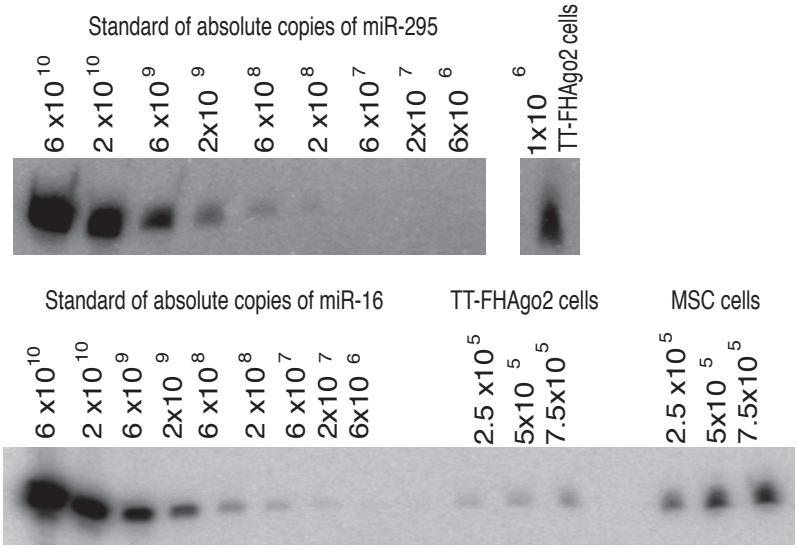
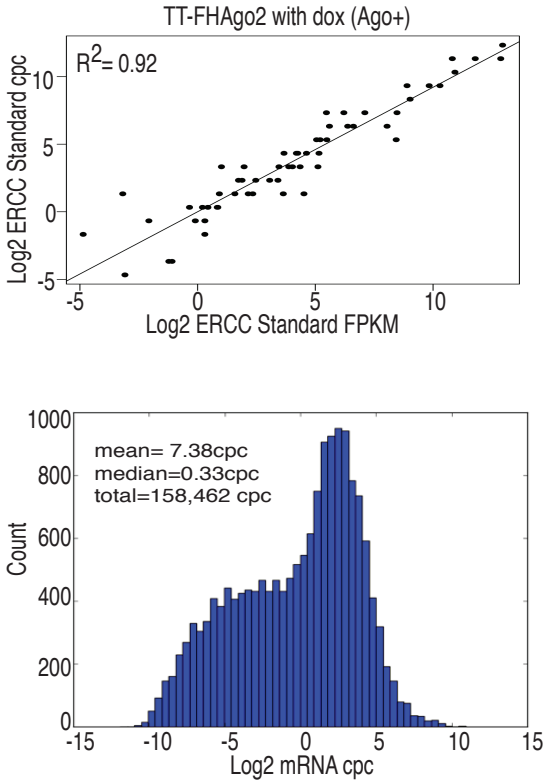
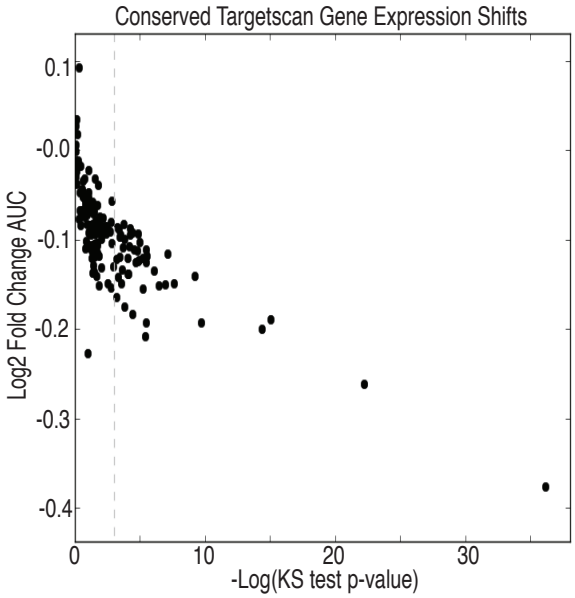
A


Standard of absolute copies of miR-295
6 x10 10
2 x10 10
6 x10 9
2x10 9
6 x10 8
2 x10 8
6 x10 7
2x10 7
6x10 6
1x10 6
TT-FHAgo2 cells
Standard of absolute copies of miR-16
TT-FHAgo2 cells
MSC cells
6 x10 10
2 x10 10
6 x10 9
2x10 9
6 x10 8
2 x10 8
6 x10 7
2x10 7
6x10 6
2.5 x10 5
5x10 5
7.5x10 5
2.5 x10 5
5x10 5
7.5x10 5

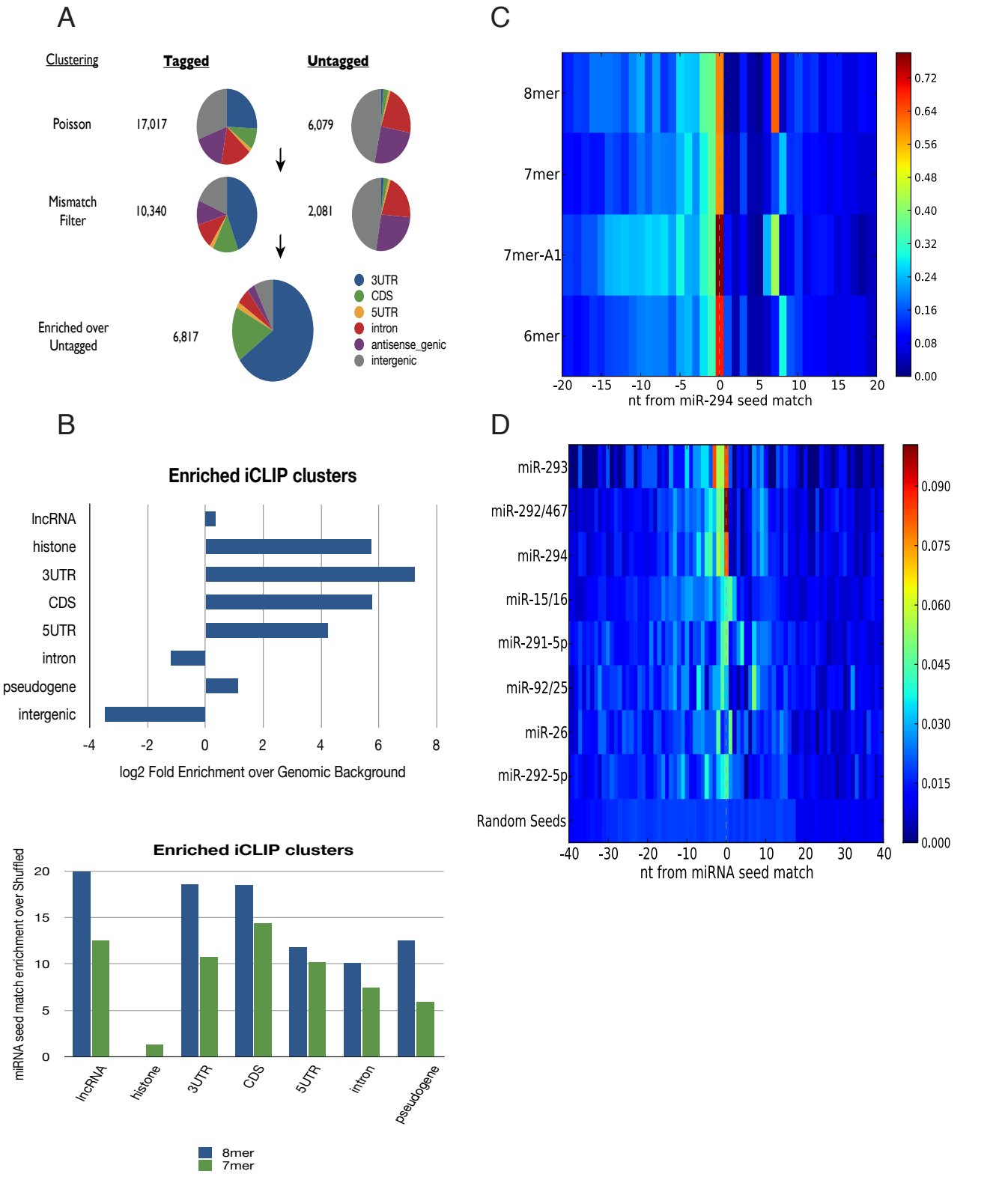
B


TT-FHAgo2 with dox (Ago+)
R2= 0.92
Log2 ERCC Standard cpc
Log2 ERCC Standard FPKM
mean= 7.38cpc
median=0.33cpc
total=158,462 cpc
Count
Log2 mRNA cpc

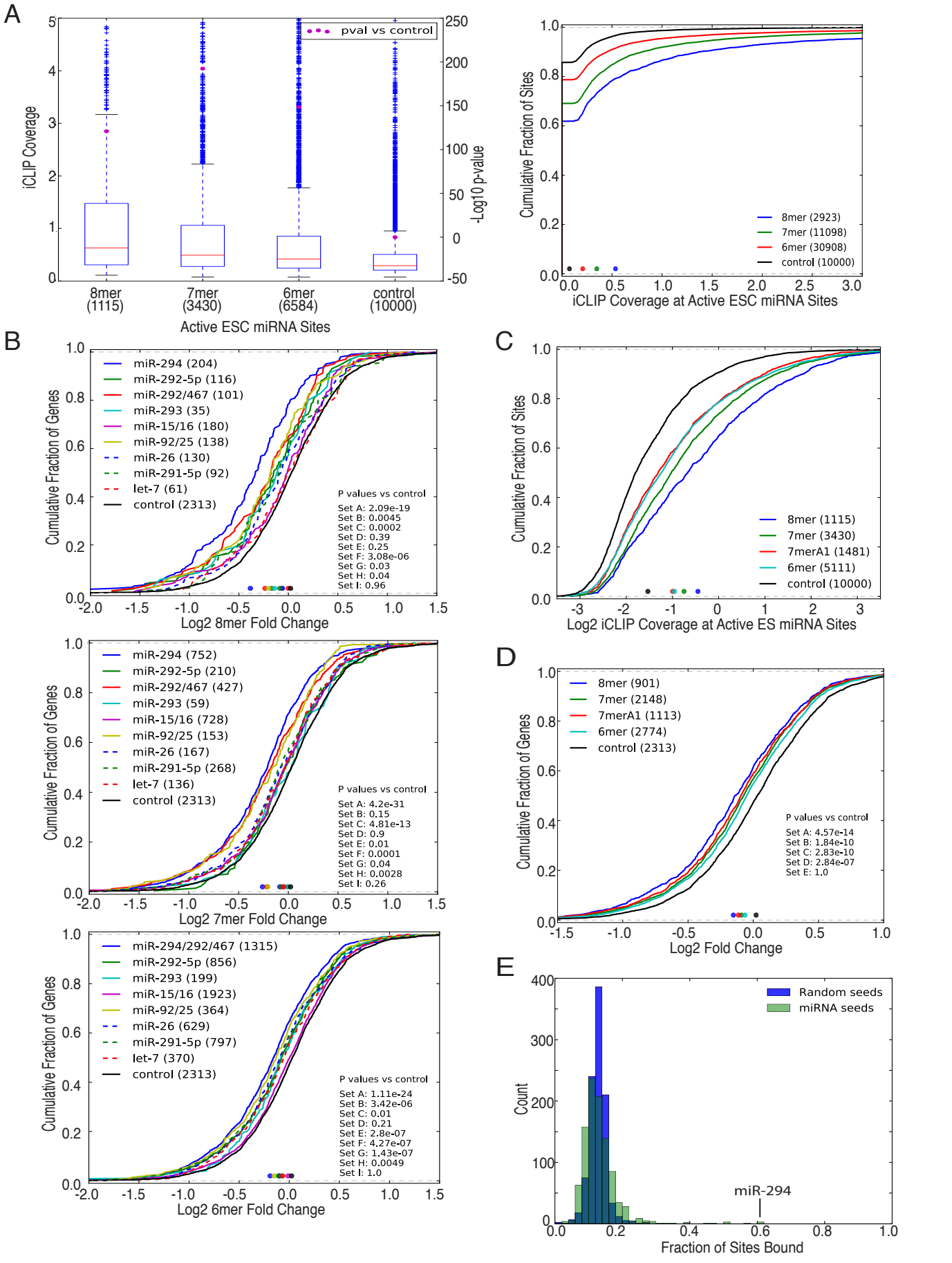
C


Conserved Targetscan Gene Expression Shifts
Log2 Fold Change AUC
-Log(KS test p-value)

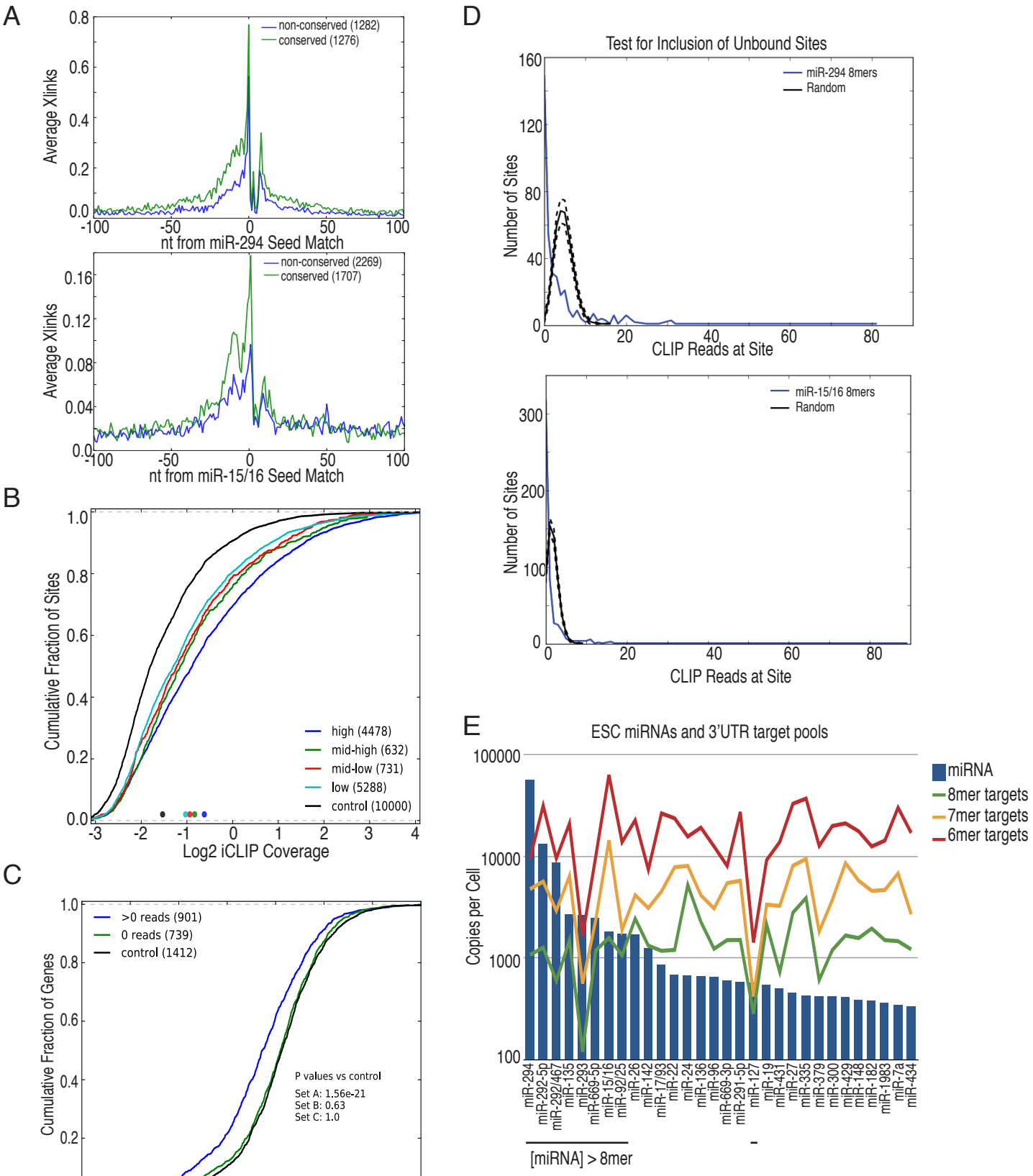
Bosson_FigureS2

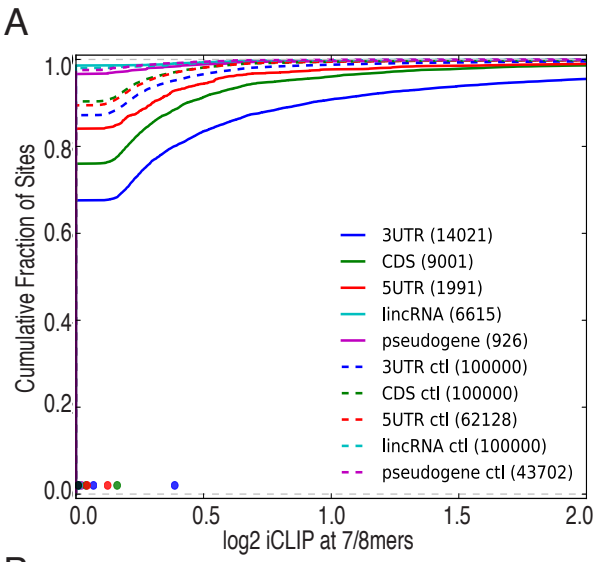


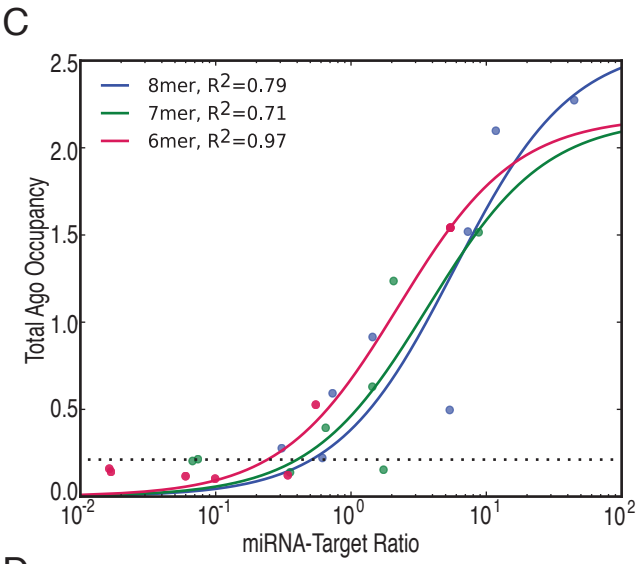
Bosson_FigureS3

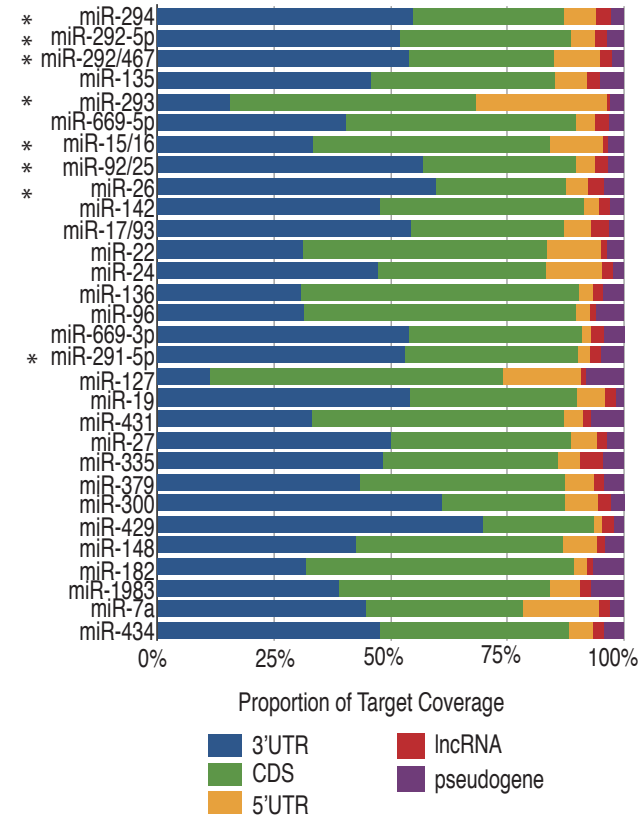


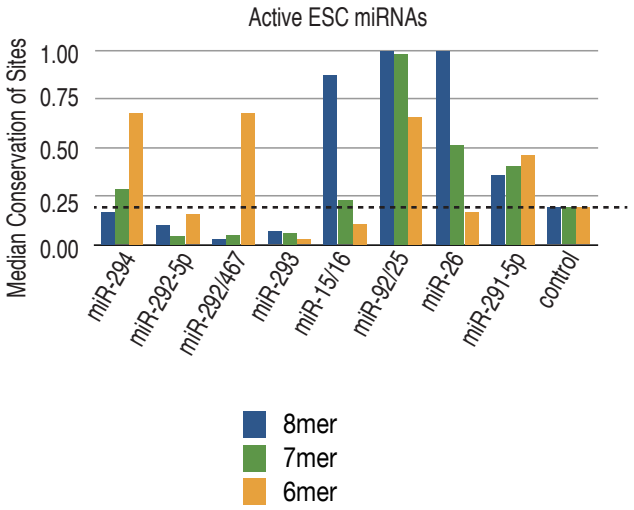
Bosson_FigureS4



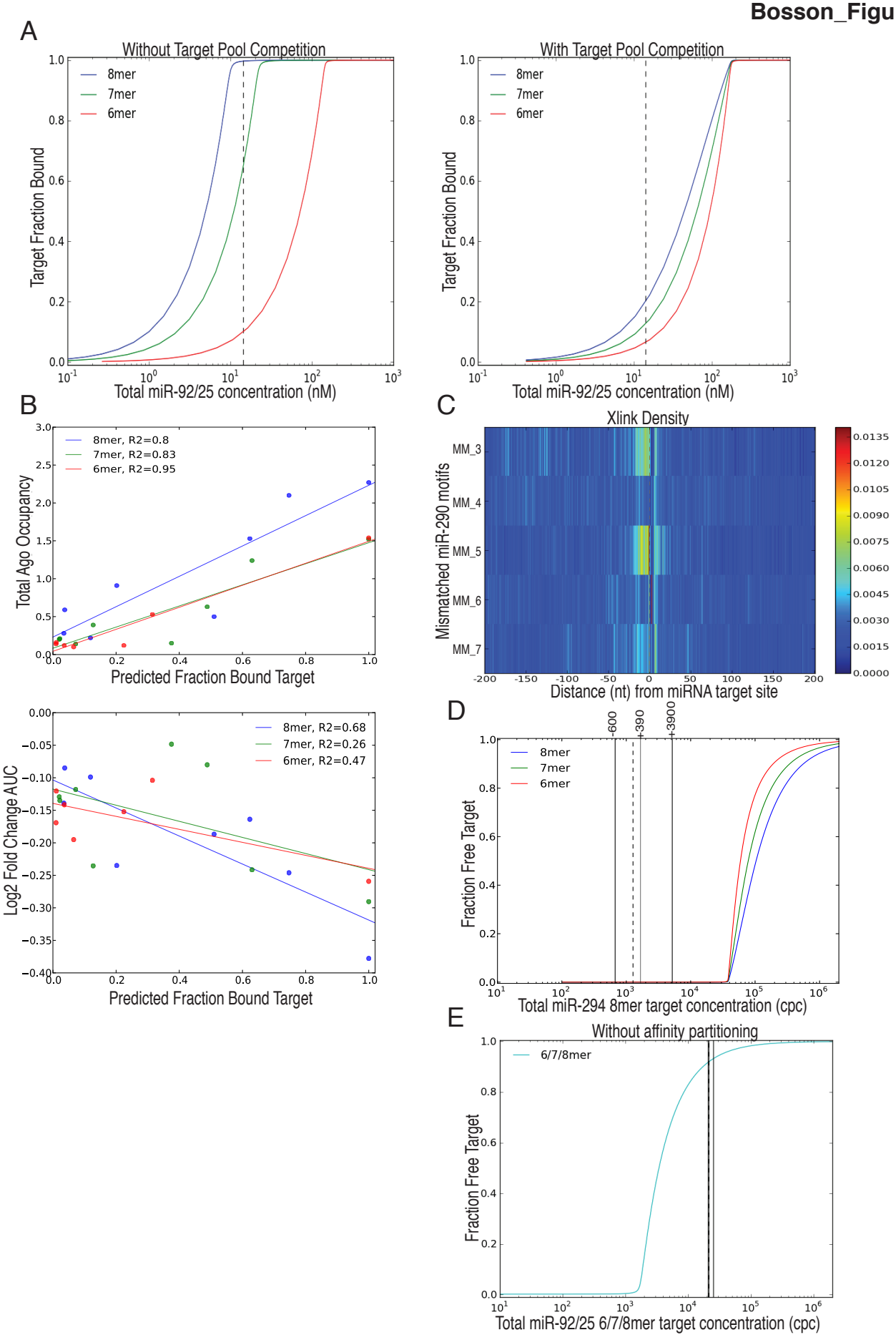

A
Cumulative Fraction of Sites
log2 iCLIP at 7/8mers
3UTR (14021)
CDS (9001)
5UTR (1991)
lincRNA (6615)
pseudogene (926)
3UTR ctl (100000)
CDS ctl (100000)
5UTR ctl (62128)
lincRNA ctl (100000)
pseudogene ctl (43702)

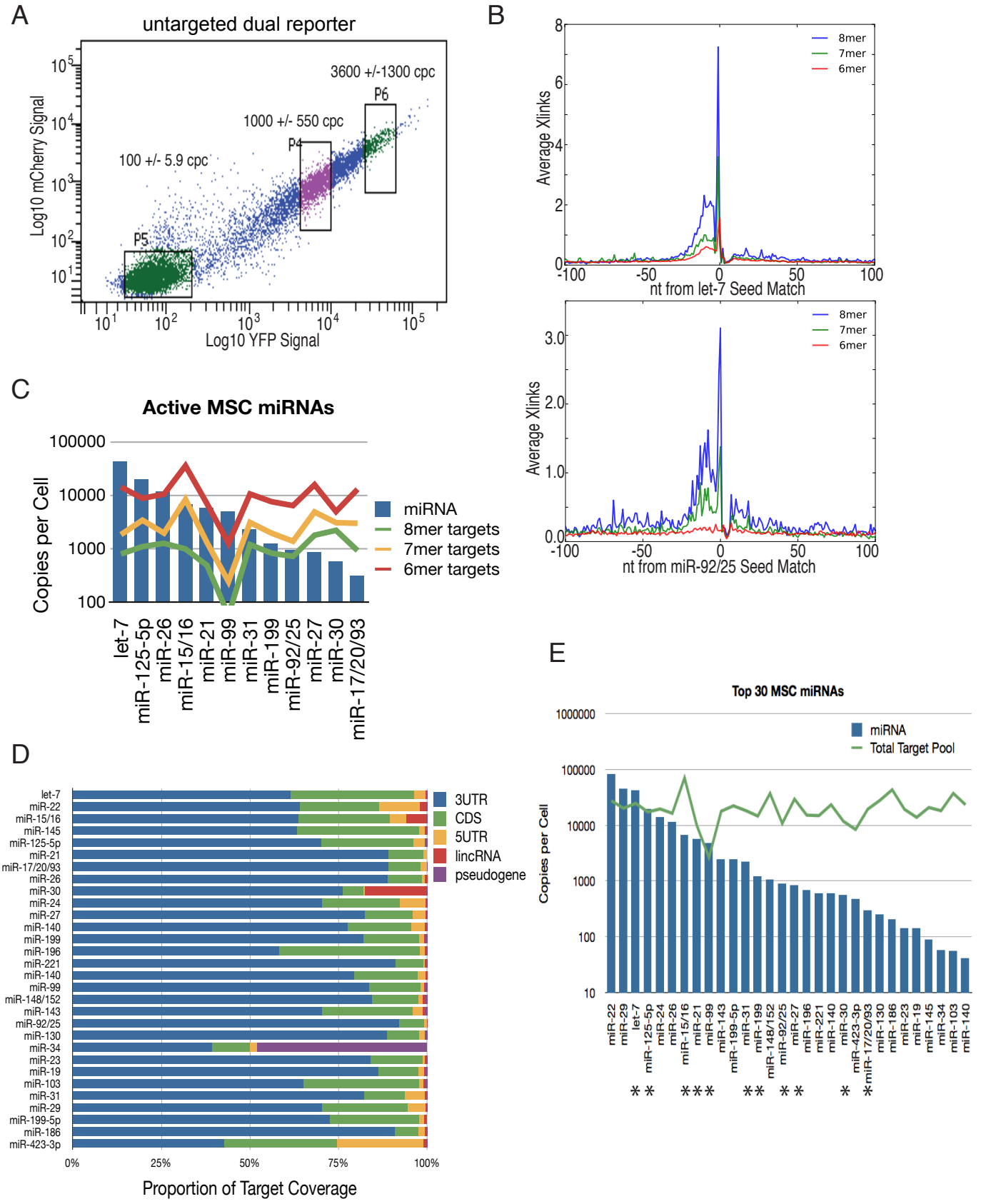

C
8mer, R2=0.79
7mer, R2=0.71
6mer, R2=0.97
Total Ago Occupancy
miRNA-Target Ratio


B
miR-294
miR-292-5p
miR-292/467
miR-135
miR-293
miR-669-5p
miR-15/16
miR-92/25
miR-26
miR-142
miR-17/93
miR-22
miR-24
miR-136
miR-96
miR-669-3p
miR-291-5p
miR-127
miR-19
miR-431
miR-27
miR-335
miR-379
miR-300
miR-429
miR-148
miR-182
miR-1983
miR-7a
miR-434
0%
25%
50%
75%
100%
Proportion of Target Coverage
3'UTR
CDS
5'UTR
lncRNA
pseudogene


D
Active ESC miRNAs
Median Conservation of Sites
miR-294
miR-292-5p
miR-292/467
miR-293
miR-15/16
miR-92/25
miR-26
miR-291-5p
control
8mer
7mer
6mer

Bosson_FigureS6




A
untargeted dual reporter
Log10 mCherry Signal
Log10 YFP Signal
100 +/- 5.9 cpc
1000 +/- 550 cpc
3600 +/-1300 cpc
P5
P4
P6
B
Average Xlinks
nt from let-7 Seed Match
nt from miR-92/25 Seed Match
8mer
7mer
6mer
C
Active MSC miRNAs
Copies per Cell
miRNA
8mer targets
7mer targets
6mer targets
let-7
miR-125-5p
miR-26
miR-15/16
miR-21
miR-99
miR-31
miR-199
miR-92/25
miR-27
miR-30
miR-17/20/93
D
3UTR
CDS
5UTR
lincRNA
pseudogene
Proportion of Target Coverage
E
Top 30 MSC miRNAs
Copies per Cell
miRNA
Total Target Pool

Supplemental Figure Legends

Figure S1. Related to Figure 1, Quantitation of miRNA and mRNA absolute levels and fold repression in presence of Ago

(A) Northern blot of synthetic and endogenous miRNAs. Synthetic single-stranded miR-295 and miR-16 RNA oligonucleotides were titrated in absolute number of molecules as indicated. TT-FHAgo2 ESCs and WT MSCs were counted, and total RNA was extracted. Total RNA corresponding to the indicated number of cells were loading on Northern. Northern blot signal was quantified with ImageQuantTL. The highest signal for each cell type (i.e. miR-295 $1x10^6$ for ESC and miR-16 $7.5x10^5$ for MSC) was fit to the standard curve from synthetic titration signal to give final copies per cell. A normalization factor was calculated by dividing the measured cpc value by the small RNAseq counts for miR-295 (ESC) or miR-16 (MSC), and all other miRNAs were normalized to this factor.
(B) (top) The known concentration of External RNA Controls Consortium (ERCC) standards in copies per cell compared to the Fragments per Kilobase per Million mapped reads (FPKM) values calculated with Cufflinks showing the linearity of RNA-seq quantitation for the indicated range of cpc. (bottom) Histogram of the log2 measured cpc for every annotated protein-coding gene in ESCs. Mean, median, and total mRNA copies per cell for ESCs indicated in top left.
(C) TT-FHAgo2 cells were starved from dox for 48 hours (to remove all Ago expression) and then either reintroduced to dox for 48 hours (Ago+) or not (Ago-). Gene repression (fold change in expression between Ago+ and Ago- cells) for all TargetScan predicted conserved targets of each of the 151 conserved TargetScan miRNA families was calculated relative to matched control genes (Supplemental Methods). X-axis is the –log10 p-value from a two-sided Kolmogorov-Smirnov test of the TargetScan gene set versus matched control gene set. p=0.001 value is indicated as dotted grey vertical line.

Figure S2. Related to Figures 1 and 2, description of Ago2 iCLIP enriched binding sites

(A) Schematic of computational pipeline to call significant Ago2 iCLIP clusters genome-wide (as described in more detail in methods) and the number of clusters identified at each filtering step. Pie charts represent proportion of clusters in indicated Ensembl annotated genomic regions. Final enriched over untagged set called using a p-value cutoff of 0.05 (binomial test).
(B) (top) Log2 fold enrichment of the number of ESC FHAgo2 iCLIP clusters in annotated Ensembl genomic regions relative to the average number of clusters in each region when randomly shuffling the clusters from each set across the genome 100 times. (bottom) Barplot of the miRNA seed match enrichment in clusters by annotated Ensembl genomic regions. Counted are the number of 8mer or 7mer seed matches for the top 99% expressed miRNAs in ESCs found in FHAgo2 iCLIP clusters from each region. These were compared to seed match counts after shuffling the cluster sequences. Fold enrichment was calculated as number of seed matches in real clusters divided by average number of seed matches found in shuffled sequence clusters (100 iterations). Clusters from all regions but histone have more seed matches than expected by chance.
(C) Heatmap of average crosslinks (Xlinks) per nt across all miR-294 target sites in expressed 3'UTRs that have at least 1 iCLIP read within 10nt. Seed match site types are

plotted individually – [AG]GCACTTA, [AG]GCACTT, GCACTT. 0 in the X-axis corresponds to the first nucleotide of the 7mer seed match (i.e. across from position 8 of the miRNA). Average Xlinks color scale is indicated to the right. Xlinks at position 7 (i.e. across from the 1st nt of the miRNA) is specifically enriched at 8mer sites.
(D) Heatmap of average Xlinks per nt across all 6/7/8mer sites in expressed 3'UTRs that have exactly 1 iCLIP read within 10nt, for each of the active ESC miRNA families, ranked by Xlink signal. 0 in the X-axis corresponds to the first nucleotide of the 7mer seed match (i.e. across from position 8 of the miRNA). Average Xlinks color scale is indicated to the right.
To test for significance of Xlink signal, we calculated the iCLIP Xlink distribution across 1000 control seed "families" of random sites in expressed 3'UTRs, with same number of sites per family as the average ESC miRNA family. As with ESC miRNA seed families, this analysis was limited to control sites with exactly 1 read within 10nt of the site. The average Xlink signal per nt of the 1000 random families is plotted on the bottom row. All active ES miRNA families exhibit significant Xlink signal at either position 0 or +1 ($p<0.01$ empirical p-value from random seed distribution).

Figure S3. Related to Figure 3, iCLIP coverage and repression of 3'UTR target sites by seed match type
(A) (left) Box plots displaying median and lower/upper quartiles of the iCLIP coverage per 3'UTR site of the indicated site type, for all active ESC miRNAs. iCLIP coverage is read count normalized to isoform expression (Supplemental Methods). Only sites with >= 1 iCLIP read were included. Number of sites in each set is indicated in parentheses. – log10 p values from two-sided wilcoxon rank-sum test of each set versus control are displayed as purple dots corresponding to the y-axis on the right.
(right) Cumulative distributions of same data as in (A) except sites with 0 reads are included to illustrate the differing fractions of non-bound sites (y-intercepts). Colored dots at bottom represent mean iCLIP coverage.
(B) Target gene expression changes corresponding to sites used in iCLIP coverage plots shown in Figure 3A-C. Cumulative distribution of log2 Ago+/Ago- fold change in expression for all genes containing 3'UTR 8mer (top), 7mer (middle), or 6mer (bottom) target sites with >=1 iCLIP read for each active ESC miRNA family, let-7, and random control sites. Target sites of the non-expressed let-7 miRNA family are included as an additional estimate of background signal in this analysis. Number of genes in each set is indicated in legend parentheses. Colored dots at bottom represent mean log2 Gene Repression. P values of two-sided Kolmogorov-Smirnov (KS) test of each gene set versus control are listed at bottom right in the same order as in the line color legend.
(C) Cumulative distribution of log2 iCLIP coverage at each 3'UTR target site for the active ESC miRNA families, plotted for each site type separately. 6mer sites are split into those with an A after the last matching nt (7merA1) or without an A (6mer). Only sites with >= 1 iCLIP read were included. Number of sites in each set is indicated in legend parentheses. Colored dots at bottom represent mean log2 iCLIP coverage. There is no statistical difference between 6mer and 7merA1 sites ($p=0.91$, two-sided KS test).
(D) Cumulative distribution of log2 Ago+/Ago- Gene Repression for all genes containing 3'UTR target sites with >=1 iCLIP read for any active ESC miRNA family. Target genes with 6mer sites are split into those with an A after the last matching nt (7merA1) or

without an A (6mer). Number of genes in each set is indicated in legend parentheses. Colored dots at bottom represent mean log2 Gene Repression. P-values of two-sided KS test of each gene set versus control are listed at bottom right. Genes with 7merA1 sites are more repressed than those with either 7mer(m8) or 6mer sites.
(E) Histogram of the fraction of 3'UTR 8mer sites with >=1 iCLIP read within 10nt for all miRNA families (green). Normalized histogram of values for 998 random seed "families" is in blue. Fraction of bound sites for the miR-294 family is indicated, which despite being expressed well above its target pool still only binds ~60% of its 8mer seed match sites in expressed 3'UTRs.

Figure S4. Related to Figure 3, target pool estimation with iCLIP
(A) Average Xlinks per nt across all conserved (green) or non-conserved (blue) 6/7/8mer target sites in expressed 3'UTRs that have at least 1 iCLIP read within 10nt, for highly expressed miR-294 family (top) and lowly expressed miR-15/16 (bottom). Conserved sites are defined as those with average PhastCons score > 0.8. Non-conserved sites are those with average PhastCons score < 0.2. Number of sites in each set is indicated in legend parentheses. 0 in the X-axis corresponds to the first nucleotide of the 7mer seed match (i.e. across from position 8 of the miRNA).
(B) Cumulative distribution of log2 iCLIP coverage at each 3'UTR 6/7/8mer target site for the active ESC miRNA families, plotted for each level of conservation separately. Conservation level boundaries are 0.2, 0.5, and 0.8 average PhastCons score. Only sites with >= 1 iCLIP read were included. Number of sites in each set is indicated in legend parentheses. Colored dots at bottom represent mean log2 iCLIP coverage.
(C) Cumulative distribution of log2 Ago+/Ago- fold change in expression for all genes containing 3'UTR 8mer target sites with >=1 iCLIP read (blue) or 0 iCLIP reads (green) within 10nt for any active ESC miRNA family. Any gene that also contained a 3'UTR 6/7/8mer target site with >=1 iCLIP read was removed from the "0 reads" and control sets. Number of genes in each set is indicated in legend parentheses. Colored dots at bottom represent mean log2 Gene Repression. Control is set of random genes with 3'UTRs that are matched with the active ESC miRNA targets for length, GC content, and expression. P values of two-sided KS test of each gene set versus control are listed at bottom right. Genes that only contain an 8mer target site with no evidence of iCLIP coverage are not significantly more repressed than control.
(D) Histograms of observed number of reads at each miR-294 (top) and miR-15/16 8mer site (bottom) and expected number of reads at each site if distributed randomly across the corresponding number of sites (Methods). Dotted black lines represent standard deviation of 100 random shuffling simulations. The number of 0 bound sites expected by chance at the given depth for each miRNA site type pair (y-intercept of black line) was added to the corresponding total target pool, as described in methods.
(E) Measured cpc values for miRNA and corresponding 3'UTR target pools of indicated site type. Same data as Figure 3D, but all of the top 30 expressed ESC miRNAs are plotted. Y-axis is log scale.

Figure S5. Related to figure 4, total target pool composition, binding kinetics, and conservation

(A) Cumulative distribution of iCLIP coverage at each 7/8mer target site for the active ESC miRNA families. Distribution for sites from each genic category plotted separately with corresponding random control sites (dashed lines). All 7/8mer seed matches (rather than only bound sites) are included to illustrate the different fractions of non-bound sites (y-intercepts). Greater than 95% of all expressed pseudogene and lincRNA 7/8mer ES miRNA sites show 0 iCLIP coverage. Number of sites in each set is indicated in legend parentheses. Colored dots at bottom represent mean iCLIP coverage.
(B) Proportion of iCLIP-estimated target abundance coming from indicated genic categories for each of the top 30 expressed ESC miRNAs. Target pool expression is not weighted by genic region iCLIP coverage here. Asterisks mark significantly active families identified in Figure 3B.
(C) Plot of miRNA and target stoichiometries per site type and Total Ago Occupancy as in Figure 5A, but values from each site type are fitted to a hyperbolic equation. R^2 values for best fit line are indicated in the legend. Non-linear least squares was used to fit the equation $y = Bmax*x/(x+K_D)$ to the data. Best fit K_D values (in "units" of miRNA-Target ratio) are 8.24, 5.25, and 3.19 for 8mers, 7mers, and 6mers, respectively. miRNA and target stoichiometries per site type are plotted against total binding estimates from iCLIP, for each active ESC miRNA. miRNA-Target Ratio is miRNA cpc divided by iCLIP-estimated cpc for each individual site type target pool. Total Ago Occupancy is iCLIP RPM across all expressed target sites of a given site type divided by iCLIP-estimated cpc for that site type target pool. The p=0.01 empirical p-value calculated from Total Ago Occupancy values of 1000 random sets of control sites is indicated by dotted line. X-axis is log scale.
(D) Conservation was calculated for each target site in expressed 3'UTRs by averaging the PhastCons score for each nt. The median target site conservation score of each site type for all the active ESC miRNAs is shown. Only sites with >= 1 iCLIP read were included. miRNAs are listed from left to right in order of expression in ESCs. We also show the median conservation score for random control sites within 3'UTRs that are matched with the active ESC miRNA targets for length, GC content, and expression. Interestingly, miR-292-5p and miR-293 target sites show decreased conservation relative to random sites. The miR-292-5p family is made of miRNAs that are from the passenger strand of the highly expressed miR-294/292/467 families, and miR-293 is expressed in the same cluster with miR-294/292 miRNAs but has a 3-shifted seed sequence relative to miR-292 (therefore it constitutes a separate family). Also, we see median conservation is inversely correlated with expression and affinity, i.e. 6mer sites of highly expressed miRNAs are more conserved than 7/8mer sites, and 8mer sites of lowly expressed miRNAs are more conserved than control. This phenomenon holds true even if looking at median conservation of all ESC expressed sites rather than just those with >=1 iCLIP reads (data not shown).

Figure S6. Related to Figure 5, mathematical model of miRNA binding is supported by iCLIP and gene expression data and predicts characteristics of miRNA threshold responses

(A) Different fraction bound predicted at endogenous miRNA concentration between.model with and without intratarget pool competition. miRNA titration curves for a representative lowly expressed active miRNA (miR-92/25) showing the relationship

between total miRNA concentration and the predicted fraction bound targets for each site type, indicated by line color. (left) Target fraction bound is calculated individually for each target pool (based on corresponding K_D) without regard for other target pools. (right) Target fraction bound is calculated for all target pools competing for the same pool of miRNA (see Methods). The endogenous miR-92/25 concentration is indicated by a dotted black vertical line. X axis is log scale.

(B) Correlation of predicted fraction bound target per site type affinity group versus iCLIP Total Ago Occupancy (top) or log2 fold change AUC (bottom) for each of the active ESC miRNAs. Total Ago Occupancy is iCLIP reads per million (RPM) across all expressed target sites of a given site type divided by iCLIP-estimated cpc for that site type target pool. log2 fold change AUC is calculated as the area under the curve between the cumulative distribution of log2 Ago+/Ago- FPKM test set values and matched control gene values, considering genes containing a target site with >= 1 iCLIP read for a given site type. The least squares linear regression best fit line for each site type is depicted in the indicated colors with the coefficient of determination noted (R^2).

(C) Heatmap of Xlink density per nt across all miR-294 7mer seed matches in expressed 3'UTRs that contain a single mismatch (MM) across from the miRNA position indicated to the left. For instance, "MM_3" corresponds to any instances of AGCAC[CGA]T sequences. 0 in the X-axis corresponds to the first nucleotide of the 7mer seed match (i.e. across from position 8 of the miRNA). Xlink density is calculated as the number of read 5' ends (Xlinks) at a given nt position, normalized to the total number of Xlinks within a +/- 200nt window around the target sites. Xlink density color scale is indicated to the right.

(D) Simulated target titration curves for the highly expressed miR-294 family showing the relationship between proportion of free targets and total 8mer target pool concentration in cpc. Dotted black vertical line indicates endogenous 8mer target pool concentration. Gray-black solid vertical lines indicate estimates of physiological ESC ceRNA perturbations, corresponding to loss (left of dotted line) of a highly expressed mRNA (200 cpc) containing 3x8mer sites (600 site cpc) or 10 and 100 -fold upregulation (right of dotted line) of an average target gene (13 cpc) containing 3x8mer sites (390, 3900 site cpc). X-axis is log scale. The miR-294 family is expressed highly enough to saturate its complete 6/7/8mer target pool and therefore is not susceptible to physiological perturbations in target concentration. Since iCLIP coverage of miR-294 sites is not equally saturated at 6mers, 7mers, and 8mers, it is likely that other, more difficult to define, non-canonical weak affinity sites meaningfully contribute to the miR-294 target pool, as discussed in the text, but are not modeled here.

(E) Simulated target titration curves for the lower expressed miR-92/25 family showing the relationship between proportion of free targets and total target pool concentration in cpc, when not partitioning the target pool into exclusive site type affinity groups. Concentration of all 6/7/8mer sites were summed together, and the 26 pM 7mer K_D was applied to all sites. Dotted black vertical line indicates endogenous total target pool concentration. Gray-black solid vertical lines indicate estimates of physiological ESC ceRNA perturbations, corresponding to loss (left of dotted line) of a highly expressed mRNA (200 cpc) containing 3 sites (600 site cpc) or 10 and 100 -fold upregulation (right of dotted line) of an average target gene (13 cpc) containing 3 sites (390, 3900 site cpc). X-axis is log scale. The predicted fraction free target for miR-92/25 using this model is

92% at the endogenous target pool concentration and only shifts to 93.5% after simulated addition of 3900 competing sites. These results are less consistent with the iCLIP and reporter data than using a model with affinity-partitioned target pools, as described in the text.

Figure S7. Related to Figure 6, iCLIP in MSCs reveals miRNA target pools with similar target pool composition and site type binding pattern as in ESCs

(A) Plot of mCherry (PE-Texas Red-A) and eYFP (FITC-A) values for ES cells transfected with untargeted dual reporter, for isolation and quantification of mCherry RNA transcripts corresponding to fluorescent protein expression levels. The full range of fluorescent values are shown to allow depiction of background signal. Black boxes indicate the gates used to isolate cells in range of eYFP expression corresponding to background (P5), "Mid" (P4) and "High" (P6) overexpression of reporters in ESCs. Values for absolute quantitation of mCherry transcripts are indicated above gate with standard deviation from 2 biological replicates.

(B) Average Xlinks per nt across all target sites in expressed 3'UTRs that have at least 1 iCLIP read within 10nt, for highly expressed let-7 family (top) and lowly expressed miR-92/25 (bottom). Distribution across each site type is plotted individually. 0 in the X-axis corresponds to the first nucleotide of the 7mer seed match (i.e. across from position 8 of the miRNA). Similarly to miR-294 in ESC, the highly expressed let-7 miRNA shows strong Xlink signal across all site type, but the lowly expressed miR-92/25 does not detectably crosslink its 6mer sites. Note the difference in scale. let-7 6mers are bound to a similar level as miR-92/25 7mers. let-7 8mers exhibit a much higher average Xlink per site signal than any miRNA in ESCs, but this may be due to increased depth of the MSC iCLIP (6-fold higher collapsed iCLIP reads in MSC dataset compared to ESC). The MSC iCLIP cloning protocol included a random nucleotide barcode in the 3' adaptor ligated to Ago-bound RNA, such that sequenced reads could be collapsed after mapping based on the random barcode and not just identical 5' and 3' mapped coordinates. This may allow for increased depth of MSC Xlink signal since many of the ESC reads with identical 5' ends would have been collapsed and only counted once.

(C) Measured copies per cell (cpc) values for miRNA and corresponding 3'UTR target pools of indicated site type for significantly active MSC miRNAs. Y-axis is log scale. As in ESC, the majority of active MSC miRNAs are expressed below their 3'UTR 6mer and 7mer target pool levels, but there are more intermediate expressed miRNAs in MSCs than in ESCs.

(D) Proportion of iCLIP reads at 7/8mer target sites coming from indicated genic categories for each of the top 30 expressed MSC miRNAs. Again, ~75% of total Ago binding signal comes from 3'UTR. One notable exception is miR-34, which has 50% of its binding in pseudogenes.

(E) Measured copies per cell (cpc) values for miRNA and corresponding total target abundance, including all iCLIP-estimated 6/7/8mer sites from all genic categories. Top 30 expressed MSC miRNAs are plotted. Y-axis is log scale. All but 5 miRNAs in MSC are expressed below their total target pool. However, the estimated MSC miRNA total target pools are smaller on average than in ESCs (21,847 cpc in MSC vs 37,139 cpc in ESC, Figure 4C). This difference may be biological, but it is also possible that the

increased depth of the MSC iCLIP dataset allowed a more accurate estimate of which target sites to include in the target abundance estimates (Methods).

Supplemental Experimental Procedures

Cell Culture

ESCs were cultured in DME/HEPES supplemented with 2mM L-glutamine, 100 Units Penicillin, 100 ug Streptomycin, 1X Non-essential amino-acids (Invitrogen), 15% Defined FBS (Hyclone-ES screened), 1000U/ml ESGRO (Chemicon) on gelatinized flasks. MSCs were cultured in α-MEM supplemented with 10% FBS, 100 Units Penicillin, and 100 ug Streptomycin. TT-FHAgo2 (Ago1-4$^{-/-}$;Cre-ER, FHAgo2) and TT-Ago2 (Ago1-4$^{-/-}$;Cre-ER,Ago2) cell lines are described in Zamudio et al., 2014. Cells were maintained in 0.1 mg/ml doxycycline (Sigma) for propagation. MSC cells used for iCLIP are described in (Gurtan et al., 2013). They contain stable integrations of dox-inducible untagged or Flag-HA tagged hAgo2 constructs, similarly to ESCs.

small RNA sequencing and quantitation

For ESC small RNA library preparation input, RNA was isolated using Trizol (Life Technologies) and then size selected on denaturing polyacrylamide gels for 18-75nt small RNA. A 5' phosphate-dependent cloning library was generated using the NEBNext Small RNA Library Prep Set for Illumina (New England Biolabs) as described in Zamudio et al., 2014. DNA sequences obtained from the Illumina HiSeq 2000 Sequencing system were split by barcode, trimmed of adapter sequence with Cutadapt (Martin, 2011) and mapped to UCSC mm9 assembly with Bowtie 1 (Langmead et al., 2009) allowing one mismatch and multiple mapping to the genome for up to 500 sites to include repeat elements. Quantitation of reads that map to multiple positions in the

genome were adjusted as described in Ruby et al., 2006 by dividing the read numbers by the number of mapping sites. The miRBase v19 miRNA annotation was used to classify and quantify miRNAs. Since we were interested specifically in the concentration of active miRNA molecules loaded into Ago complexes, we required all considered miRNAs to additionally be enriched over background in FHAgo2 immunoprecipitated small RNA-seq samples as previously reported (Zamudio et al., 2014). MSC small RNA-seq data was taken from Gurtan et al., 2012. The average miRNA expression value of two biological replicates was used for ESC and MSC. For all analyses, miRNAs were grouped into families based on shared 7mer seed sequence. Expression values from each member of a miRNA family were summed. Below is a table of the miRNAs comprising each of the active ESC and MSC miRNA families detailed in the text. All individual miRNAs expressed above 10 copies per cell are listed, in order of decreasing expression.

miRNA family	seed match	miRNAs
ESC		
miR-294	AGCACTT	mmu-miR-294-3p, mmu-miR-291a-3p, mmu-miR-295-3p, mmu-miR-302a-3p,mmu-miR-302b-3p, mmu-miR-302d-3p
miR-292-5p	GTTTGAG	mmu-miR-292-5p, mmu-miR-290-5p, mmu-miR-293-5p
miR-292/467	GGCACTT	mmu-miR-467a-5p, mmu-miR-292-3p, mmu-miR-290-3p
miR-293	GCGGCAC	mmu-miR-293-3p
miR-15/16	TGCTGCT	mmu-miR-16-5p, mmu-miR-15b-5p, mmu-miR-15a-5p, mmu-miR-497-5p, mmu-miR-195a-5p
miR-92/25	GTGCAAT	mmu-miR-25-3p, mmu-miR-32-5p, mmu-miR-92a-3p, mmu-miR-92b-3p, mmu-miR-363-3p
miR-26	TACTTGA	mmu-miR-26a-5p, mmu-miR-26b-5p
miR-291-5p	CTTTGAT	mmu-miR-291a-5p
MSC		

let-7	CTACCTC	mmu-let-7c-5p, mmu-let-7f-5p, mmu-let-7b-5p, mmu-let-7i-5p, mmu-let-7a-5p, mmu-let-7e-5p, mmu-let-7d-5p, mmu-let-7g-5p, mmu-miR-98-5p
miR-125-5p	CTCAGGG	mmu-miR-125b-5p, mmu-miR-125a-5p, mmu-miR-351-5p
miR-26	TACTTGA	mmu-miR-26a-5p, mmu-miR-26b-5p
miR-15/16	TGCTGCT	mmu-miR-16-5p, mmu-miR-322-5p, mmu-miR-15b-5p, mmu-miR-497-5p, mmu-miR-15a-5p, mmu-miR-195a-5p
miR-21	ATAAGCT	mmu-miR-21a-5p
miR-99	TACGGGT	mmu-miR-99b-5p, mmu-miR-99a-5p, mmu-miR-100-5p
miR-31	TCTTGCC	mmu-miR-31-5p
miR-199	ACTACTG	mmu-miR-199a-3p, mmu-miR-199b-3p
miR-92/25	GTGCAAT	mmu-miR-32-5p, mmu-miR-25-3p, mmu-miR-92a-3p, mmu-miR-92b-3p
miR-27	ACTGTGA	mmu-miR-27b-3p, mmu-miR-27a-3p
miR-30	TGTTTAC	mmu-miR-30c-5p, mmu-miR-30a-5p, mmu-miR-30d-5p, mmu-miR-30b-5p, mmu-miR-30e-5p
miR-17/20/93	GCACTTT	mmu-miR-93-5p, mmu-miR-20a-5p, mmu-miR-17-5p, mmu-miR-106b-5p

For copies per cell estimates, we quantified RNA signal from Northern blots comparing a titration of a synthetic RNA standard (IDT) of known concentration to the signal from total RNA corresponding to known absolute cellular number. Copies per cell values for miR-295 in ESC and miR-16 in MSC were obtained in this manner, as described in Figure S1. These values were then used to normalize all miRNA read values from the small RNA sequencing.

RNA sequencing and quantitation

For ES, strand-specific mRNA sequencing libraries were prepared using either the UTP (Parkhomchuk and Borodina, 2009) method or TruSeq sample preparation kit from Illumina. Muliplexing barcode sequences were incorporated during the PCR amplification for UTP prepared samples. For MSC, RNA was extracted with Qiazol (Qiagen), and the Illumina TruSeq mRNA kit was used for sequence preparation. For all MSC and ESC libraries, paired-end sequence reads were generated with the Illumina

HiSeq 2000 system and separated based on library barcodes. RNA reads were mapped with Tophat v2.0.9 (Trapnell et al., 2009) to the Ensembl NCBIM37 build gene annotation. Isoform estimates from Cufflinks v2.1.1 (Trapnell et al., 2010) were used for Ensembl annotated isoforms. The average expression value of two biological replicates was used. To estimate copies per cells, a separate RNA-seq biological replicate was performed for ES and MSCs with spike-in standard RNAs. Cells were counted before extracting total RNA. 1.5ng External RNA Controls Consortium ERCC (Life Technologies) mix 1 RNA spike-ins of known concentrations were added to the cellular RNA post oligo dT purification step, corresponding to ~1-5% of total poly A RNA input into Illumina sequencing prep. Before mapping with Tophat, reads for spike-in experiments were trimmed to 40 nt length. FPKM expression values were converted based on regression fit for the known concentration of standards, and this was related to cell counts by total RNA input into library preparations.

iCLIP library preparation

TT-FHAgo2 and TT-Ago2 ESCs were induced to approximately WT levels of Ago2 expression with 2.5 ug/ml doxycycline (Sigma) 24 hours before harvesting. TT-FHAgo2 and TT-Ago2 MSCs were induced with 0.1 ug/ml doxycycline for 24 hours, then 1 ug/ml doxycycline for an additional 24 hours. iCLIP was performed with a tandem Flag-HA immunoprecipitation (IP) as described in (Jangi et al., 2014) and (Gurtan et al., 2013), except that ESC lysates were treated with a lower 1:200 dilution of RNase I (Ambion, AM2295), rather than 1:1000. ESC cloned iCLIP cDNA libraries were PCR amplified for 28 cycles, MSC libraries were PCR amplified for 23 cycles. For ESCs, a second replicate was performed with ON-bead 3' linker ligation (Konig et al.,

2011) instead of the OFF-bead 3' linker ligation described in Jangi et al., 2014. The two replicates were highly overlapping (>55% of genome-wide significant clusters overlap) with identical trends in miRNA site coverage. The OFF-bead ligated dataset with higher depth (2.5x more clusters called) was used exclusively for all presented analyses. For MSCs, two biological replicates with identical protocols were performed. Again, the two replicates highly overlapped (70% of genome-wide significant clusters overlap), but highest depth replicate 1 (3x more called read clusters than replicate 2) was exclusively used for all presented analyses, to be consistent with ESC analysis.

Summary of sequencing datasets

Sample name	Type	Total reads	GEO Accessions
TT-FHAgo2 (Tagged ES)	iCLIP	34,553,896	
TT-Ago2 (UnTagged ES)	iCLIP	22,049,552	
WT_Ago2CLIP_Tagged (MSC)	iCLIP	29,182,433	GSE44163/GSM1116239
WT_Ago2CLIP_Untagged (MSC)	iCLIP	23,416,533	GSE44163/GSM1116240
TT-FHAgo2_+dox48h_rep1	RNAseq	49,065,986	
TT-FHAgo2_+dox48h_rep2	RNAseq	53,621,134	
TT-FHAgo2_+dox48h_spikeIn	RNAseq	33,118,743	
MSC_rep1	RNAseq	16,588,780	GSE61031
MSC_rep2	RNAseq	17,980,365	GSE61031
MSC_spikeIn	RNAseq	28,002,946	
Dox2.5_1_130611	smallRNAseq	29,400,725	GSE50595/GSM1224440
Dox2.5_2_130611	smallRNAseq	24,496,606	
MSC_smallRNA_r1	smallRNAseq	17,791,160	GSE36978/GSM907767
MSC_smallRNA_r2	smallRNAseq	21,559,229	GSE36978/GSM907775

iCLIP mapping and clustering

Sequenced reads were trimmed for adapter sequence and mapped to the UCSC mouse genome build mm9 using Bowtie1 (Langmead et al., 2009). Reads mapping to >1 location or with >2 mismatches were discarded. To control for PCR bias in ESC iCLIP, we collapsed reads based on identical 5' and 3' end mapped coordinates. For MSC iCLIP, the RT primer contained 4 random barcode nucleotides that correspond to

positions 2-5 of the sequenced reads. Mapped reads were therefore collapsed based on identical 5' and 3' end positions and matching random barcode nucleotides. These uniquely mapping collapsed reads were used for all iCLIP analyses presented here. To include reads mapping across exon-exon junctions, we extracted splice junction coordinates from the isoforms produced by Tophat from RNA-seq data and used custom python scripts to generate a fasta file of the sequences +/- 45 nt of these splice junctions. We then mapped any previously unmapped iCLIP reads to these exon-exon junction sequences and added any splice junction mapping reads to the total mapped dataset. The number of exon-exon junction mapping reads for each iCLIP dataset using this method were: TT-FHAgo2 27,122; TT-Ago2 10,468; WT_Ago2CLIP_Tagged(MSC) 65,272; WT_Ago2CLIP_Untagged(MSC) 11,657.

After initially calling putative Ago2 bound regions by identifying clusters of more overlapping reads than a null poisson distribution model with cut-off p value < 0.001 (**Fig. S2A, "Poisson"**), we filtered out any clusters that had the same mismatched nucleotide in >70% of its reads (**Fig. S2A, "Mismatch Filter"**). This filter likely removed abundant small RNAs that persist through the iCLIP protocol and are mismapped due to the allowance of 2 genomic mismatches. Two thirds of the poisson clusters from Untagged (TT-Ago2) cells do not pass this filter. Finally, the remaining clusters were required to have significantly more tagged reads than untagged reads (binomial, $p<0.05$) after median-based normalization of the read counts in both libraries, using background regions found in both tagged and untagged datasets (similar to in (Zamudio et al., 2014) (**Fig. S2A, "Enriched over Untagged"**). Annotation of genomic regions was taken from Ensembl NCBIM37 build annotation.

Meta-site, gene expression, and iCLIP coverage analyses

All bioinformatics analyses were performed with custom python and R scripts. miRNA binding sites for meta-site analysis were identified by seed match searches across Ensembl-annotated 3'UTR isoform sequences. Conservation scores per miRNA binding site were calculated as average placental mammal PhastCons score (Siepel et al., 2005) across the seed match nucleotides. TargetScan miRNA family targets were obtained from TargetScan Mouse version 6.1 (Garcia et al., 2011). An expression cutoff of 0.1 FPKM was designated for all isoforms or genes considered in the analyses.

For all miRNA target site coverage analyses, the number of iCLIP reads overlapping the seed match +/- 10 nt flanks were quantified. "iCLIP coverage" was calculated as the number of iCLIP reads at a site (or in a gene) normalized to the WT expression of the isoform containing the site (or of the gene). We normalized each value to expression based on the slope and intercept of the linear regression best fit line of isoform WT FPKM versus number of iCLIP reads in the 3'UTR for all expressed 3'UTRs.

Control sites used for background estimation were randomly distributed across 3'UTRs expressed above 0.1 FPKM in the corresponding dataset. For analyses in Figures 2B and S3E , these control sites were divided into 1000 control seed "families" of random sites, with same number of sites per family as the average ESC miRNA family. Control gene sets, such as in Figures S3 and S4C, were always random, expressed genes that have a matching distribution of 3'UTR characteristics including length, GC content, and WT gene expression compared to genes containing bound 6/7/8mer target sites of the active ESC miRNAs. Control sites used as background estimation in iCLIP coverage

analyses, such as in Figures 3A-C, S3, and S4B, were the set of random control sites within these matched control genes. Control sites in Figure S5A were random sites within expressed (>0.1 FPKM) isoforms of each corresponding genic category.

Target Pool cpc estimations

Every 6/7/8mer target site was associated with a copies per cell (cpc) value based on the gene isoform that contains it. We then used iCLIP coverage to determine which sites are likely accessible and therefore counted as part of the target pool as follows: First, we summed the cpc of every site with at least one iCLIP read. Second, since sites with 0 iCLIP reads could be either truly inaccessible or not covered by iCLIP due to low coverage, we determined the number of 0-read sites to include in the total target pool based on iCLIP sampling depth. For each miRNA and site type pair, we calculated the total number of sites (S) in expressed 3'UTRs and the total number of reads across those sites (R). We then ran 100 simulations of shuffling R reads across S bins to estimate the number of bins (sites) expected to have 0 reads by chance, given the depth of read coverage for that miRNA site type pair. Histograms of random read distribution and observed read distribution across miR-294 and miR-15/16 8mers are shown in Figure S4C. The number (N) of 0 bound sites expected by chance at the given depth for each miRNA site type pair were included in the corresponding total target pool by adding [N * average expression of 0 bound sites for that miRNA site type pair] to the target abundance value. When calculating the final target abundance values for each miRNA we adjusted the expression value of each considered target site from a non-3'UTR region by the average iCLIP coverage per site across the corresponding region relative to 3'UTR (**Figure S5A**), to reflect their probability of Ago-miRNA interaction.

Reporter plasmid construction

Bi-directional pTRE-Tight-BI (Clontech) eYFP and mCherry reporter and rtTA constructs were as described in (Mukherji et al., 2011). mCherry 3'UTR inserts were designed based on an endogenous miR-293 8mer site with a log2 iCLIP coverage value of 1.96 located in the 3'UTR of the Sirt7 gene. The 8mer sequence +/- 20 nt of flanking sequence from the Sirt7 3'UTR was placed in a tandem array of 3 repeats, such that 3 8mer sites were separated by 40 nt spacers. The miR-293 seed match site (GCGGCACA) was then mutated to correspond to different miRNAs: miR-294 AGCACTTA; miR-92/25 GTGCAATA; let-7 CTACCTCA. The 144 bp 3'UTR insert sequences were then synthesized by IDT, along with the complementary strand, with engineered restriction enzyme overhangs for HindIII and SalI. Complementary DNA oligos were annealed and cloned into the 3'UTR of mCherry. All constructs were sequence-confirmed.

Flow cytometry

AB2.2 WT ESCs or clone 12 WT MSCs were plated in 12-well plates 24 hours before transfection with 200 ng reporter plasmid, 200 ng rtTA plasmid, and 1200 ng pWhitescript plasmid as carrier DNA. Lipofectamine2000 (Invitogen) was used for transfections per manufacturer protocol. After 4 hours in transfection mix, cells were switched to media with 1ug/ml doxycycline (Sigma). Assays were performed 24 hours post-transfection.

FACs measurements were taken with FACS LSR II HTS (BD Biosciences) and data processed with Flowjo software to yield eYFP and mCherry values for each cell. These values were normalized for background fluorescence by subtracting the mean plus two standard deviation of signal from untransfected control and binned by eYFP

expression levels as previously described (Mukherji et al., 2011). For absolute quantitation of mCherry transcripts, transfections were done in 6-well plates and sorted on FACSAria (BD Biosciences) instrument to isolate nearly 10e5 cells per gate. Cells were washed twice in 5ml of PBS and RNA isolated with RNeasy columns (Qiagen) with on-column DNase treatment. The cellular levels were normalized for total RNA concentration per cell and compared to a mCherry standard generated by spiking-in an *in vitro* transcribed mCherry transcript. The mCherry transcript was produced with the MAXIscript T7 kit (Life Technologies), purified on a denaturing gel and quantified. The *in vitro* transcribed mCherry transcripts were then added at various cpc levels to mock transfected RNA for cDNA synthesis with random hexamer oligos in the SuperscriptIII First strand synthesis system (Life Technologies). The mCherry transcripts in all samples were quantified with PowerSyber Green PCR mix on 7500 Real-Time PCR system (Applied Biosystems) using mCherry primers described in (Mukherji et al., 2011).

Mathematical model and simulations

Our mathematical model for binding is based on basic biochemical principles of equilibrium thermodynamics described by Michaelis-Menten kinetics. The fraction bound θ_i of a given target pool i is calculated by the hyperbolic equation:

Eq.1 $$\theta_i = [\text{Ago-miRNA}]_{free} / (K_i + [\text{Ago-miRNA}]_{free})$$

where K_i is the dissociation constant (K_D) for the Ago-miRNA:target interaction of the given target pool affinity group. 6mer and 8mer K_D values were calculated relative to *in vitro* 7mer K_D values using relative average iCLIP coverage across all expressed 3'UTR sites of each type (**Figure S3A, right**). Copies per cell concentrations were converted into molar amounts for simulations based on cellular volumes of (ESC=500um^3 ,MSC=

2000um^3) and cytoplasmic proportion (ESC=0.40 ,MSC=0.75) (Milo et al., 2010). As miRNAs require Ago for stability (Zamudio et al., 2014), we assumed all quantified miRNA molecules are in complex with Ago (i.e. [Ago-miRNA] = [miRNA]).

All target pools compete for the same pool of total miRNA, thus free miRNA is related to total miRNA by the following equation:

Eq.2 $$[\text{miRNA}]_{total} = [\text{miRNA}]_{free} + [8\text{mer}]_{total} * \theta_{8mer} + [7\text{mer}]_{total} * \theta_{7mer} + [6\text{mer}]_{total} * \theta_{6mer}$$

For the miRNA titration curves in Figure S6A comparing total miRNA concentration to fraction bound targets, Eq.2 was solved numerically for simulated $[\text{miRNA}]_{free}$ values across the range 10^{-4} to 10^{6}, and fraction bound of each site type target pool was calculated by inserting $[\text{miRNA}]_{free}$ into Eq. 1. Total target pool concentrations for each affinity group were the endogenous values we measured using all iCLIP-estimated sites from all genomic regions as described in the text (**Table 1**).

For the target titration curves in Figure 5C and Figures S6D and S6E, Eq. 2 was rearranged to:

Eq.3 $$[8\text{mer}]_{total} = ([\text{miRNA}]_{total} - [\text{miRNA}]_{free} - [7\text{mer}]_{total} * \theta_{7mer} - [6\text{mer}]_{total} * \theta_{6mer}) / \theta_{8mer}$$

Eq. 3 was solved numerically for simulated $[\text{miRNA}]_{free}$ values across the range 10^{-9} to 10^{8}, inputting measured endogenous concentrations for $[\text{miRNA}]_{total}$, $[7\text{mer}]_{total}$, and $[6\text{mer}]_{total}$ and K_D values as described above. Corresponding fraction bound of each site type target pool was then calculated by inserting $[\text{miRNA}]_{free}$ into Eq. 1.

Supplemental References

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat Struct Mol Biol *18*, 1139–1146.

Gurtan, A.M., Lu, V., Bhutkar, A., and Sharp, P.A. (2012). In vivo structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. Rna *18*, 1116–1122.

Gurtan, A.M., Ravi, A., Rahl, P.B., Bosson, A.D., JnBaptiste, C.K., Bhutkar, A., Whittaker, C.A., Young, R.A., and Sharp, P.A. (2013). Let-7 represses Nr6a1 and a mid-gestation developmental program in adult fibroblasts. Genes Dev *27*, 941–954.

Jangi, M., Boutz, P.L., Paul, P., and Sharp, P.A. (2014). Rbfox2 controls autoregulation in RNA-binding protein networks. Genes Dev.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2011). iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. J Vis Exp.

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.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal.

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers--the database of key numbers in molecular and cell biology. Nucleic Acids Res *38*, D750–D753.

Mukherji, S., Ebert, M.S., Zheng, G.X.Y., Tsang, J.S., Sharp, P.A., and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. Nat Genet *43*, 854–859.

Parkhomchuk, D., and Borodina, T. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids ….

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. elegans. Cell *127*, 1193–1207.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034–1050.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511–515.

Zamudio, J.R., Kelly, T.J., and Sharp, P.A. (2014). Argonaute-bound small RNAs from promoter-proximal RNA polymerase II. Cell *156*, 920–934.