occasionally for 3 minutes.
7. Centrifuge the tube for 15s, ramping to max speed.
8. Carefully transfer supernatant to new tube.
9. Set-up kinase reaction for both linkers:

Volume (µL)	Component
5	**RT-PCR water** (Ambion)
5	**10x Kinase Buffer** (KinaseMax, Ambion)
24	**Dephosphorylated linker RNA** (from step 8, above)
12	**γ-ATP** (3,000Ci/mmol, 10mCi/mL, BLU502A, Perkin Elmer)
4	**T4 PNK** (KinaseMax, Ambion)

10. Incubate 37°C for 30 minutes.
11. Add **3 µL of 100µM ATP** (10 mM ATP diluted 1:100) to kinase reactions.
12. Incubate at 37°C for additional 5 minutes.
13. Resuspend resin in 2 **illustra MicroSpin G-25 columns** (GE) by vortexing.
14. Pre-spin columns for 1 min at 750 x g.
15. Place columns in new tubes.

16. Without touching the resin with the pipette tip, apply one kinase reaction to the center of a G-25 column and the other kinase reaction to the other G-25 column.
17. Spin columns for 2 min at 735 x g (orient slope of resin in centrifuge so that the same slope is maintained).
18. Approximate percentage of label in the flow through versus in the column using a Geiger counter (should be ~25-50% incorporation).
19. Leave eluted radioactive linkers on ice behind a shield.

5. Washes and CIP Treatment

1. After IP, place tubes on magnet, remove the supernatant, and save as "depleted." This can be used to check the IP efficiency by Western Blot.
2. [ ] [ ] Wash beads 2x with 1X PXL.
3. [ ] [ ] Wash beads 2x with 5X PXL.
4. [ ] [ ] Wash beads 2x with 1X PNK Buffer, transfer in last wash to new tube.
5. Make CIP Master Mix:

Volume (µL)	Component
350	**RT-PCR water** (Ambion)
42	**10x Dephosphorylation Buffer** (Roche)
10.5	**RNasin Plus** (Promega)
16	**Alkaline Phosphatase** (1U/µL, Roche, 10713023001)

6. Add **60 µL CIP Master Mix** to each tube.
7. Incubate at 37°C for 20 minutes (1200 rpm for 15 seconds every 2 minutes).
8. [ ] Wash 1X with 1X PNK Buffer.
9. [ ] [ ] Wash 2X with 1X PNK+EGTA Buffer.
10. [ ] [ ] Wash 2X with 1X PNK Buffer. Transfer to new tube in last wash.

6. 3' RNA Linker Ligation

1. Make two tubes of Ligation Master Mix (one for linker 1 and one for linker 2)

Volume (µL)	Component
126	**RT-PCR water** (Ambion)
25.6	**10mM ATP** (NEB)
25.6	**10x T4 RNA ligase buffer** (Fermentas)
16	**BSA** (1 mg/mL, Fermentas, comes with ligase)
6.4	**RNAsin Plus** (Promega)
6.4	**T4 RNA ligase** (10U/µL, Fermentas)
50	**Radioactively labeled linker 1 or linker 2** (from section 4)

2. Remove last wash from beads and add **80 µL of the appropriate Ligation Master Mix** to each tube of beads.
3. Incubate at 16°C for 2 hours (1200 rpm for 15 seconds, every 2 minutes).
4. Add **3µL of appropriate linker** (20pmol/µL; with Phosphate on 5'end) to each sample, and incubate overnight (16°C, 1200 rpm for 15 seconds, every 2 minutes).

7. Day 2, PNK Treatment

1. [ ] Wash 1x with 1X PXL buffer (**Note: all washes of beads from this point forward are radioactive and should be disposed of as radioactive waste. Most of the radioactivity will be in the initial washes)**.
2. [ ] Wash 1x with 5X PXL buffer.
3. [ ] [ ] [ ] Wash 3x with 1X PNK buffer.
4. Make PNK Master Mix.

Volume (µL)	Component
406	**RT-PCR water** (Ambion)
56	**10x PNK Buffer** (NEB)
56	**ATP** (10mM)
14	**RNAsin Plus** (Promega)
28	**T4 PNK** (10,000U/mL, NEB)

5. Add **80 µL PNK Master Mix** to each tube of beads.
6. Incubate at 37°C for 20 minutes (1200 rpm for 15 seconds, every 2 minutes).
7. [ ] [ ] [ ] Wash 3x with 1X PNK buffer. Transfer in last wash.

8. SDS-PAGE and transfer to nitrocellulose

1. Make Loading Dye Master Mix.

Volume (µL)	Component
158	**PNK buffer** (used for washes)
158	**NuPAGE LDS Sample Buffer 4x** (Invitrogen, used at 2x)
35	**NuPAGE Sample Reducing Agent 10x** (Invitrogen)

2. Remove last wash from beads and add **55 µL of the appropriate Ligation Master Mix** to each tube of beads.
3. Incubate at 70°C, 1000RPM for 10 minutes.
4. During this time set-up two NuPAGE 4-12% Bis-Tris Gel 1.0 mm, 9 well (Invitrogen) with NuPAGE MOPS SDS Running Buffer. Set-up gel in an ice bucket surrounded by ice.
5. Load each sample into 2 wells of the gel. Run samples with different linkers next to each other (to prevent inter-library contamination).
6. Load **15 µL Full Range Molecular Weight Marker** (GE) into the wells between different samples.
7. Run gel ~2 hours at 150V.
8. For transfer, use NuPAGE transfer buffer and 13% methanol. Transfer to **Protran BA85 0.45 µm nitrocellulose (Whatman)**. Surround the transfer cell with ice in an ice bucket. Use a stirbar.
9. In our old BioRad transfer cell (no longer sold), transfer 130V for 2.5 hours (different voltage may be required for other transfer cells).
10. Rinse membrane in Nuclease Free PBS (Diluted from Ambion 10x).
11. Place nitrocellulose in section of page protector. Tape down in a film cassette. Place several luminescent stickers for orientation.
12. Expose film 1-3 days at -80C.

9. RNA Release

1. Overlay film on nitrocellulose (in page protector). Use film to identify position of lanes on nitrocellulose. To mark landmarks to guide cutting, a needle can be used to pierce through film and nitrocellulose.
2. Using a new scalpel blade for each sample, cut nitrocellulose corresponding to sample from ~100kD-150kD (will contain both miRNA and digested mRNA fragements, see image at the end of the protocol).
3. Cut nitrocellulose piece into many small fragments (~3x3mm) and use the scalpel blade to transfer into a new 1.5 mL tube.
4. Mix **800 µL Proteinase K** (20mg/mL solution, recombinant, PCR-grade, Roche) with **3.2 mL PK buffer** (final solution is 4mg/mL Proteinase K).
5. Pre-incubate this stock at 37°C for 12 minutes to digest RNAses.
6. Add **200 µL of diluted proteinase K** to each tube of nitrocellulose.
    a. Incubate 20 minutes at 37°C, 1000 rpm.
7. Make PK/7M urea solution by adding some of the **diluted proteinase K pre-incubated in step 5** to **0.84 g urea** to a final volume of **2 mL**. Heat at 37°C to bring into solution.
    a. Add **200 µL PK/7M urea** solution to each sample.
    b. Incubate 20 mins at 37°C, 1000rpm.
8. To each sample, add **400 µL Phenol** (BioReagent, Saturated with 0.1M citrate buffer, pH 4.3, Sigma)—avoid using aqueous phase that sits on top of phenol.
9. To each sample, add **130 µL chloroform:isoamyl acohol 49:1** (BioUltra, for molecular biology, Fluka, Sigma).
10. Vortex well, then incubate 37°C for 20 mins, 1000rpm. Vortex several times during this incubation.
11. Spin tubes at maximum speed in microcentrifuge for 15 minutes at 4°C.
12. Remove aqueous phase to a new tube.
13. Add **40 µL or ~10% of volume 3M Sodium Acetate pH 5.2** (buffer kit, Ambion) to the aqueous layer.
14. Add **0.7 µL glycogen (5mg/mL, Ambion)**.
15. Add **1 mL of 1:1 mix of Ethanol:Isopropanol**.
16. Vortex samples.
17. Precipitate by incubating overnight at -20°C.

10. 5' RNA Linker Ligation and DNAse treatment

1. Spin samples (30 minutes, 4°C , maximum speed in a microcentrifuge).
    1. Dispose of supernatant.
2. Add **150 µL ice cold 75% ethanol**.
    1. Spin samples (10 minutes, 4°C , maximum speed in a microcentrifuge).
    2. Dispose of supernatant.
3. Add **150 µL ice cold 75% ethanol**.
    1. Gently vortex.

2. Spin samples (10 minutes, 4°C , maximum speed in a microcentrifuge).
3. Carefully dispose of as much supernatant as possible.

4. Let dry about 10 minutes until there is no visible liquid in tube and pellet appears as gel-like droplet.
5. While samples are drying, make RNA Ligation Master Mix

Volume (µL)	Component
44.25	**RT-PCR water** (Ambion)
7.5	**10X T4 RNA** ligase buffer (Fermentas)
6.75	**BSA** (10 µg/µL, Fermentas, comes with ligase)
7.5	**ATP** (10 mM)
7.5	**5' linker** @ 5 pmol/µL
1.5	**T4 RNA ligase** (10U/µL, Fermentas)

6. Resuspend each sample in **10 µL RNA Ligation Master Mix**. Using the pipette tip, run the liquid droplet over side of tube that RNA was spun down against at least 10 times to recover the maximum amount of RNA.
7. Incubate 16°C for 2 hours.
8. Add **0.2 µL T4 RNA Ligase** to each sample
9. Incubate an additional 3 hours at 16°C.
10. Make DNAse Master Mix.

Volume (µL)	Component
553	**RT-PCR water** (Ambion)
77	**10X RQ1 DNAse Buffer** (Promega)
35	**RNasin Plus** (Promega)
35	**RQ1 RNase-Free DNAse** (1U/µL, Promega)

10. Add **100 µL DNAse Master Mix** to each sample.
11. Incubate 37°C for 20 minutes.
12. To each tube add:
    1. 300 µl **RT-PCR water** (Ambion)
    2. 300 µl **Phenol** (BioReagent, with 0.1M citrate buffer, pH 4.3, Sigma)
    3. 100 µl **Chloroform:Isoamyl Alcohol 49:1** (BioUltra, for molecular biology, Fluka, Sigma)
13. Shake tubes vigorously for 15 seconds, let rest for 1 minute, and repeat 2x [ ] [ ] [ ]
14. Spin in microcentrifuge at 12,000xg for 15 minutes.
15. Remove aqueous layer to new tube.
16. To aqueous layer add:
    1. **45 µl 3M Sodium Acetate pH 5.2** (buffer kit, Ambion)
    2. **0.6 µl glycogen** (5mg/mL, Ambion)

3. **1 mL 1:1 mix of ethanol:isopropanol**.

17. Vortex samples.
18. Precipitate by incubating overnight at -20°C.

11. RT-PCR

1. Spin samples (30 minutes, 4°C, maximum speed in a microcentrifuge).
    1. Dispose of supernatant
2. Add **150 µL 75% ice cold ethanol**.
    1. Spin samples (10 minutes, 4°C , maximum speed in a microcentrifuge).
    2. Dispose of supernatant.
3. Add **150 µL 75% ice cold ethanol**.
    1. Gently vortex
    2. Spin samples (10 minutes, 4°C , maximum speed in a microcentrifuge).
    3. Carefully dispose of as much supernatant as possible
4. Let dry about 10 minutes until there is no visible liquid in tube and pellet appears as gel-like droplet.
5. Make dNTP dilution

Volume (µL)	Component
200	**RT-PCR water** (Ambion)
17.2	**10mM dNTPs** (mixed from 100mM dNTP set, Invitrogen)

6. Make RT Primer Mix for each RT primer that will be used

Volume (µL)	Component
31	**dNTP dilution (see step 5)**
2	**5 µM RT primer (make sure RT primer is compatible with 3'linker used)**

7. Resuspend precipitated samples in **6.75 µL RT Primer Mix**
8. Place at 65°C for 5 minutes.
9. Place on ice.
10. Make RT Master Mix

Volume (µL)	Component
4	**0.1M DTT**
16	**5x Superscript Buffer** (Invitrogen)
2	**RNasin Plus** (Promega)
4	**Superscript III** (200U/µL, Invitrogen)

11. Add **3.25 µL of the RT Master Mix** to each sample.
12. Transfer samples to PCR strips.
13. Place in prewarmed thermocycler with following program:
    1. 52.5°C 45:00
    2. 55°C 15:00
    3. 90°C 5:00

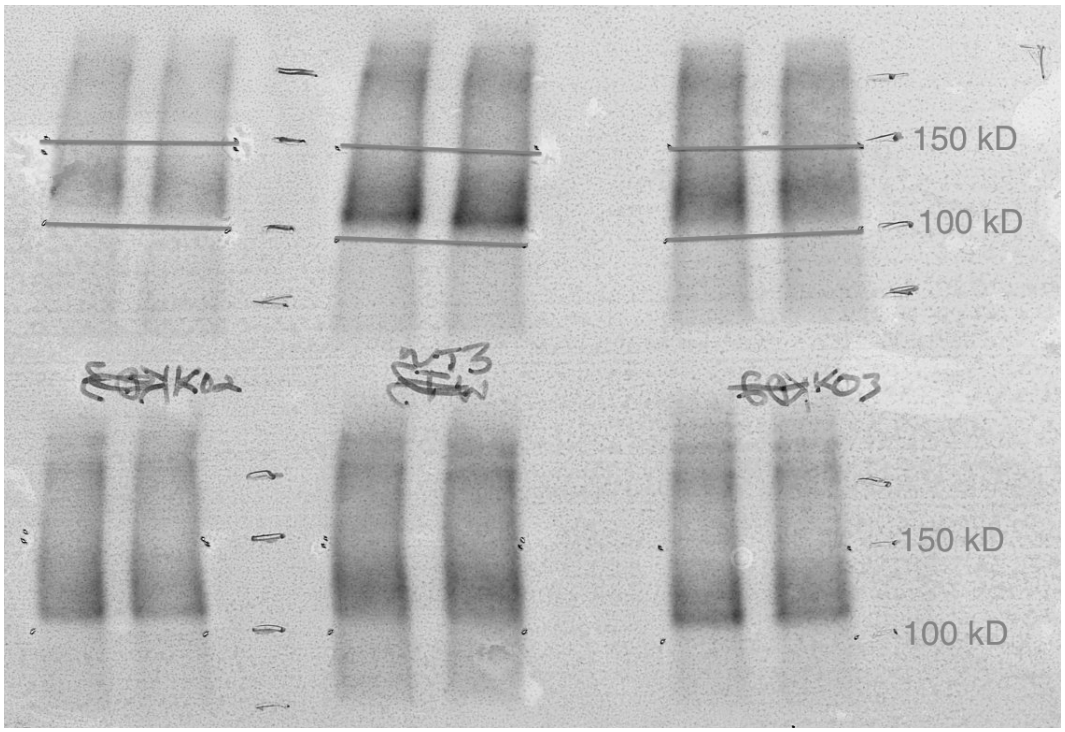
4. 4°C forever

14. Place RT reactions on ice.
15. Make the PCR Master Mix

Volume (µL)	Component
575	**RT-PCR water** (Ambion)
196.5	**5X Phusion HF Buffer** (NEB)
19.8	**10mM dNTPs**
45	**5' PCR Primer** 10µM
45	**3' PCR Primer** 10µM
9.75	**50x SybrGreen** (Diluted in RT-PCR water from SYBR Green I Nucleic Acid Gel Stain—10,000x, Invitrogen), final in PCR is 0.5x
9.75	**Phusion High-Fidelity DNA Polymerase** (2,000U/mL, NEB)

16. For each sample, add 37 µL of PCR Master Mix to three wells of a PCR strip (compatible with the real-time thermocycler). Put the aliquots for each sample into their own strip so that they can be easily removed as a unit from the thermocycler when optimal amplification has occurred.
17. Add **3.3 µL of RT Reaction** to each aliquot of **PCR Master Mix**.
18. Place samples into a pre-warmed real-time thermal cycler that can be easily opened and closed during thermocycling (eg. BioRad CFX 96 detection module) with the following program:
    1. 98°C 1:30
    2. 98°C 0:12
    3. 59.6°C 0:30
    4. 72°C 0:10
    5. Plate Read (Sybr Green)
    6. 72°C 0:10
    7. Go to Step 2, 35x
19. Aliquots corresponding to a sample should be removed when amplification for that sample leaves the exponential phase (at the first sign of a plateau, typically between 19-23 cycles). Samples should be removed from the thermal cycler after the plate read, but before the denaturation step. If temperature starts ramping to 98°C, wait for next cycle to remove samples.
20. Pool amplified aliquots from a single sample.
21. Stopping point—samples can be frozen or protocol can be continued.
    1. **Never** take amplified material back to the bench where earlier steps of the CLIP protocol are performed.
22. Sample volume can be reduced by ~20-40% in a vacuum centrifuge to facilitate gel loading for gel extraction.
23. Make 4.5% **MetaPhor Agarose Gel (Lonza)** with **0.5-1 µg/mL ethidium bromide**.
24. Run ½ of amplified sample slowly on gel (eg. 80V) and reserve ½ as back-up
25. Cut from 170 basepairs-215 basepairs with a new scalpel (this may require re-staining the agarose gel with ethidium bromide), see image at end of protocol.

    1. Note: there will likely be a band at ~150 nt that contains no insert, this material should be avoided. Run the gel until this band can be distinguished from the smear of amplified material.
26. Extract the DNA from Agarose using QIAquick Gel Extraction Kit (Qiagen) (this may require multiple columns per slice).
27. Combine equal amounts of all libraries to be sequenced together.
28. Sequence extracted libraries using Read 1 Sequencing primer (Custom) and Index Sequencing Primer (Illumina) with an 11 nucleotide index read (Note: These primers have only been tested on an Illumina GAIIX)



Film from CLIP experiment. 72 hour exposure of two pieces of nitrocellulose containing 6 replicates. The top has been marked to show the region of nitrocellulose that would be cut. The bottom has been left unmarked.

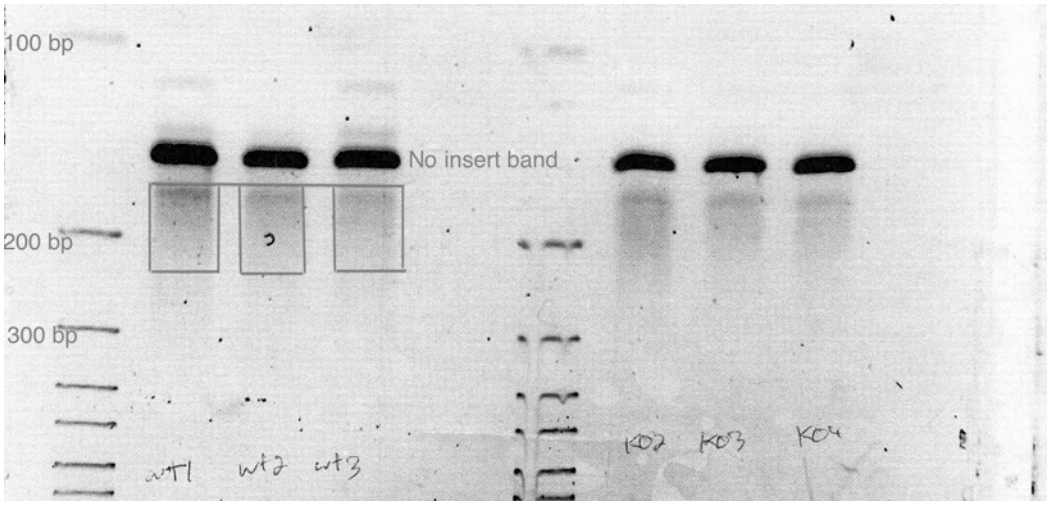


Image of gel prior to gel extraction. The left side has been marked to show how the gel would be cut for gel extraction. The right side has been left unmarked. The gel can be run for more time to improve separation of the "no insert band" from the desired material.

Name	Type	Sequence	Modifications and Purification
3' Linker 1	RNA	/5Phos/rGrUrG rUrCrU rUrUrA rCrArC rArGrC rUrArC rGrGrC rGrUrC rG/3ddC/	5'Phos,3'Dideoxy-C,RNase-Free HPLC
3' Linker 2	RNA	/5Phos/rGrUrG rUrCrA rCrUrG rArUrA rGrCrA rArCrC rCrGrG rUrGrC rU/3ddC/	5'Phos,3'Dideoxy-C,RNase-Free HPLC
5' Linker	RNA	rCrGrA rCrCrA rGrCrA rUrCrG rArCrU rCrArG rArArG	RNase-Free HPLC
RT primer_1.1	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNCGCTA**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_1.2	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNGACAC**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_1.3	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNTGTGC**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_1.4	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNATATC**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_1.5	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNAGGAT**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_1.6	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNCAAGT**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_1.7	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNGCTTT**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_1.8	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNGGCGT**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_1.9	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNAGTTG**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_1.10	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNCACCG**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCCGACGCCGTAGCTGTGTAAA	PAGE
RT primer_2.1	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNTCAAA**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
RT primer_2.2	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNAATCA**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
RT primer_2.3	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNGTGGA**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
RT primer_2.4	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNCTTAG**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
RT primer_2.5	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNGGACG**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
RT primer_2.6	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNACCGG**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
RT primer_2.7	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNTAGTG**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
RT primer_2.8	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNCTGTT**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
RT primer_2.9	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNACACT**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
RT primer_2.10	DNA	CAAGCAGAAGACGGCATACGAGAT**NNNNNNTTAGG**GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCAGCACCGGGTTGCTATCAGT	PAGE
5' PCR primer	DNA	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT GGT ACT CCG ACC AGC ATC GAC TCA GAA G	PAGE
3' PCR primer	DNA	CAA GCA GAA GAC GGC ATA CGA GAT	PAGE
Read 1 Seq Primer	DNA	ACA CTG GTA CTC CGA CCA GCA TCG ACT CAG AAG	Standard Desalting
Index Seq Primer	DNA	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC	Same as Illumina, no need to order

RT primer_1.x are for linker 1 and RT primer_2.x are for linker 2.
RT primers X.1-7 have indexes that differ from each other at ≥ 4 nt. RT primers X.8-10 differ from each other at ≥ 4 nt. Between these groups, indexes differ by ≥ 3 nt.
Note: These libraries have only been tested on an Illumina GAIIX.

Primers used for Luciferase Constructs

Target	Primers	Fragment size (nt)
Cep135_Mutation	CAATGTGTTTATACTGTGAATTGCTCCTTCATGCTCACATTATG CTTTAATAATC	NA
	GATTATTAAAGCATAATGTGAGCATGAAGGAGCAATTCACAGT ATAAACACATTG	
Gas2l3_Mutation	GTATCCAACTTAACTTATATTTTCTCAAGTTGGTAAATGTCAGC ATAAACCCACTCTGTTTGC	NA
	GCAAACAGAGTGGGTTTATGCTGACATTTACCAACTTGAGAAA ATATAAGTTAAGTTGGATAC	
Unc119b_Mutation1	CCCCCTAGTTCACAGAGGTCGGCTAGTGAAGTAAAAGATTAG	NA
	CTAATCTTTTACTTCACTAGCCGACCTCTGTGAACTAGGGGG	
Unc119b_Mutation2	CAGTGTTTGTTCCCTGCCCGGGCGTTCACAGAGGTATTATAG	NA
	CTATAATACCTCTGTGAACGCCCGGGCAGGGAACAAACACTG	
Gimap3_Mutation	CCTATACCCCTCAATGCGTCGTCCAAAACTCGGCGTTTCAGCC	NA
	GGCTGAAACGCCGAGTTTTGGACGACGCATTGAGGGGTATAGG	
Zc3h11a_Mutation	GTAGGCCTCTTTTGTATATAACCTCTATTATTTTCGTCCAAAAA AAAAAAAAGGTCTGTTC	NA
	GAACAGACCTTTTTTTTTTTTTGGACGAAAATAATAGAGGTTAT ATACAAAAGAGGCCTAC	
Trib1_Mutation	CTCATAATTCTCACACATGGGTAATCTAATGACAGGCAGGCCC	NA
	GGGCCTGCCTGTCATTAGATTACCCATGTGTGAGAATTATGAG	
Unc119b_UTR	ctcgagTGCCTATAATGGAGGCCAGT	488
	gcggccgcCCTAACTGCATAGGCGAGGA	
Tgm2_UTR	ctcgagGCCAGAGAACTGGGAGTCAG	667
	gcggccgcCTGAAGCTTGCACAGACCAA	
Cep135_UTR	ctcgagTGGTTTTCAGTGAGTTTCTGC	581
	gcggccgcTGAGGGAGTCCTAACTCTCTGG	
Igf2r_UTR	ctcgagCTCGATGATGCGTCTGTCAT	936
	gcggccgcAACACTGGCCCTACACCAAC	
Zc3h11a_UTR	ctcgagATGCTCTGGGGTGTATTGGA	771
	gcggccgcTGCCCTCCTTTAAAACAATAGC	
Trib1_UTR	ctcgagTGAGTACCAGGAGGACAGTGA	502
	gcggccgcGGTTGAGTCTGCAGGGAGAG	
Gas2l3_UTR	ctcgagCTTGCCTGGCATTCTCTAGG	550
	gcggccgcGCCACGAGTCTGGTCTTTCT	
Rasa2_UTR	ctcgagCCGTGCTTTCCTGAAAATGT	713
	gcggccgcCCCTGGGTTTTGTCCTTTTT	
Gimap3_UTR	ctcgagTTGTGGATTGGGATCACTCA	644
	gcggccgcTGGGAGAGAAAAGAGAAGCAA	
Wasl_UTR	ctcgagTTCATGTGGCTTGAGCTCTG	751
	gcggccgcCATGATGAATTGCCACCAGT	
Tnrc6a_UTR	ctcgagGTTGCCGTCTGCTTTCAGAT	838
	gcggccgcAGTCAGCTTCGTGTCCCAAT	
Prkar1a_UTR	ctcgagCGGAACATCCAGCAGTACAA	654
	gcggccgcCCAAAACATTTCAGGGTTGG	

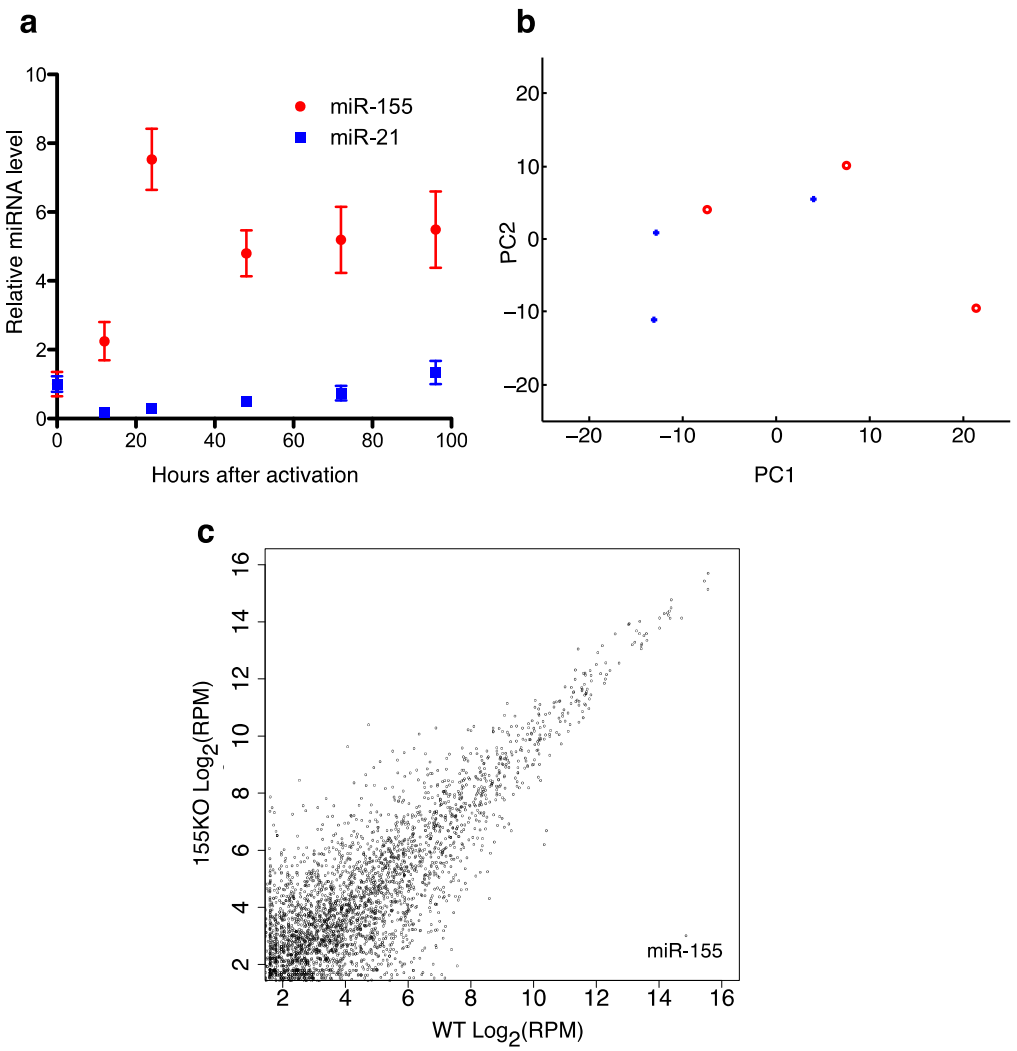


Figure S1. Characterization of wild type and miR-155 deficient activated T cells, Related to Figure 1

a) miR-155 levels during T cell activation. Mouse $CD4^+$ T cells were isolated and activated with 1 ug/mL CD3 and CD28 antibodies in the presence of IL-2 for the indicated number of hours. miR-155, miR-21, and U6 RNA were quantified by qRT-PCR. miRNA levels at each time point were normalized to U6 and miRNA levels are shown as a fold change relative to those in cells at 0 hours. Each replicate came from cells isolated from an independent animal and error bars represent SEM (n=4); data are representative of three independent experiments.

b) miR-155-deficient (155KO) and –sufficient T cells (WT) have similar transcriptomes. Principal component analysis of microarray expression data from WT and 155KO T cells activated for 3 days. Each point represents gene expression from a single array performed with cells from a separate animal. The two components represent ~80% of the total variance.

c) Levels of miRNAs bound to Argonaute are unaffected in miR-155 deficient cells. Reads mapping to miRNAs were analyzed for their seed (nt 2-7) and each seed is plotted separately to capture 5' end heterogeneity that may affect targeting (Wu et al., 2007). Data summarizes six representative WT and six representative 155KO libraries. All reads mapping to miRNAs were used to calculate the denominator for reads per million (RPM). RPM for each miRNA seed was calculated separately for each library and then averaged across libraries.

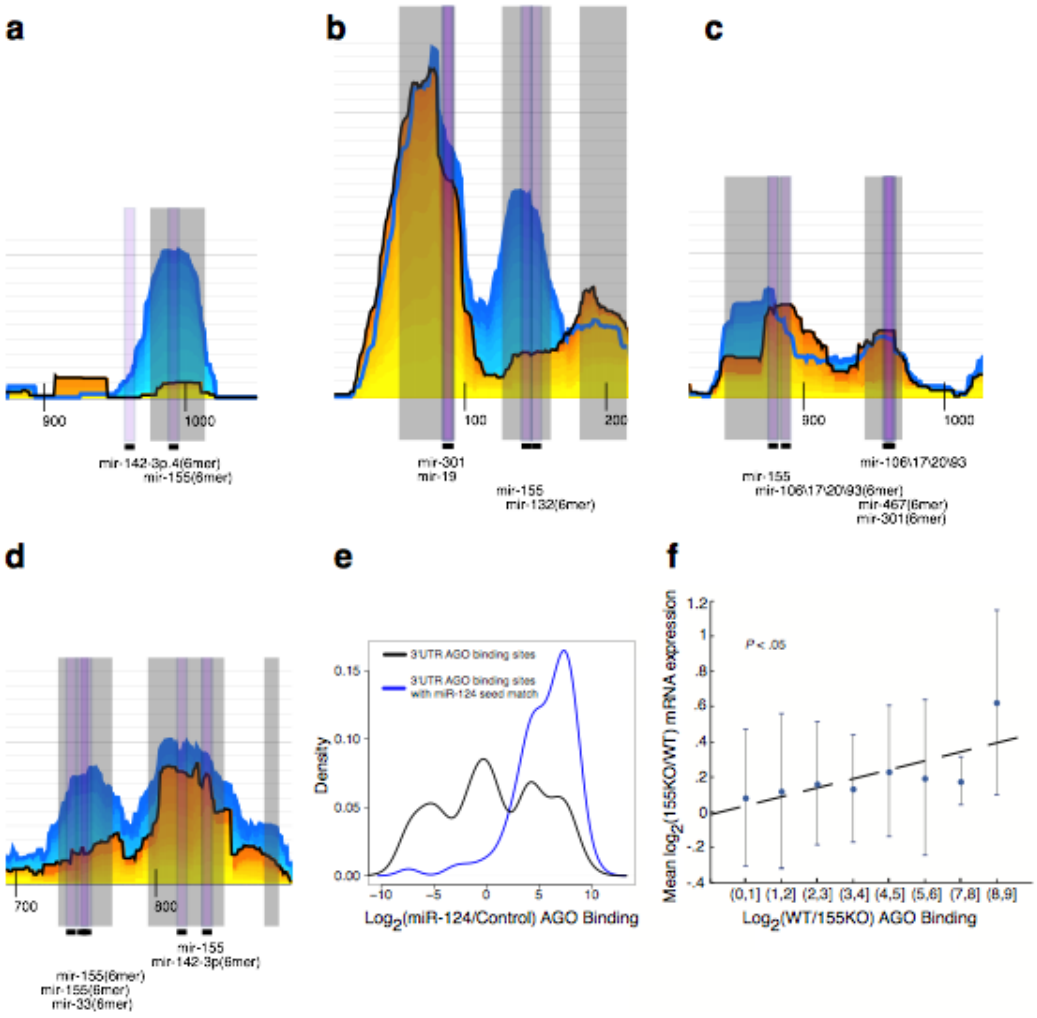


Figure S2. Efficacy of novel peak calling algorithm and effect of differential binding on differential gene expression, Related to Figure 2

(a-d) AGO binding at miR-155 binding sites in the a) *MAF*, b) *JARID2*, c) *SKI*, and d) *PICALM* genes revealed by dCLIP. While identifying clusters of overlapping reads suffices for detecting differential binding in an isolated peak such as that seen in a), many sites such as those seen in b), c), and d), have overlapping binding sites that cannot be discriminated by calling clusters of overlapping reads. Grey rectangles indicate AGO binding sites called by the edge detection method, which readily identifies overlapping binding sites as distinct. Reads from the 12 replicates of each genotype are stacked; blue shades are WT replicates, and yellow shades are 155KO replicates. Black bars indicate the location of seed matches of highly expressed miRNAs. Coordinates along the horizontal axis indicate nucleotides from the 3'UTR start.

e) The edge detection algorithm identifies binding sites in other CLIP data sets. Greater AGO binding is seen at AGO binding sites containing miR-124 seeds in miR-124 transfected HeLa cells than in control transfected cells. Peak calling was performed on published HITS-CLIP data using miR-124 and control transfected HeLa cells (Chi et al., 2009). Kernel density estimation is plotted for sites with coverage from at least 3 of 4 replicates in control transfected or miR-124 transfected cells. The average size for these sites is 30.2 nucleotides (relative to 46nt and 62nt used in (Chi et al., 2009)). The number of binding sites in the 3'UTR with miR-124 seed matches (nt 2-7) is 250 for sites with at least 2

replicates, 80 for binding sites with at least 3 replicates, and 30 for binding sites with all 4 replicates.
f) Association of differential AGO binding with level of miR-155 dependent regulation for canonical miR-155 sites. Linear regression fitting mRNA expression changes between mir-155-sufficient and –deficient cells to binned differential AGO binding ($R^2 = 0.5284$, $F = 6.7237$, $p = 0.0410$; R^2 ,F, and p were calculated using the Matlab regress function and assess whether differential AGO binding significantly contributes to the model of differential mRNA expression).

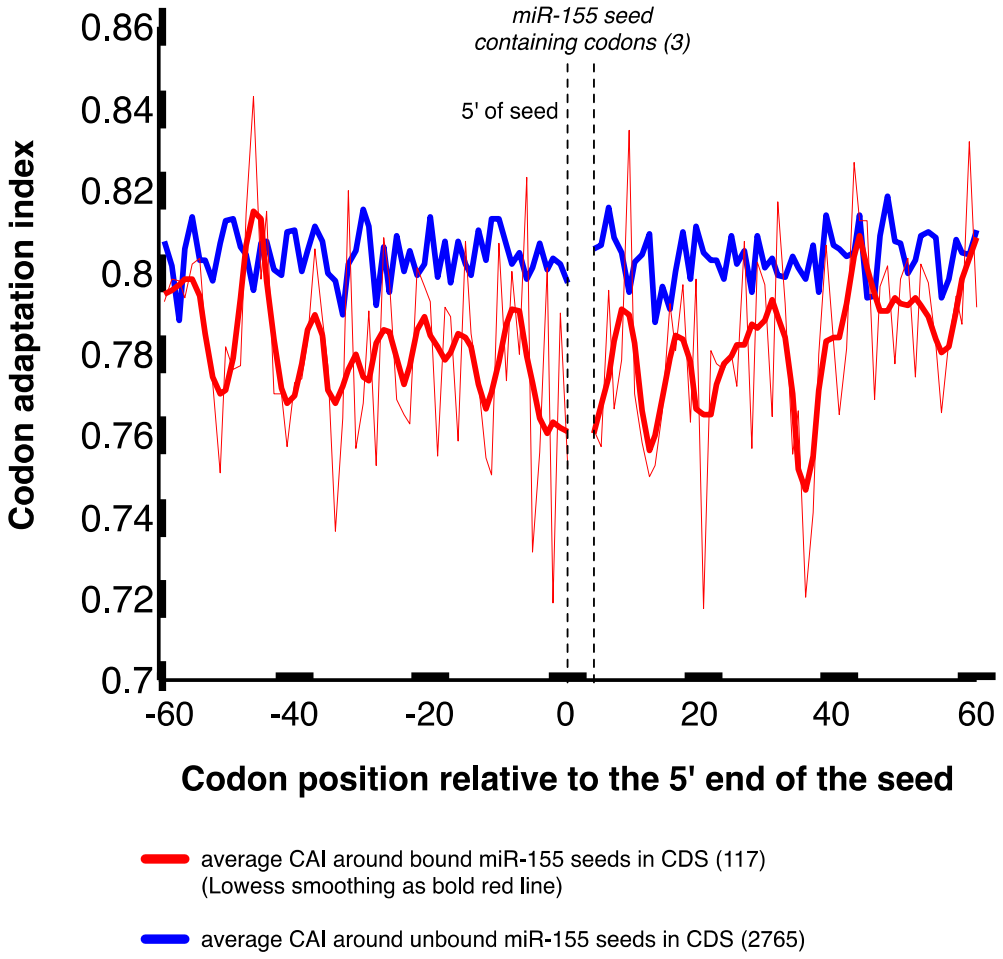


Figure S3. Rare codons are preferentially found surrounding Argonaute bound miR-155 seeds, Related to Figure 2

The average codon adaptation index (Sharp and Li, 1987) at locations proximal to AGO bound and unbound miR-155 seed matches within coding sequences.

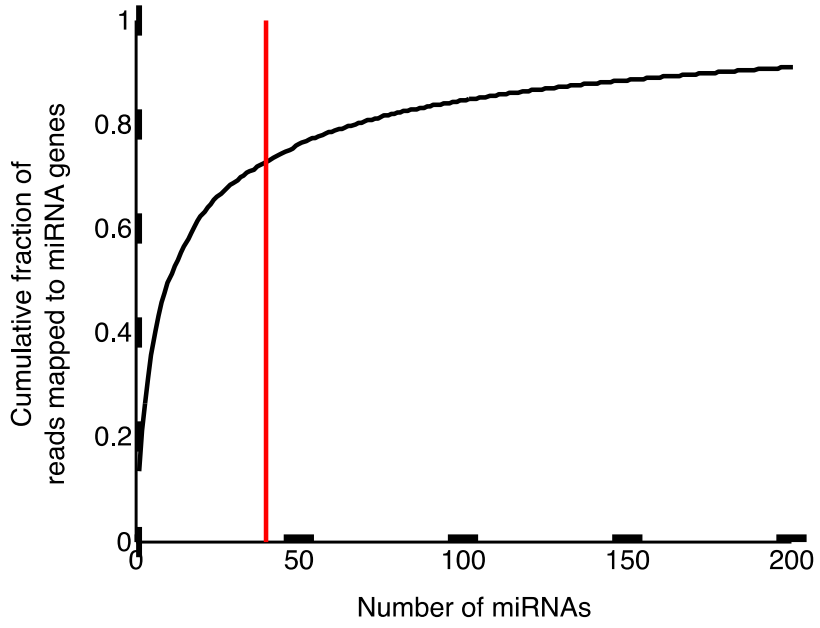


Figure S4. Distribution of miRNAs bound to Argonaute, Related to Figure 3
The fraction of reads from dCLIP libraries aligning to specific miRNAs out of all reads aligning to miRNAs. For assigning binding sites to miRNAs, we considered seeds of the top 40 miRNAs (denoted by the red line), which comprise more than 75% of all reads to miRNAs. Beyond the first ~20 miRNAs, each additional miRNA accounts for a minimal fraction of total miRNA bound to Argonaute.

6mer 3' Motifs with a single mismatch

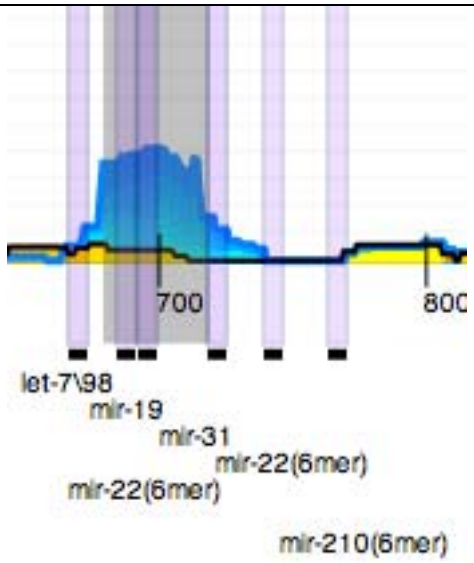
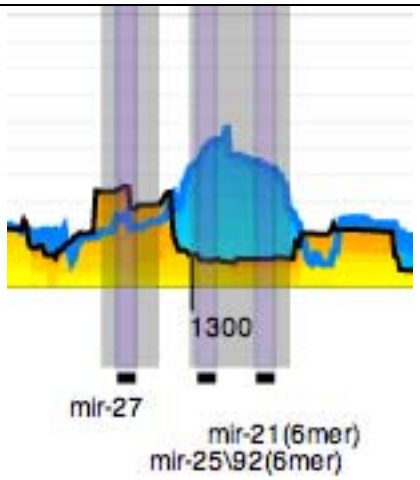
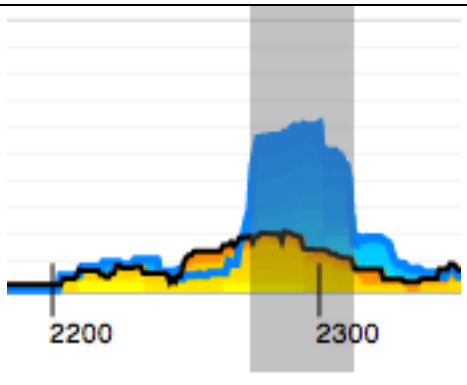
		Noncanonical			Canonical		
Position	Motif	# (of 114)	Percent	Enrichment (p-val)	# (of 130)	Percent	Enrichment (p-val)
1	CATTAA	16	14.0%	**$7.70x10^{-3}$**	10	7.7%	$4.56x10^{-1}$
2	GCATTA	11	9.6%	$3.83x10^{-2}$	7	5.4%	$5.27x10^{-1}$
3	AGCATT	8	7.0%	$4.85x10^{-1}$	9	6.9%	$4.94x10^{-1}$
4	TAGCAT	10	8.8%	$7.24x10^{-2}$	8	6.2%	$3.59x10^{-1}$
5	TTAGCA	9	7.9%	$2.01x10^{-1}$	12	9.2%	$6.80x10^{-2}$
6	ATTAGC	4	3.5%	$6.71x10^{-1}$	7	5.4%	$2.65x10^{-1}$
7	AATTAG	7	6.1%	$5.06x10^{-1}$	5	3.8%	$8.83x10^{-1}$
8	CAATTA	11	9.6%	$5.09x10^{-2}$	13	10.0%	$2.82x10^{-2}$
9	ACAATT	6	5.3%	$8.21x10^{-1}$	12	9.2%	$2.07x10^{-1}$
10	CACAAT	5	4.4%	$6.43x10^{-1}$	10	7.7%	$9.72x10^{-2}$
11	TCACAA	7	6.1%	$2.99x10^{-1}$	13	10.0%	**$9.52x10^{-3}$**
12	ATCACA	12	10.5%	**$2.03x10^{-3}$**	12	9.2%	**$5.91x10^{-3}$**
13	TATCAC	13	11.4%	**$7.56x10^{-4}$**	12	9.2%	**$6.75x10^{-3}$**
14	CTATCA	10	8.8%	$2.39x10^{-2}$	10	7.7%	$5.13x10^{-2}$
15	CCTATC	7	6.1%	$1.53x10^{-2}$	8	6.2%	**$9.81x10^{-3}$**
16	CCCTAT	9	7.9%	**$3.71x10^{-3}$**	5	3.8%	$2.67x10^{-1}$
17	CCCCTA	10	8.8%	**$7.87x10^{-5}$**	5	3.8%	$1.05x10^{-1}$
18	ACCCCT	10	8.8%	**$7.43x10^{-4}$**	9	6.9%	**$6.50x10^{-3}$**

7mer 3'Motifs with a single mismatch

		Noncanonical			Canonical		
Position	Motif	# (of 114)	Percent	Enrichment (p-val)	# (of 130)	Percent	Enrichment (p-val)
1	GCATTAA	6	5.3%	$1.17x10^{-2}$	1	0.8%	$8.81x10^{-1}$
2	AGCATTA	5	4.4%	$3.73x10^{-2}$	3	2.3%	$3.44x10^{-1}$
3	TAGCATT	5	4.4%	$1.00x10^{-1}$	2	1.5%	$7.69x10^{-1}$
4	TTAGCAT	5	4.4%	$5.08x10^{-2}$	4	3.1%	$1.94x10^{-1}$
5	ATTAGCA	2	1.8%	$4.57x10^{-1}$	4	3.1%	$1.02x10^{-1}$
6	AATTAGC	2	1.8%	$4.27x10^{-1}$	1	0.8%	$8.10x10^{-1}$
7	CAATTAG	0	0.0%	1	0	0.0%	1
8	ACAATTA	1	0.9%	$8.69x10^{-1}$	4	3.1%	$2.01x10^{-1}$
9	CACAATT	2	1.8%	$4.19x10^{-1}$	1	0.8%	$8.05x10^{-1}$
10	TCACAAT	1	0.9%	$7.82x10^{-1}$	2	1.5%	$5.17x10^{-1}$
11	ATCACAA	3	2.6%	$1.43x10^{-1}$	7	5.4%	**$8.55x10^{-4}$**
12	TATCACA	4	3.5%	$7.15x10^{-2}$	5	3.8%	$3.37x10^{-2}$
13	CTATCAC	4	3.5%	$2.70x10^{-2}$	5	3.8%	**$9.85x10^{-3}$**
14	CCTATCA	3	2.6%	$3.27x10^{-2}$	5	3.8%	**$1.32x10^{-3}$**
15	CCCTATC	4	3.5%	**$3.90x10^{-3}$**	1	0.8%	$5.08x10^{-1}$
16	CCCCTAT	6	5.3%	**$6.10x10^{-5}$**	1	0.8%	$5.21x10^{-1}$
17	ACCCCTA	8	7.0%	**$1.79x10^{-6}$**	2	1.5%	$2.25x10^{-1}$

Figure S5. Enrichment of motifs complementary to the 3' end of miR-155 near canonical and non-canonical seed motifs, Related to Figure 4

There is similar 3' complementarity in both canonical and non-canonical sites. Exact 6mer seed matches (canonical) and inexact 6mer seed matches (1nt mismatch, non-canonical) were identified in miR-155 dependent sites ($p < 0.05$). The number of sites with an inexact (1 nt mismatch) motif in a region located 3-18nt 5' of the seed were counted. Enrichment is calculated by Fisher's exact test and is relative to the same sized region chosen from peaks lacking miR-155 seed matches.

Gene	Binding Site	Predicted binding site	Visualization
Igf2r	680-720	ATCTCTGTAGTTGCAGCTCTTT**GCACGAA**TCCCTCTGGGTGTTGCTCTTGCCA	
Rasa2	1300-1338	TCTCACTAGCTGCAATCCCCTGTCGTTG**AGCTTAA**GCTATAGTTCAAAATA	
Wasl	2275-2314	CATATCTAAATGCTGTGTGATAGACAATTTCAGCTAGCCACATATT**GTATTA**	

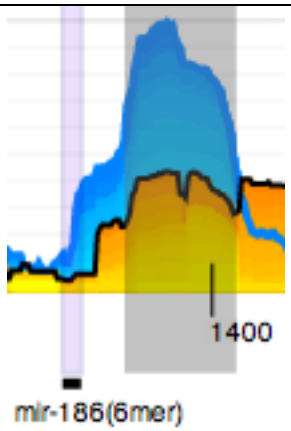
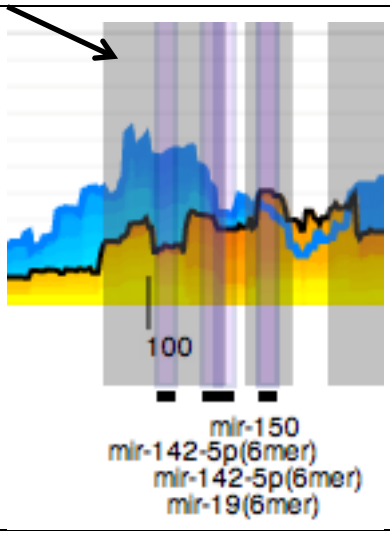
Tnrc6a	1754-1793	GAACTTGTCC**ACAATCA**CCCTGAGA TAATGTGAGAACAGTGGGAACTGTAG C	
Prkar1a	84-130	TGGCCACTGCTTCACAGCTTCTTGTC TCTTTATACTAAAAGTT**GCTTTA**TTGC ACCATT	

Figure S6. Non-canonical miR-155-binding sites, Related to Figure 5
Non-canonical sites from luciferase assays shown in Figure 5b that are not shown in Figure 3. Bolded sequence indicates predicted site of miR-155 binding.

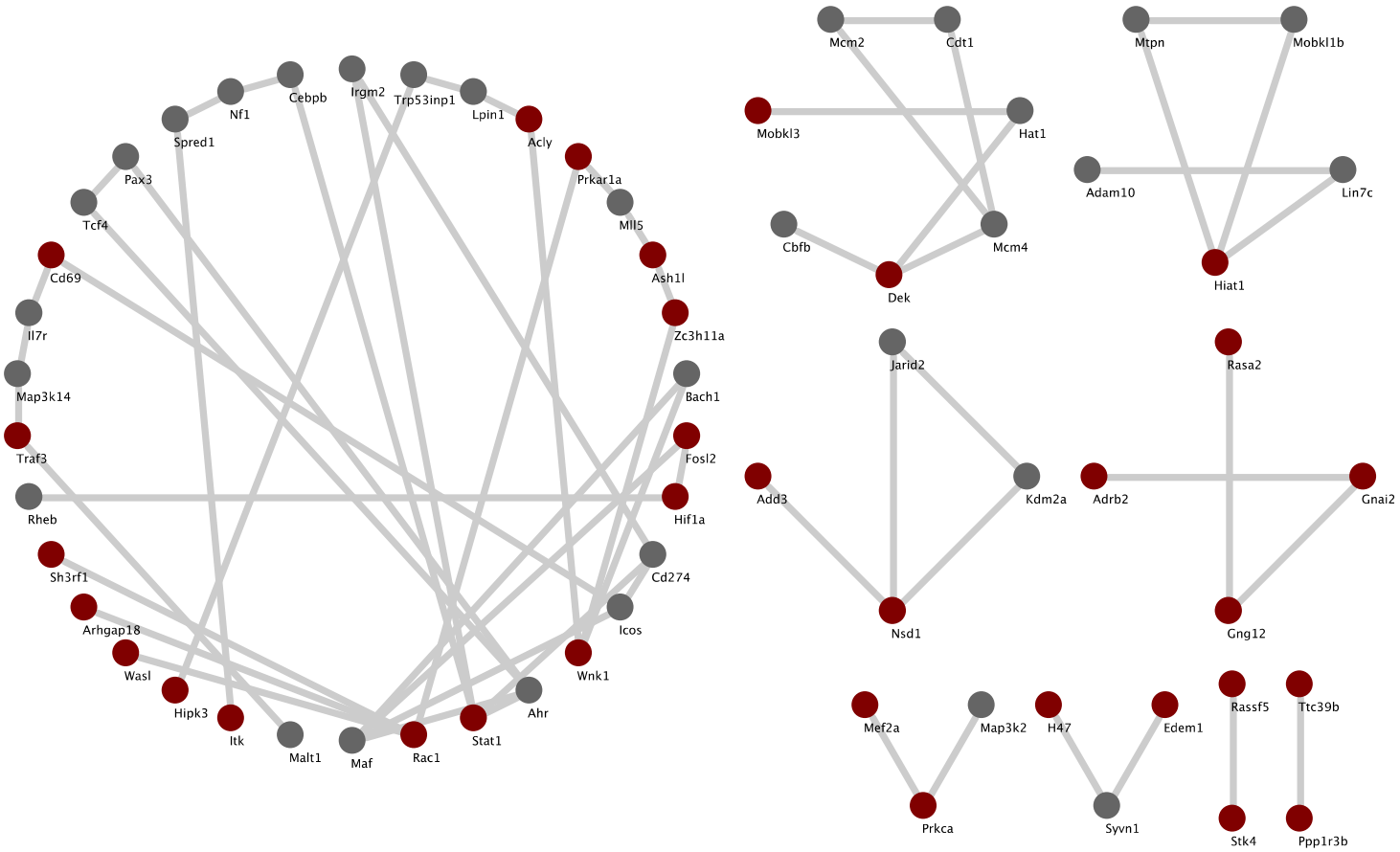


Figure S7. miR-155 targets multiple sets of functionally related genes, Related to Table 1

We examined 175 genes with miR-155 dependent sites ($P < .01$) in the context of known and predicted gene-gene interactions using the STRING database (default settings) (Szklarczyk et al., 2011) and identified connected sub-networks of genes. We display all (9) connected sub-network components with at least two members. Red circles represent genes with a putative miR-155 non-canonical site. Grey circles represent genes with only canonical miR-155 sites. Lines between circles represent known or predicted interactions.

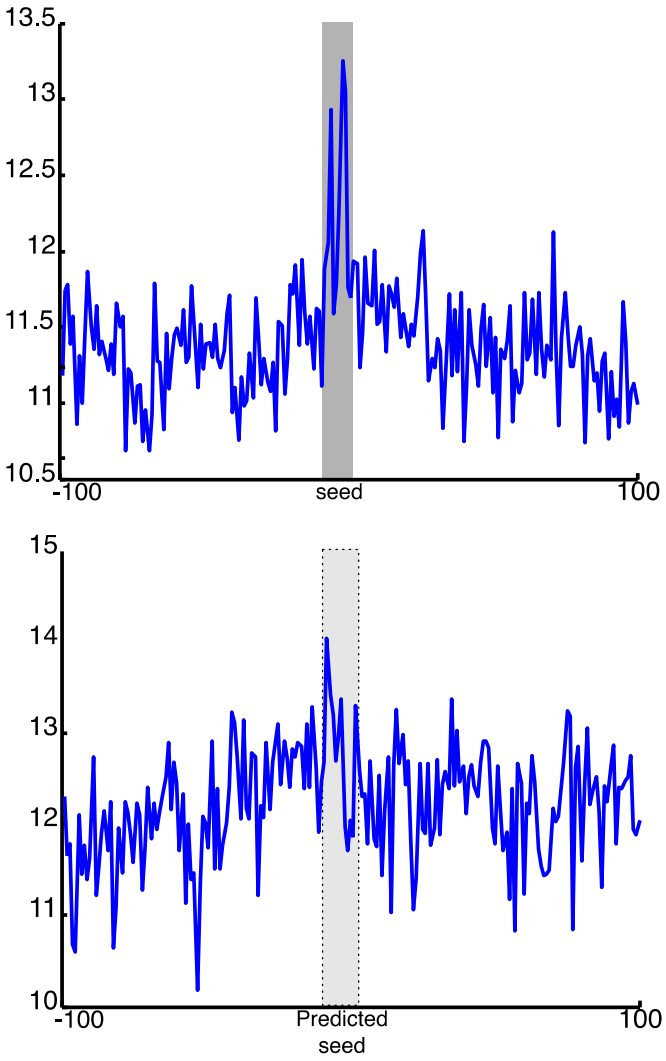


Figure S8. Conservation of canonical and non-canonical target sites, Related to Figure 4

Conservation of regions surrounding AGO bound canonical and non-canonical miR-155 sites. Regions were centered on canonical or non-canonical miR-155 seeds. 23-way multiple species alignments were used to calculate the median number of species with a nucleotide identical to the mouse nucleotide at a given position relative to the seed.

Supplemental References

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479-486.

Conway, J.H., and Sloane, N.J.A. (1986). Lexicographic codes: Error-correcting codes from game theory. IEEE Trans. Inf. Theor. *32*, 337-348.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). Human MicroRNA targets. PLoS Biol *2*, e363.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Mu, P., Han, Y.C., Betel, D., Yao, E., Squatrito, M., Ogrodowski, P., de Stanchina, E., D'Andrea, A., Sander, C., and Ventura, A. (2009). Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. Genes Dev *23*, 2806-2811.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res *15*, 1281-1295.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P.*, et al.* (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res *39*, D561-568.

Wu, H., Neilson, J.R., Kumar, P., Manocha, M., Shankar, P., Sharp, P.A., and Manjunath, N. (2007). miRNA profiling of naive, effector and memory CD8 T cells. PLoS One *2*, e1020.

Yu, C.N., Joachims, T., Elber, R., and Pillardy, J. (2008). Support vector training of protein alignment models. J Comput Biol *15*, 867-880.

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol *29*, 607-614.