



Published in final edited form as:

Mol Cell. 2014 November 6; 56(3): 347–359. doi:10.1016/j.molcel.2014.09.018.

Endogenous miRNA and Target Concentrations Determine Susceptibility to Potential ceRNA Competition

Andrew D. Bosson^{1,2,3}, Jesse R. Zamudio^{1,3}, and Phillip A. Sharp^{1,2,*}

¹David H. Koch Institute for Integrative Cancer Research, Cambridge, MA 02139, USA

²Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

SUMMARY

Target competition (ceRNA crosstalk) within miRNA-regulated gene networks has been proposed to influence biological systems. To assess target competition, we characterize and quantitate miRNA networks in two cell types. Argonaute iCLIP reveals that hierarchical binding of high- to low-affinity miRNA targets is a key characteristic of in vivo activity. Quantification of cellular miRNA and mRNA/ncRNA target pool levels indicates that miRNA:target pool ratios and an affinity partitioned target pool accurately predict in vivo Ago binding profiles and miRNA susceptibility to target competition. Using single-cell reporters, we directly test predictions and estimate that ~3,000 additional high-affinity target sites can affect active miRNA families with low endogenous miRNA:target ratios, such as miR-92/25. In contrast, the highly expressed miR-294 and let-7 families are not susceptible to increases of nearly 10,000 sites. These results show differential susceptibility based on endogenous miRNA:target pool ratios and provide a physiological context for ceRNA competition in vivo.

INTRODUCTION

MicroRNAs (miRNAs) are ~21–23 nt RNAs that guide Argonaute (Ago) proteins to targets for posttranscriptional gene repression (Bartel, 2004; Meister, 2013). Ago-miRNA target recognition is predominantly based on sequence complementarity to positions 2–7 of the miRNA, termed the seed (Bartel, 2009). A miRNA's repressive activity varies according to target site accessibility and binding affinity. The binding affinity is mainly related to the extent of miRNA seed base pairing, from the higher abundant but lower affinity 6-mer seed match to less abundant and increasingly higher affinity 7-mer and 8-mer sites (8-mer = 7-

*Correspondence: sharppa@mit.edu.

³Co-first author

AUTHOR CONTRIBUTIONS

A.D.B. performed the iCLIP experiments, analyses of target pool quantitation, and coverage with input from J.R.Z. and created the mathematical binding model. J.R.Z. analyzed the small RNA-seq and RNA-seq data and the FACS experiments with input from A.D.B. A.D.B., J.R.Z., and P.A.S. designed the study and prepared the manuscript.

ACCESSION NUMBERS

The National Center for Biotechnology Information Gene Expression Omnibus accession number for ESC iCLIP and RNA sequencing data described in this paper is GSE61348.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and seven figures and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2014.09.018>.

mer match with adenosine across from position 1 of the miRNA) (Friedman et al., 2009; Nielsen et al., 2007).

In addition to target site affinity, Ago binding is biochemically linked to the relative cellular concentration of the miRNA and the total target pool (miRNA:target ratio) (Arvey et al., 2010; Brown et al., 2007; Doench and Sharp, 2004; Garcia et al., 2011; Mullokandov et al., 2012). Previous analysis of miRNA repression levels showed miRNAs with low miRNA:target ratios confer minimal repression (Arvey et al., 2010; Garcia et al., 2011). Absolute levels of miRNAs for various cell types are estimated in the range of tens to 120,000 copies per cell (cpc) (Bissels et al., 2009; Calabrese et al., 2007; Denzler et al., 2014; Lim et al., 2003; Mukherji et al., 2011). Estimating total target concentrations has proven more difficult. Target genes can be computationally predicted based on miRNA seed match conservation and other factors affecting site accessibility (Bartel, 2009), but biochemical target identification by Ago immunoprecipitation (IP) suggests even more extensive miRNA targeting (Chi et al., 2009; Hafner et al., 2010; Leung et al., 2011; Loeb et al., 2012; Zisoulis et al., 2010). These in vivo Ago-binding studies provided thousands of statistically significant target sites, but have not been used to quantify the total target pools of individual miRNAs.

miRNA repression can be inhibited by perturbing endogenous miRNA:target ratios through overexpression of RNAs with multiple high-affinity miRNA binding sites, termed miRNA “sponges” (Ebert et al., 2007). Interestingly, miRNA target competition by sponges occurs in a threshold-like manner (Mukherji et al., 2011) similar to other biological systems of molecular titration (Brewster et al., 2014; Buchler and Louis, 2008; Levine and Hwa, 2008). The salient property of titration-mediated regimes in biology is nonlinear input-output responses that occur near the buffering molecule concentration. In the case of miRNAs, as the target pool surpasses the threshold set by the buffering miRNA concentration plus K_D of the miRNA:target interaction, smaller changes in targets can result in larger changes in the concentration of free, i.e., unrepressed target (Mukherji et al., 2011). This relationship makes target affinity and cellular abundance estimates important to determining responses to miRNA or target pool perturbations.

Based on target competition, the competing endogenous RNA (ceRNA) hypothesis proposes a layer of gene regulation mediated by transcripts with shared miRNA binding sites (Salmena et al., 2011; Tay et al., 2014). In this hypothesis, ceRNAs (RNAs targeted by the same miRNA) exhibit indirect positively correlated expression. For example, as one ceRNA increases, it titrates away miRNA from repressing other ceRNAs, and increases expression of all ceRNAs in the network (Sumazin et al., 2011; Tay et al., 2011). Multiple types of endogenous ceRNAs have been reported, including protein-coding transcripts, long noncoding RNAs (lncRNAs), pseudogenes, and circular RNAs (circRNAs) (reviewed in Tay et al., 2014). Many ceRNA interactions have been linked to development or various disease states (Cesana et al., 2011; de Giorgio et al., 2013; Kallen et al., 2013; Karreth and Pandolfi, 2013; Ling et al., 2013). However, recent quantitation of ceRNA effects for miR-122 in liver suggested no competition would occur in vivo because of the large relative abundance of miRNA target pools (Denzler et al., 2014). Thus, despite the preponderance of examples attributing various phenotypes to individual ceRNA interactions, the hypothesis remains

controversial. Mathematical models predict ceRNA effects within an optimum regimen where the miRNA and targets are near equimolar concentrations (Ala et al., 2013; Bosia et al., 2013; Figliuzzi et al., 2013), but it remains unclear if endogenous miRNA networks are in these susceptible regimes.

In this report, we use Ago2 individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) (König et al., 2010) and absolute quantification of miRNAs and mRNAs to quantitatively assess miRNA-regulation genome-wide. We find that miRNA:target ratios determine both the accumulation of Ago across its different affinity sites and the susceptibility of target repression to target competition. This quantification reveals endogenous miRNA families susceptible to competition, which we then validate with single-cell miRNA reporter assays.

RESULTS

Quantifying miRNA Regulation with mRNA-Seq, Small RNA-Seq, and iCLIP

Ago-miRNA complexes bind targets according to the target site's equilibrium dissociation constant (K_D) and the concentrations of both free miRNA and target (Figure 1A, top). Each miRNA has hundreds of targets (potential ceRNAs) with different affinities and expression levels that essentially act as competitive inhibitors that reduce free miRNA (Figure 1A, bottom). Due to the buffering properties of molecular titration, these competition effects depend on both the interaction affinity and miRNA:target ratio (Buchler and Louis, 2008; Mukherji et al., 2011).

We assessed miRNA regulation using four genome-wide measurements to quantify miRNAs, characterize Ago-RNA interactions, estimate miRNA target pools, and measure repression. Specifically, we first performed small RNA-seq, poly-A RNA-seq, and iCLIP within a single mouse embryonic stem cell (ESC) line. To assure high-confidence Ago-RNA interactions, these were performed in cells expressing a single doxycycline (dox)-inducible *Ago2* gene in an endogenous Ago1-4 null background (either TT-FHAgO2 [FLAG-HA Ago2 epitope tagged] or TT-Ago2 [Ago2 untagged]) (Zamudio et al., 2014) (Figure 1B). To calculate miRNA cpc, small RNA-seq miRNA counts were normalized to miR-295 cpc quantified by northern blot. miRNAs with shared 7-mer seed sequences were summed into miRNA families because their targets will be highly overlapping (Figure S1A available online; Table 1). To measure global target RNA concentrations, we performed poly-A RNA-seq with synthetic spike-in RNAs and used transcript isoform levels estimated with Cufflinks (Trapnell et al., 2010) because isoforms may differ in miRNA seed match content. The total protein-coding mRNA content for ESCs was calculated at 158,000 cpc (Figure S1B). To measure miRNA-mediated gene repression, we performed mRNA-seq from TT-FHAgO2 with (Ago+) or without (Ago-) Ago proteins and observed upregulation of predicted miRNA targets upon Ago loss (Figure S1C).

With these absolute measurements of cellular RNA and miRNA-mediated repression, we next turned to directly measuring an individual miRNA's targeting activity by quantitating actual binding events at each target site. FHAgO2 iCLIP provided high confidence genome-wide characterization of Ago binding events. After UV-crosslinking, FHAgO2 was purified

by tandem FLAG-HA IP, and associated RNAs were sequenced and mapped to the mouse genome. All IP experiments were performed in parallel in the untagged TT-Ago2 cell lines for a negative control (Figure S2). iCLIP reads and crosslink sites show highly specific Ago interactions, demonstrated for the three prime untranslated region (3' UTR) of p21 and Casp2 genes in Figure 1C. We developed a computational pipeline to identify statistically significant clusters of Ago2 iCLIP reads above untagged background (Figure S2A; Experimental Procedures). The resulting 6,817 regions constitute a set of high-confident Ago2 bound sites in ESCs. Contrary to previous reports of abundant intronic binding (Chi et al., 2009; Hafner et al., 2010; Leung et al., 2011), our stringently identified clusters were depleted in introns relative to genomic background but still enriched in long noncoding RNA and pseudogene categories, as well as the expected exonic categories (Figure S2B). The 5' UTR, 3' UTR, and CDS mapping clusters represent 85, 1,811, and 675 different genes, respectively. The nucleotide resolution of RNA-protein crosslink sites (König et al., 2010; Sugimoto et al., 2012) is seen by the strong peak of crosslinking signal immediately upstream of miRNA seed match sites for the highly expressed miR-294 family (Figure 1D). In fact, cross-linking patterns are distinguishable for various site types by the presence of additional RNA-protein interactions at the first position of miR-294 8-mer seed matches confirming different target interactions (Figure S2C). Finally, we find increased iCLIP coverage (iCLIP reads per million [RPM]/FPKM of gene isoform) is correlated with increased Ago repression (Figure 1E). The quantitative, specific, and stringent nature of the Ago2 iCLIP data suggests iCLIP reads can be used to characterize Ago2-miRNA target binding events.

Ago-miRNA-Bound Sites Are Correlated with Potential Target Pool Size and Distinguish Active miRNA Seed Families

To extend our analysis to include even weak affinity sites, we quantified the number of Ago2 iCLIP reads at all miRNA seed matches within expressed 3' UTRs. The inclusion of sites with even one iCLIP read was supported by meta-analysis that showed a crosslink peak above background immediately upstream of the seed match at one-read sites (Figure S2E). With these sites quantified, we found that the number of bound 7/8-mer sites for a given miRNA family more resembled its total number of potential target sites rather than miRNA expression level (Figure 2A). This suggests that Ago-miRNA binding spreads across the entire accessible target pool even at low miRNA concentrations, possibly accumulating at 7/8-mer target sites and spreading to lower-affinity sites at higher miRNA concentrations.

We next identified miRNA seed families that exhibit significant accumulation across their target pool by comparing average iCLIP coverage per site. The iCLIP coverage for the top 30 expressed ESC miRNA families was compared to a control distribution of randomly selected 3' UTR sites (random seeds). To predict the maximal number of active ESC miRNAs, we only used high-affinity 8-mer sites in this analysis. We found 11 ESC miRNA families exhibit average iCLIP coverage above background ($p < 0.01$) (Figure 2B). Plotting the crosslink density profile around 7/8-mer target sites for the top 30 expressed ESC miRNAs confirmed appreciable Ago binding only at the identified active miRNA seeds (Figure 2C). For subsequent analysis, we excluded three lowly expressed active seed families, miR-17/93 (GCACTTT), miR-19 (TTTGAC), and miR-148 (TGCACTG) that

shared shifted or mismatched seeds with the highly expressed miR-294/292 families ([A/G]GCACTT) to alleviate any crossover signal. This crossover signal can be seen by the -1 and +2 max crosslink positions of miR-17/93 and miR-19 seed matches, respectively, corresponding exactly to the -1 and +2 shifts in seed sequence relative to miR-294 (Figure 2C, bottom three rows). Finally, to confirm that significant crosslinking signal translates to repressive activity, we compared gene repression of conserved 7/8-mer target sites to average iCLIP coverage and found a strong correlation (Figure 2D). This result validates use of iCLIP coverage to assess miRNA binding across the target pool and identifies multiple seed families conferring repression in stem cells that we use to characterize miRNA-regulation.

Target Pool Abundance, miRNA Concentration, and Individual Target Binding Affinity Determine Ago-miRNA Coverage In Vivo

To investigate how miRNAs interact with their target pool, we split target sites into mutually exclusive 8-mer, 7-mer, and 6-mer pools to represent high- to low-affinity targets. We found the average iCLIP coverage across active ESC miRNA target sites decreases with decreasing miRNA complementarity, supporting that iCLIP coverage reflects Ago binding affinity (Figure S3A). Further, plotting iCLIP coverage per 8-mer target site for each active ESC miRNA individually revealed varying degrees of increased coverage over background control sites (Figure 3A). Interestingly, iCLIP coverage per site for some miRNAs begins to collapse to near background levels at 7-mer (Figure 3B) and 6-mer target sites (Figure 3C). Across 6-mer targets, only three miRNA families—miR-292/467 and miR-294, which share a 6-mer seed, and miR-293—are strongly detected above background. This showed that target affinity groups, approximated by seed match type, are differentially bound by miRNA seed families. As expected, we found target repression also differs by seed type for individual miRNA seed families, with stronger affinity sites conferring more repression in general (Figure S3B). For this and any further analysis, we included 7-mer-A1 sites (6-mer match with adenosine across from position 1; Bartel, 2009) with the 6-mer class because we observed no difference in binding between these classes when analyzing all active ESC miRNAs in aggregate (Figure S3C).

We next developed an approach to estimate target pool concentrations for each affinity group to explore the relationship between concentration and binding. The inclusion of all predicted sites would greatly overestimate target numbers because 40% of predicted miR-294 8-mer target sites are not bound despite miR-294 levels being significantly higher than its 6/7/8-mer target pool (Figure S3E, see below). Alternatively, restriction to only conserved targets underestimates the target pool because the majority of iCLIP bound sites are nonconserved (Figure S4A), although conserved sites do exhibit ~30% higher iCLIP coverage on average than nonconserved sites (Figure S4B). Therefore, we used iCLIP to inform which target sites to include in the target pool in a two-step process. First, the isoform expression value for any target site with at least one iCLIP read was added to the target pool. We confirmed this target pool inclusion criteria by comparing the repression for mutually exclusive 8-mer targets with and without iCLIP reads (Figure S4C) and found no significant evidence for repression at 0 read sites compared to matched controls, indicating that our approach appropriately captured the majority of actively repressing miRNA-target interactions. Second, because we are

interested in defining the entire potential target pool that is accessible to miRNA, not just the pool that is being detectably repressed, we also include a portion of the sites with no iCLIP reads. The 0 read sites could be due to either true site inaccessibility or low sampling depth and therefore were included or not based on the depth of read coverage for each miRNA seed family and site type (Figure S4D; Supplemental Experimental Procedures).

This method provided target cpc estimates of 1,200 8-mers, 5,000 7-mers, and 22,000 6-mers on average for 3' UTR sites of each active miRNA ESC family (Figure 3D, lines). For reference to other cell types, these values correspond to 0.8%, 3.2%, and 14.1% of the total mRNA concentration in the cell. We found target pool levels can vary greatly, as exemplified by the ~10-fold lower cpc values for miR-293. The miR-293 low target abundance is likely because its seed match contains a CG dinucleotide, which is depleted within 3' UTR sequences. Six of eight active ESC miRNA seeds were expressed at concentrations greater than their 3' UTR 8-mer target pool. The minimum expression value of active ESC miRNA families is 580 cpc, given by miR-291-5p. However, this value is likely an underestimate of the miRNA concentration required for detectable activity in ESCs, because the significant crosslink signal of miR-291-5p may be partially explained by its 7-mer sequence similarity (CTTTGAT) to the higher expressed miR-292-5p (GTTTGAG). Examining the top 30 ESC seed families, only the top eight are above the concentration of their 3' UTR 8-mer target pool, with the exception of another CG dinucleotide miRNA, miR-127, which only has 78 possible 8-mer sites in expressed 3' UTRs (Figure S4E). The low miRNA:target ratios relative to even just the 8-mer target pool likely explain the undetected binding for the majority of the top 30 miRNAs. However some miRNAs, including miR-135 and miR-669-5p, exceed their 3' UTR target pools, but do not show significant binding signal. Highly expressed miRNAs with undetectable canonical activity have been reported and could be due to many factors, including subcellular localization or post-transcriptional modification of the miRNA (Hwang et al., 2007; Jones et al., 2009; Mullokandov et al., 2012).

We found that Ago binding signal is related to the miRNA: target ratio partitioned by affinity group. Only the three miRNAs (miR-294, miR-292, miR-293) expressed in excess of their 3' UTR 6-mer pool (Figure 3D) show strong iCLIP coverage across 6-mer target sites (Figure 3C). As confirmation, a crosslink peak is clearly detectable across all three site types of the highly expressed miR-294 family (Figure 3E, top), whereas miR-92/25 only shows strong crosslink signal across 7-mer and 8-mer sites, as predicted by its low 6-mer miRNA-Target ratio (Figures 3E, bottom, and 3D). These observations support hierarchical binding within target pools with the extent of lower affinity coverage dependent on miRNA:target ratios.

Total Endogenous miRNA Target Pools Are Predominately Composed of 3' UTR Sites and Are Larger Than Most miRNA Concentrations

Ago-miRNA complexes target a diverse set of transcript classes and exonic regions in addition to 3' UTR that should be included in target pool calculations (Figure 4A; Figure S2B). To estimate the complete cellular target pool for each miRNA, we compiled seed matches for the top 30 expressed ESC miRNAs across Ensembl annotated 3' UTR, CDS, 5'

UTR, lncRNA, and pseudogene classes. Total iCLIP coverage across 7/8-mer target sites in these categories is dominated (~75%) by 3' UTR sites (Figure 4B). The pseudogene and lncRNA classes make up only 1.7% of the average miRNA's 7/8-mer binding pool in ESCs. The lower proportion of iCLIP coverage across non-3' UTR classes is partially due to lower average iCLIP coverage per site in these regions relative to 3' UTR (Figure S5A) and may reflect differential accessibility to Ago-miRNA complexes. Accordingly, we adjusted the non-3' UTR contributions to total target pool number to reflect their relative Ago binding (Experimental Procedures). This resulted in CDS, 5' UTR, pseudogene, and lncRNA sites being added to the total target pool as 42%, 32%, 3%, and 5.4% of their measured expression value, respectively. After including all regions, the total target abundances for the top 30 expressed ES miRNAs ranged from approximately 17,000 to 140,000 cpc (excluding the two families with CG-containing seed matches) and are above the corresponding miRNA levels for all but the miR-294 and miR-292/467 families (Figure 4C; no weighting by class, Figure S5B bottom).

In Vivo Ago Binding Is Determined by Stoichiometric miRNA:Target Ratios

Having estimated the total target pools, we set to correlate the miRNA:target ratio for a given miRNA and site type with Ago binding. We defined Total Ago Occupancy, analogous to “fraction bound” in in vitro binding assays, as $\text{Target site iCLIP RPM} / \sum \text{target site FPKM}$ across the corresponding target sites. Plotting miRNA:target ratio versus Total Ago Occupancy for each active ESC miRNA family by site type revealed that Ago binding above background occurs as the miRNA:target pool ratio approaches 1 and increases hyperbolically with increasing miRNA:target ratio (Figure 5A). In addition, we found the maximum Total Ago Occupancy for 8-mers is higher than 6/7-mers supporting hierarchical binding within target pools and accumulation at high-affinity sites. To estimate when 50% of a given affinity target pool will be bound, we fit hyperbolic curves to the data points and found half-maximal miRNA:target ratios of 8.2, 5.3, and 3.2 for 8-mers, 7-mers, and 6-mers, respectively (Figure S5C). Although the half-maximal miRNA:target ratios for 7-mers and 6-mers are lower than for 8-mers, the values actually correspond to larger absolute miRNA levels required for 50% binding because the 7-mer and 6-mer target abundances are larger. In addition, because miRNA and target concentrations are far above the K_D of binding (Wee et al., 2012, see below), half-maximal miRNA:target ratios above 0.5 reflects endogenous competition between site types. For instance, miRNA levels must be 8.2 times the level of 8-mer targets in the cell to achieve 50% binding of 8-mers because the large 6-mer and 7-mer target pools compete for the same pool of available miRNA molecules. These half-maximal ratios may be slightly overestimated, however, because miR-292-5p skewed the fitted curve to the right with Ago occupancy values that are lower than expected based on its miRNA:target ratios. The lower occupancy of miR-292-5p sites is expected due to lower overall conservation of its target sites, indicating these targets are not as biologically active (Figure S5D). Still, the half-maximal miRNA:target ratios reported here provide estimates of miRNA expression sufficient for strong binding to different affinity groups.

Mathematical Model of Target Competition Predicts Response to Changes in Target Pool Concentration

Having confirmed the influence of both target site affinity and miRNA:target ratios on Ago-miRNA binding, we turned to evaluating how the relative miRNA:target ratio will affect ceRNA susceptibility. We reasoned that active miRNA families with lower miRNA:target ratio might be more sensitive to target competition. As a first test, we created a simple equilibrium binding model to incorporate intratarget pool site type competition (details in Experimental Procedures). In our model, each site type target pool competes for the same pool of free miRNA based on their concentration and K_D (Figure 5B), effectively reducing the fraction bound calculated for each individual target pool (Figure S6A). We input our measured endogenous miRNA and target pool cellular concentrations as a starting point for model simulations. A previously measured *in vitro* K_D of 26 pM was used for 7-mer sites (Wee et al., 2012), and the 8-mer and 6-mer K_D was 15.4 and 53.7 pM, respectively, calculated relative to 7-mer iCLIP coverage (Figure S3A, right). However, due to intratarget pool competition, the effective total miRNA concentrations for half maximal target binding are in the midnanomolar range (Figure S6A), nearer the cellular concentrations of miRNA and target mRNA molecules (Table 1). We found strong correlations between the predicted fraction bound of each active miRNA's target pools and both the corresponding iCLIP Total Ago Occupancy values and repression values (Figure S6B), indicating the reasonableness of the model as well as our *in vivo* estimates of miRNA:target ratios and K_D values.

Of note, to accurately predict target pools we limited our target pool inclusion to canonical 6/7/8-mer pools. As a result, our binding model predicts full saturation of miR-294 6/7/8-mers because it is expressed almost three times above its total 6/7/8-mer target pool. However, our iCLIP and repression values do not support equal saturation of miR-294 6/7/8-mers. For exceptional miRNAs like miR-294 that are expressed substantially above 6-mer seed matches, other noncanonical sites likely contribute meaningfully to the competing target pool. Such evidence of noncanonical binding sites has been reported elsewhere (Helwak et al., 2013; Lal et al., 2009; Loeb et al., 2012) and can be seen in our iCLIP data as crosslinking peaks across miR-294 mismatched seed sites (Figure S6C). Until all noncanonical binding classes and relative affinities are clearly defined, the exceptional miR-294 family, which is expressed highly enough to buffer against 8-mer target increases of ~1.5 orders of magnitude and likely not in the range of physiological perturbation, is not accurately modeled (Figure S6D).

For all other miRNA seed families, we used the model to predict Ago-miRNA binding changes in response to changes in high-affinity target concentrations. The fraction of free targets as a function of total 8-mer target pool levels is shown for a representative lowly expressed active miRNA family, miR-92/25 in Figure 5C. The simulation was performed across over two orders of magnitude above and below the endogenous concentration (vertical dotted black line) and showed the expected nonlinear titration response to changes in 8-mer total target pool levels. To assess physiological perturbations in ESCs, we indicate 10- and 100-fold induction of an average expressed target gene containing 3×8 -mer sites (390 or 3,900 cpc gained) or complete knockdown of a very highly expressed target with 3×8 -mer sites (600 cpc lost) (Figure 5C, vertical solid gray-black lines). The model predicts 8-

mer targets of the lower expressed miR-92/25 family are endogenously in a sensitive region. All active miRNAs other than the highly expressed miR-294/292 families exhibit qualitatively similar susceptibility to target pool changes as for miR-92/25 (data not shown). Responses like these that are more susceptible to physiological perturbation would likely occur with lower overall changes in expression due to lower overall fraction bound (Figure 5C, fraction 8-mers bound ~20%). These comparisons highlight predicted differences in ceRNA responses and differential susceptibility based on miRNA:target ratios.

Single-Cell Measurement of miRNA Activity upon Target Induction Demonstrates Differential Response to Target Pool Increase

To test our model's prediction of differential target competition susceptibility based on miRNA:target ratio, we used a single-cell reporter system to simultaneously estimate target overexpression and the corresponding effect on miRNA activity in vivo. A bidirectional promoter was used to transcribe identical levels of nontargeted control eYFP mRNA and targeted mCherry mRNA, which contains three 8-mer 3' UTR miRNA binding sites to a selected miRNA. An untargeted mCherry mRNA was used as a control to measure mCherry/eYFP levels in the absence of miRNA regulation. We measured fluorescence per cell with flow cytometry and used eYFP levels to assess overexpression. By binning eYFP expression into 100 bins, we calculated the miRNA-mediated fold repression ($mCherry_{\text{untargeted}}/mCherry_{3 \times \text{sites}}$) across a range of inductions (Figure 6A). We selected three active miRNA families that represent high (miR-294), intermediate-low (miR-293), and low (miR-92/25) miRNA:target pool ratios for this experiment (Table 1). By calculating the average fold repression of the first ten bins, we estimated the endogenous repression levels as 10-fold, 6.2-fold, and 3.7-fold repression for miR-294, miR-293, and miR-92/25, respectively (Figure 6A, bottom, "Low"), consistent with the different miRNA:target pool ratios. Importantly, when comparing the average fold repression in the middle 10 eYFP expression bins ("Mid") to endogenous repression levels (first ten bins, "Low"), we see that repression of the two lower miRNA:target ratio reporters, miR-293 and miR-92/25, is reduced from 6.2- to 4.4-fold and from 3.7- to 2.4-fold, respectively, but the high miRNA:target ratio reporter, miR-294, is unaffected (Figure 6A, bottom). Even after "High" reporter induction (last ten bins) the miR-294 reporter is not derepressed, whereas average fold repression of the miR-293 and miR-92/25 reporters drops to 2.3- and 1.5-fold, respectively (Figure 6A). These results confirm the potential ceRNA susceptibility of low miRNA:target pool ratios and the insensitivity of high miRNA:target ratios. To estimate the overexpression achieved with the reporters, nontargeted mCherry transcripts were quantified after cell sorting in the ranges of induction corresponding to "Mid" and "High" overexpression of reporter (Figure 6A) and gave absolute values of 1,000 and 3,600 cpc, respectively (Figure S7A). The incorporation of three miRNA binding sites per reporter mRNA puts our target pool induction estimates for detectable miR-92/25 target derepression at 3,000 additional high-affinity 8-mer cpc in these assays. This provides a direct quantitative test of in vivo competition for endogenous miRNA target pools with reasonable estimates of the perturbation levels required for titration effects.

To show our approach can be extended to other systems, we performed the identical analyses in mesenchymal stem cells (MSCs) using Ago2-iCLIP and miRNA and target pool

measurements (Figures S7B–S7E). MSCs had similar Ago binding characteristics in regard to iCLIP coverage per site type and target pool composition. Twelve miRNA families were identified as active above background. Interestingly, we again observed miRNA families predicted to exist in two classes of ceRNA susceptibility. We found let-7 endogenous miRNA-Target ratio is distinctly higher than most other active miRNAs (Table 1), similar to the miR-294 family in ESCs. The majority of other active miRNAs, including miR-92/25, are at lower miRNA:target ratios relative to their total 6/7/8-mer pools and may be more sensitive to physiological ranges of ceRNA competition. We confirmed the differential susceptibility of let-7 and miR-92/25 families by our single-cell reporter system and demonstrated again that only the lower expressed miR-92/25 family responded to this range of overexpression (Figure 6B).

DISCUSSION

Using integrated analysis of small RNA-seq, mRNA-seq, and iCLIP data, we present a biochemical-based quantitative assessment of endogenous miRNA and target concentrations. This is performed in two different mouse cell lines to observe how miRNA:target ratio quantitatively determines miRNA binding across a range of target site affinities. Only 8–12 miRNA seed families in ESCs and MSCs exhibit detectable binding activity, and most of these active miRNAs are not expressed highly enough to appreciably bind their weaker-affinity sites (6-mer target pool). We create a simple target competition model for equilibrium binding, which predicts up to ~30% derepression effects that can be reasonably attributed to physiological levels of high-affinity ceRNA induction (Figure 5C). The principles of this model are validated with comparison to in vivo measured binding (iCLIP) and single-cell miRNA target reporter assays, which reveal up to 35% derepression effects upon addition of 3,000 additional target sites per cell (Figure 6A). Importantly, the reporter assays demonstrate that only active miRNA families with low total miRNA:target ratios are susceptible to ceRNA inductions even up to approximately 10,000 additional target copies per cell. Together, this analysis provides a quantitative context to evaluate miRNA activity in general and the possibility of physiological ceRNA crosstalk in particular.

Previous estimates of miRNA target abundance vary widely and predict very different responses to ceRNA regulation. Wee et al. (2012) discuss a rough estimate of ~500 miRNA target transcripts per cell (50 conserved targets per miRNA \times 10 cpc), and most ceRNA mathematical models tend to follow a similar low range (Ala et al., 2013; Bosia et al., 2013; Figliuzzi et al., 2013). These target pool underestimates amplify the likely effects of ceRNA regulation. By utilizing Ago2 iCLIP-based identification of accessible target sites, our estimates range from 17,000–140,000 6/7/8-mer target cpc per miRNA seed family. However, as noted, restricting the target pool of the highly expressed miR-294 family to 6/7/8-mer seed matches likely underestimates its total pool because the miRNA concentration is high enough to significantly accumulate at 6-mers and spread to lower-affinity sites. The target pool estimates reported here are in a range 10-fold lower than the apparent miRNA target abundances estimated recently in primary hepatocytes by Denzler et al. (2014), but we note that the total mRNA cpc in the cell types analyzed here are approximately 10-fold lower than in hepatocytes, most likely due to the fast dividing nature of ESC and MSC cell lines. Interestingly, we find proportionally little binding in nonprotein-

coding transcripts and no unusually high-coverage CDS sites that would suggest significant competition by exon-derived circRNAs (Hansen et al., 2013; Memczak et al., 2013) in the two nonpathological cell types analyzed here.

Molecular titration regimes canonically require the buffering molecule, i.e., miRNA, to be initially at higher concentrations than the titrant, i.e., target sites, to see threshold release effects (Buchler and Louis, 2008). In contrast to this, we find that total 6/7/8-mer target pool concentration surpasses miRNA concentration for almost all miRNA families. Similarly, Denzler et al. (2014) recently reported that miR-122 target abundance in primary hepatocytes is above miR-122 levels, despite the high expression of miR-122. In addition, by examining mouse livers after metabolic shifts in gene expression or introduction of a potential ceRNA, they showed that due to the large miR-122 target pool, meaningful ceRNA induction required physiologically unreasonable increases in mRNA expression that approach the size of the entire target pool. Additionally in their system, reducing miR-122 expression 3-fold with Antagomirs did not measurably alter the level of ceRNA induction necessary for detectable derepression (Denzler et al., 2014). Despite the differences in experimental systems, our identification of differential ceRNA susceptibility regimens is in agreement with Denzler et al. (2014) regarding the measured behavior of highly abundant miRNAs. Specifically, we find similar results of insensitivity to large target inductions for the highly expressed miR-294 and let-7 families, and our mathematical binding simulation would predict minimal changes in miR-294 high-affinity target binding with a 3-fold miRNA reduction. Although Denzler et al. (2014) probe a more physiological cellular state, our study extends the current understanding of ceRNA regulation to a more generalizable model by globally examining the potential for competition for multiple miRNA:target ratios and using highly sensitive single cell reporters to measure miRNA repression in vivo. Our binding simulations and reporter assays surprisingly demonstrate that targets of active miRNA families with low miRNA:target ratio, like miR-92/25, can be appreciably derepressed at levels of high-affinity ceRNA induction far below that of the entire target pool (Figure 6, ~3,000 copies of 8-mer reporter sites, ~15% of total 6/7/8-mer pool). The threshold target level necessary for derepression in this reporter system also depends on miRNA concentration because miR-92/25 and miR-294 have similar total 6/7/8-mer target abundances (~20,000 cpc), but require very different amounts of reporter expression to show consequential titration effects. Two salient features captured by our model based on fundamental biochemical principles and in vivo Ago binding characteristics help explain target competition susceptibility:

First, the total target pool is partitioned into hierarchical affinity classes that do not compete equally. Because high-affinity target sites are more favorably bound, meaningful competition can occur without approaching expression levels of the total pool of weak and strong sites combined. We find using a mathematical model that accounts for competition within a hierarchical affinity target pool best captures our experimental observations. Specifically, considering a single pool of 6/7/8-mer targets (Figure S6D) similarly as discussed in Denzler et al. (2014), rather than partitioning into affinity groups, predicts less binding for miRNA families with low total miRNA:target ratio than seen in our iCLIP data or inferred from repression data (e.g., <10% binding of miR-92/25 8-mers) and also predicts minimal effects by titration in contradiction to our reporter results. Second, the size of the “effective” target

pool that is meaningfully sequestering miRNA and raising the threshold of derepression is dependent on the miRNA concentration. For active miRNA families with low miRNA:target ratio, the low-affinity targets are mostly unbound and do not actively compete with a potential high-affinity ceRNA. Thus for high-affinity sites, the effective target pool is smaller, and thresholds of derepression can be reached with high-affinity ceRNA induction. Conversely, as miRNA concentration increases, Ago-miRNA complexes spread to weaker and weaker affinity sites, as discussed for miR-294, and the effective target pool size grows too large to be influenced by physiological ranges of ceRNA induction even with high-affinity sites. A minimal theoretical model of posttranscriptional regulation incorporating different target affinities predicts similar titration regimes (Figliuzzi et al., 2013). In this way, both the miRNA concentration and the size and affinities of the competing target pool determine the threshold of target competition.

In addition to miRNA:target ratio, the absolute concentration of the effective target pool also affects ceRNA susceptibility, particularly for exceptional miRNA families with uncommonly high or low target abundance. For example, a very low miRNA: target ratio family such as miR-15/16 is sensitive to ceRNA perturbations but would likely still require unphysiological target increases to appreciably affect repression because its absolute target abundance even for just 7/8-mers is three times higher than average. Conversely, an intermediate-low miRNA:target ratio family such as miR-293 has a high enough miRNA:target ratio to significantly bind its 6-mer targets, but the exceptionally small absolute size of its target pool allows meaningful ceRNA competition in physiological ranges.

The data presented here point to the existence of two common regimens of potential ceRNA susceptibility characterized by different miRNA:target pool ratios and their relationship with the range of target affinities (Figure 6C). Highly expressed miRNA families are likely not susceptible to derepression of their targets by ceRNA competition, due to the buffering capacity provided by the high miRNA and target pool concentrations. Intriguingly, the highly expressed miRNA families described here, miR-294 and let-7, play important master roles in enforcing cell identity (Marson et al., 2008; Melton et al., 2010) and may be positioned to sustain extreme fluctuations in target expression. In contrast, active miRNA families with low miRNA:target ratio may be susceptible to titration by a high-affinity ceRNA. In this model, the affinity of potential ceRNAs is paramount. Target genes can exhibit increased affinity in multiple ways, including favorable sequence context around the miRNA binding site (Bartel, 2009); presence of multiple, cooperatively spaced binding sites (Broderick et al., 2011; Doench et al., 2003; Grimson et al., 2007; Saetrom et al., 2007); and dynamic recruitment by Ago-interacting RNA binding proteins (Kim et al., 2009; van Kouwenhove et al., 2011). We propose that any meaningful ceRNA crosstalk would likely occur through selective communication between genes with multiple high-affinity sites of low, but still significantly repressive, miRNA: target ratio families. Future in vivo investigation of the ceRNA hypothesis should rely on quantitative analysis of the relative perturbation achieved in the context of endogenous partitioned miRNA:target pool ratios to confirm functional regulation.

EXPERIMENTAL PROCEDURES

See Supplemental Experimental Procedures for details.

Quantification of miRNA and mRNA Levels

The small RNA-seq data used for quantitation were from GSE50595 and GSE36978 for ESC and MSC, respectively. Strand-specific poly-A RNA sequencing libraries were prepared using either the UTP (Parkhomchuk and Borodina, 2009) method or TRUseq sample preparation kit from Illumina. External RNA Controls Consortium RNA spike-in standards (Life Technologies) were used to determine cellular concentration of RNA.

iCLIP

iCLIP was performed in both untagged and Flag-HA-tagged Ago2 cell lines by tandem Flag-HA immunoprecipitation essentially as described in Jangi et al., 2014, from approximately 200 million cells following doxycycline induction.

Single-Cell Reporter Assay

miR-294 (AGCACTTA), miR-92/25 (GTGCAATA), miR-293 (GCGGCACA), or let-7 (CTACCTCA) 8-mer sites were cloned into the bidirectional pTRE-Tight-BI (Clontech) eYFP and mCherry reporter constructs described in Mukherji et al., 2011. Reporter constructs and rtTA plasmids were transfected and induced with doxycycline 4 hr posttransfection. Fluorescence-activated cell sorting measurements were taken 24 hr posttransfection and data were processed with FlowJo software.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Sara Gosline, Bryan Matthews, and Ernest Fraenkel for providing MSC RNA-seq data (GEO GSE61031) supported by NIH-ICBP grant U54CA112967. We thank members of the Sharp laboratory for helpful discussions and experimental assistance, specifically Timothy Kelly for generating the doxycycline-inducible Flag-HA-Ago2 construct and performing library sample preps from TT-FHAgo2 cells. We thank David Bartel and Sean McGeary for helpful discussions. We thank Alla Leshinsky, Richard Cook, and the rest of the Swanson Biotechnology Center at the David H. Koch Institute for Integrative Cancer Research at MIT for Illumina and other sequencing services. We thank Stuart Levine and the staff of the BioMicro Center at MIT for sequencing of RNA-seq and CLIP-seq samples. We thank Mervelina Saturno-Condon and the staff of the KI Flow Cytometry for assistance with FACS experiments. We thank A.J. Bhutkar for helpful discussions of computational analyses. This work was supported by NIH grants RO1-GM34277 and R01-CA133404, NCI grants PO1-CA42063 and F32CA139902, and the NCI Cancer Center Support (core) grant P30-CA14051.

REFERENCES

- Ala U, Karreth FA, Bosia C, Pagnani A, Taulli R, Léopold V, Tay Y, Provero P, Zecchina R, Pandolfi PP. Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proc. Natl. Acad. Sci. USA.* 2013; 110:7154–7159. [PubMed: 23536298]
- Arvey A, Larsson E, Sander C, Leslie CS, Marks DS. Target mRNA abundance dilutes microRNA and siRNA activity. *Mol. Syst. Biol.* 2010; 6:363. [PubMed: 20404830]

- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004; 116:281–297. [PubMed: 14744438]
- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009; 136:215–233. [PubMed: 19167326]
- Bissels U, Wild S, Tomiuk S, Holste A, Hafner M, Tuschl T, Bosio A. Absolute quantification of microRNAs by using a universal reference. *RNA*. 2009; 15:2375–2384. [PubMed: 19861428]
- Bosia C, Pagnani A, Zecchina R. Modelling Competing Endogenous RNA Networks. *PLoS ONE*. 2013; 8:e66609. [PubMed: 23840508]
- Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R. The transcription factor titration effect dictates level of gene expression. *Cell*. 2014; 156:1312–1323. [PubMed: 24612990]
- Broderick JA, Salomon WE, Ryder SP, Aronin N, Zamore PD. Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. *RNA*. 2011; 17:1858–1869. [PubMed: 21878547]
- Brown BD, Gentner B, Cantore A, Colleoni S, Amendola M, Zingale A, Baccarini A, Lazzari G, Galli C, Naldini L. Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. *Nat. Biotechnol.* 2007; 25:1457–1467. [PubMed: 18026085]
- Buchler NE, Louis M. Molecular titration and ultrasensitivity in regulatory networks. *J. Mol. Biol.* 2008; 384:1106–1119. [PubMed: 18938177]
- Calabrese JM, Seila AC, Yeo GW, Sharp PA. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA*. 2007; 104:18097–18102. [PubMed: 17989215]
- Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011; 147:358–369. [PubMed: 22000014]
- Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*. 2009; 460:479–486. [PubMed: 19536157]
- de Giorgio A, Krell J, Harding V, Stebbing J, Castellano L. Emerging roles of competing endogenous RNAs in cancer: insights from the regulation of PTEN. *Mol. Cell. Biol.* 2013; 33:3976–3982. [PubMed: 23918803]
- Denzler R, Agarwal V, Stefano J, Bartel DP, Stoffel M. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol. Cell*. 2014; 54:766–776. [PubMed: 24793693]
- Doench JG, Sharp PA. Specificity of microRNA target selection in translational repression. *Genes Dev.* 2004; 18:504–511. [PubMed: 15014042]
- Doench JG, Petersen CP, Sharp PA. siRNAs can function as miRNAs. *Genes Dev.* 2003; 17:438–442. [PubMed: 12600936]
- Ebert MS, Neilson JR, Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods*. 2007; 4:721–726. [PubMed: 17694064]
- Figliuzzi M, Marinari E, De Martino A. MicroRNAs as a selective channel of communication between competing RNAs: a steady-state theory. *Biophys. J.* 2013; 104:1203–1213. [PubMed: 23473503]
- Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 2009; 19:92–105. [PubMed: 18955434]
- Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of Isy-6 and other microRNAs. *Nat. Struct. Mol. Biol.* 2011; 18:1139–1146. [PubMed: 21909094]
- Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*. 2007; 27:91–105. [PubMed: 17612493]
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp A-C, Munschauer M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010; 141:129–141. [PubMed: 20371350]
- Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013; 495:384–388. [PubMed: 23446346]

- Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*. 2013; 153:654–665. [PubMed: 23622248]
- Hwang H-W, Wentzel EA, Mendell JT. A hexanucleotide element directs microRNA nuclear import. *Science*. 2007; 315:97–100. [PubMed: 17204650]
- Jangi M, Boutz PL, Paul P, Sharp PA. Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes Dev*. 2014; 28:637–651. [PubMed: 24637117]
- Jones MR, Quinton LJ, Blahna MT, Neilson JR, Fu S, Ivanov AR, Wolf DA, Mizgerd JP. Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat. Cell Biol*. 2009; 11:1157–1163. [PubMed: 19701194]
- Kallen AN, Zhou X-B, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C, Yi J-S, Zhang H, et al. The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol. Cell*. 2013; 52:101–112. [PubMed: 24055342]
- Karreth FA, Pandolfi PP. ceRNA cross-talk in cancer: when cebling rivalries go awry. *Cancer Discov*. 2013; 3:1113–1121. [PubMed: 24072616]
- Kim HH, Kuwano Y, Srikantan S, Lee EK, Martindale JL, Gorospe M. HuR recruits let-7/RISC to repress c-Myc expression. *Genes Dev*. 2009; 23:1743–1748. [PubMed: 19574298]
- König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol*. 2010; 17:909–915. [PubMed: 20601959]
- Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O'Day E, Chowdhury D, Dykxhoorn DM, Tsai P, Hofmann O, et al. miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to “seedless” 3'UTR microRNA recognition elements. *Mol. Cell*. 2009; 35:610–625. [PubMed: 19748357]
- Leung AKL, Young AG, Bhutkar A, Zheng GX, Bosson AD, Nielsen CB, Sharp PA. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat. Struct. Mol. Biol*. 2011; 18:237–244. [PubMed: 21258322]
- Levine E, Hwa T. Small RNAs establish gene expression thresholds. *Curr. Opin. Microbiol*. 2008; 11:574–579. [PubMed: 18935980]
- Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. The microRNAs of *Caenorhabditis elegans*. *Genes Dev*. 2003; 17:991–1008. [PubMed: 12672692]
- Ling S, Birnbaum Y, Nanhwan MK, Thomas B, Bajaj M, Ye Y. MicroRNA-dependent cross-talk between VEGF and HIF1a in the diabetic retina. *Cell. Signal*. 2013; 25:2840–2847. [PubMed: 24018047]
- Loeb GB, Khan AA, Canner D, Hiatt JB, Shendure J, Darnell RB, Leslie CS, Rudensky AY. Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. *Mol. Cell*. 2012; 48:760–770. [PubMed: 23142080]
- Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*. 2008; 134:521–533. [PubMed: 18692474]
- Meister G. Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet*. 2013; 14:447–459. [PubMed: 23732335]
- Melton C, Judson RL, Blelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature*. 2010; 463:621–626. [PubMed: 20054295]
- Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013; 495:333–338. [PubMed: 23446348]
- Mukherji S, Ebert MS, Zheng GXY, Tsang JS, Sharp PA, van Oudenaarden A. MicroRNAs can generate thresholds in target gene expression. *Nat. Genet*. 2011; 43:854–859. [PubMed: 21857679]
- Mulloikandov G, Baccarini A, Ruzo A, Jayaprakash AD, Tung N, Israelow B, Evans MJ, Sachidanandam R, Brown BD. High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat. Methods*. 2012; 9:840–846. [PubMed: 22751203]
- Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*. 2007; 13:1894–1910. [PubMed: 17872505]

- Parkhomchuk D, Borodina T. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* 2009; 37:e123. [PubMed: 19620212]
- Saetrom P, Heale BSE, Snøve O Jr, Aagaard L, Alluin J, Rossi JJ. Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res.* 2007; 35:2333–2342. [PubMed: 17389647]
- Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell.* 2011; 146:353–358. [PubMed: 21802130]
- Sugimoto Y, König J, Hussain S, Zupan B, Curk T, Frye M, Ule J. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* 2012; 13:R67. [PubMed: 22863408]
- Sumazin P, Yang X, Chiu H-S, Chung W-J, Iyer A, Llobet-Navas D, Rajbhandari P, Bansal M, Guarnieri P, Silva J, Califano A. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell.* 2011; 147:370–381. [PubMed: 22000015]
- Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, Karreth F, Poliseno L, Provero P, Di Cunto F, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell.* 2011; 147:344–357. [PubMed: 22000013]
- Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature.* 2014; 505:344–352. [PubMed: 24429633]
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 2010; 28:511–515. [PubMed: 20436464]
- van Kouwenhove M, Kedde M, Agami R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat. Rev. Cancer.* 2011; 11:644–656. [PubMed: 21822212]
- Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell.* 2012; 151:1055–1067. [PubMed: 23178124]
- Zamudio JR, Kelly TJ, Sharp PA. Argonaute-bound small RNAs from promoter-proximal RNA polymerase II. *Cell.* 2014; 156:920–934. [PubMed: 24581493]
- Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, Yeo GW. Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* 2010; 17:173–179. [PubMed: 20062054]

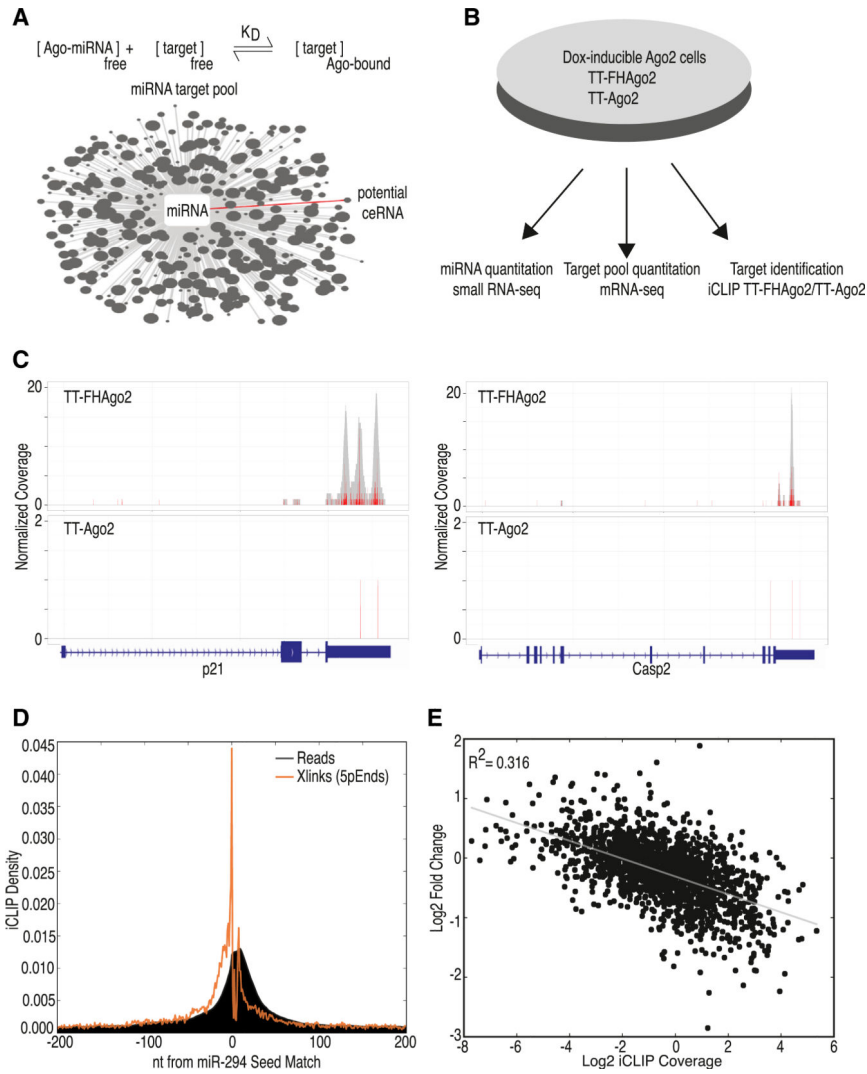


Figure 1. In Vivo miRNA and Target Pool Quantitation for Prediction of ceRNA Effects
 (A) Steady-state equation governing miRNA interactions (top) and illustration of a potential ceRNA gene in context of a miRNA's total target pool (bottom). Nodes depict target genes and size represents different expression levels.
 (B) Experimental system for quantifying miRNA regulation networks. TT-FHago2 and TT-Ago2 ESCs lack endogenous Ago1-4 and express only FLAG-HA-hAgo2 or untagged hAgo2 under a doxycycline inducible promoter.
 (C) iCLIP read coverage at p21 and Casp2 genes from TT-FHago2 or TT-Ago2 ESCs. iCLIP read 5' ends predict the exact crosslink site and are indicated in red. Coverage is normalized to mean values of background RNA to allow comparison between data sets, as described in the Experimental Procedures. Scale represents 0–20 for TT-FHago2 coverage and 0–2 for TT-Ago2 coverage.
 (D) iCLIP read or crosslink (Xlink) density across all miR-294 seed match sites (GCACTT) in expressed 3' UTRs. Xlink sites represent 5' end of sequenced reads. iCLIP density represents read coverage or number of Xlinks at a given nt position/total read coverage or number of Xlinks within the ± 200 nt window surrounding the target sites; 0 in the x axis
 (E) Scatter plot of Log₂ Fold Change vs Log₂ iCLIP Coverage with $R^2 = 0.316$.

corresponds to the first nucleotide of the 7-mer seed match (i.e., across from position 8 of the miRNA).

(E) Correlation between \log_2 iCLIP coverage per gene and \log_2 gene repression, for each gene containing at least one 3' UTR iCLIP cluster. iCLIP coverage represents the number of reads across all 3' UTR clusters normalized to wild-type (WT) gene expression (Supplemental Experimental Procedures). Gene repression is the \log_2 Ago+/Ago- fold change/median Ago+/Ago- value of a random control set (Supplemental Experimental Procedures). The least-squares linear regression best-fit line is depicted in gray with the coefficient of determination noted (R^2).

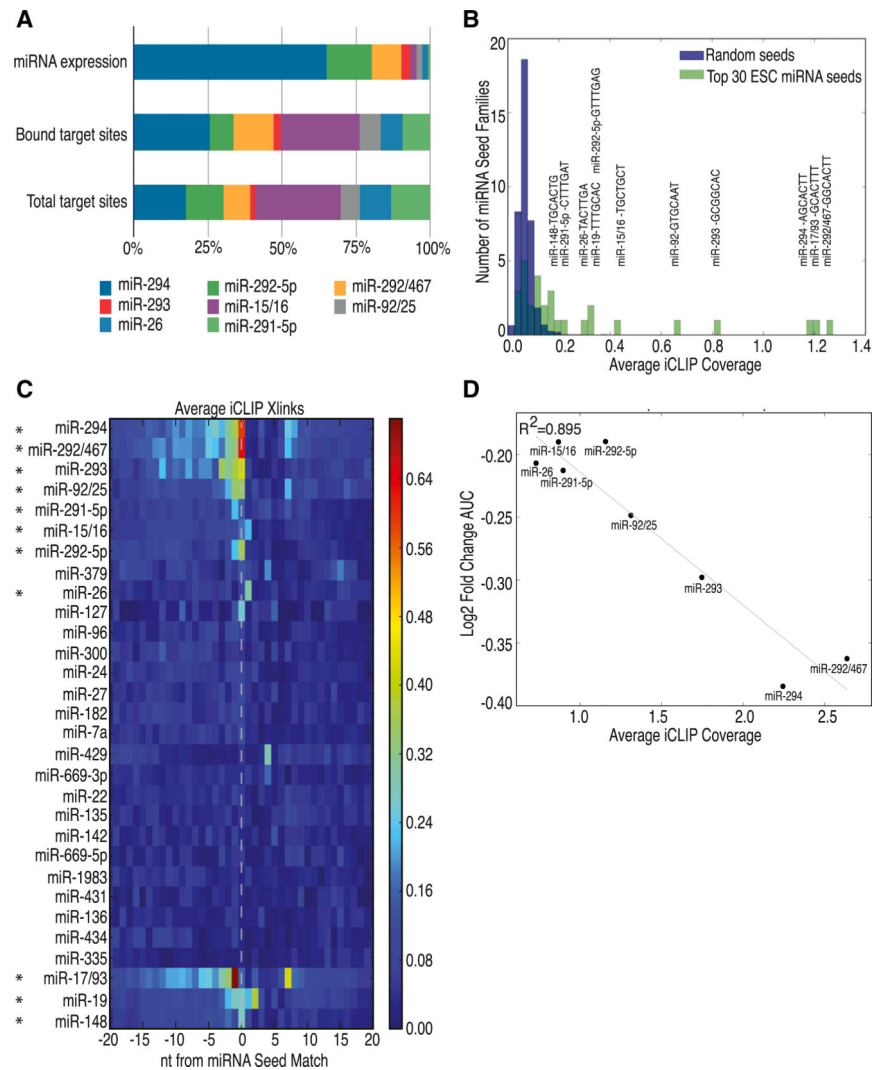


Figure 2. Quantitative iCLIP Coverage across Entire Target Pool Identifies Active ESC miRNAs
 (A) Proportion of miRNA expression, bound target sites (≥ 1 iCLIP read), or total target sites for the indicated miRNA seed families.

(B) Histogram of the average iCLIP coverage per expressed 3' UTR 8-mer site for each of the top 30 expressed ESC miRNA families (green). miRNA family name and 7-mer seed match is indicated above corresponding value for families with signal over background control ($p < 0.01$). The histogram for 998 random seed “families” used to estimate background iCLIP coverage levels is in blue.

(C) Heatmap of average crosslinks (Xlinks) per nucleotide across all 7/8-mer sites in expressed 3' UTRs (≥ 1 iCLIP read within 10 nt of the seed match) for each of the top 30 expressed miRNA families in ESCs. These are ranked by Xlink signal. The three miRNA families with seed matches similar to highly expressed miRNAs are moved to the bottom; 0 in the x axis corresponds to the first nucleotide of the 7-mer seed match. Average Xlinks color scale is indicated to the right. Asterisks mark significantly active families identified in Figure 3B.

(D) Plot of average iCLIP coverage per conserved 3' UTR 7/8-mer site versus Ago-dependent expression change of the corresponding gene set, calculated as the area under the curve (AUC) between the cumulative distribution of \log_2 Ago+/Ago- FPKM test set values and matched control gene values. Only significantly active ESC miRNA families are plotted. Genes considered have a bound 3' UTR 7/8-mer site with PhastCons score > 0.8 . The least-squares linear regression best-fit line is depicted in gray with the coefficient of determination noted (R^2).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

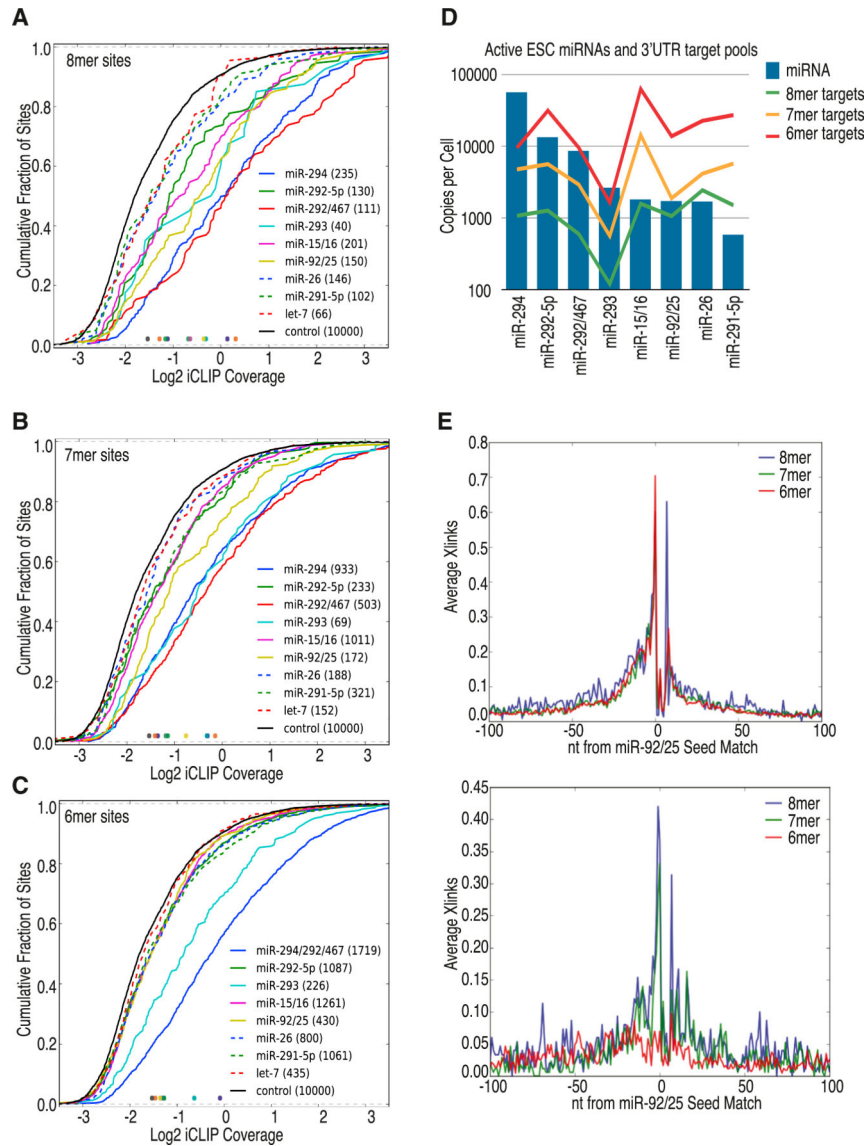


Figure 3. Relative Target Pool Abundance Influences Site Type Binding Pattern

(A) Cumulative distribution of log₂ iCLIP coverage at each 8-mer target site for the active ESC miRNA families and the nonexpressed let-7 miRNA family (included as an additional estimate of background signal in this analysis). 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. 3' UTR matched control set is included (Experimental Procedures).

(B) Same as (A) for 7-mer sites.

(C) Same as (A) for 6-mer sites.

(D) Measured copies per cell (cpc) values for miRNA and corresponding 3' UTR target pools of indicated site type for significantly active ESC miRNAs. y axis is log scale.

(E) Average crosslinks (Xlinks) per nucleotide across all target sites in expressed 3' UTRs (≥ 1 iCLIP read within 10 nt), for highly expressed miR-294 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 7-mer seed match.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

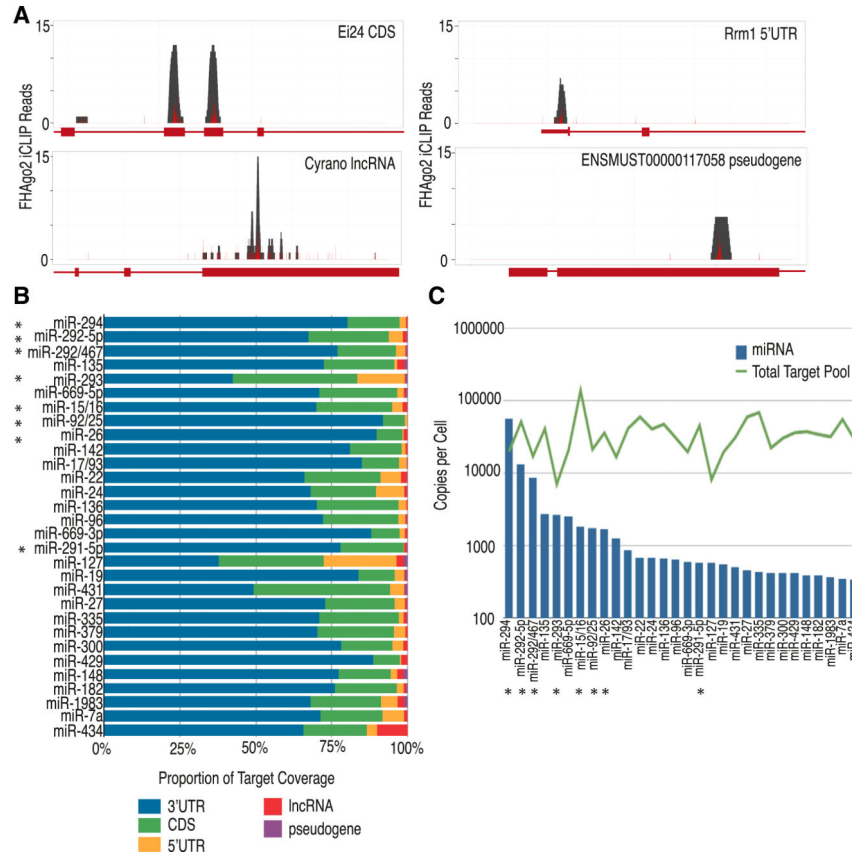


Figure 4. Total Target Pools Are Dominated by 3' UTR Binding and Above the Concentration of Most miRNAs

(A) Coverage plot of FHAgO2 iCLIP read (black) and crosslink (Xlink) (red) coverage at the indicated non-3' UTR ESC targets.

(B) Proportion of iCLIP reads at 7/8-mer target sites coming from indicated genic categories for each of the top 30 expressed ESC miRNAs. iCLIP reads at sites with ambiguous genic annotation are counted in both categories. Asterisks mark significantly active families identified in Figure 3B.

(C) Measured copies per cell (cpc) values for miRNA and corresponding total target abundance (6/7/8-mer sites from all genic categories). Top 30 expressed miRNAs are plotted. miRNA family cpc is a sum of all miRNAs sharing the same 6-mer seed, thus miR-294 and miR-292/467 have identical values. The y axis is log scale.

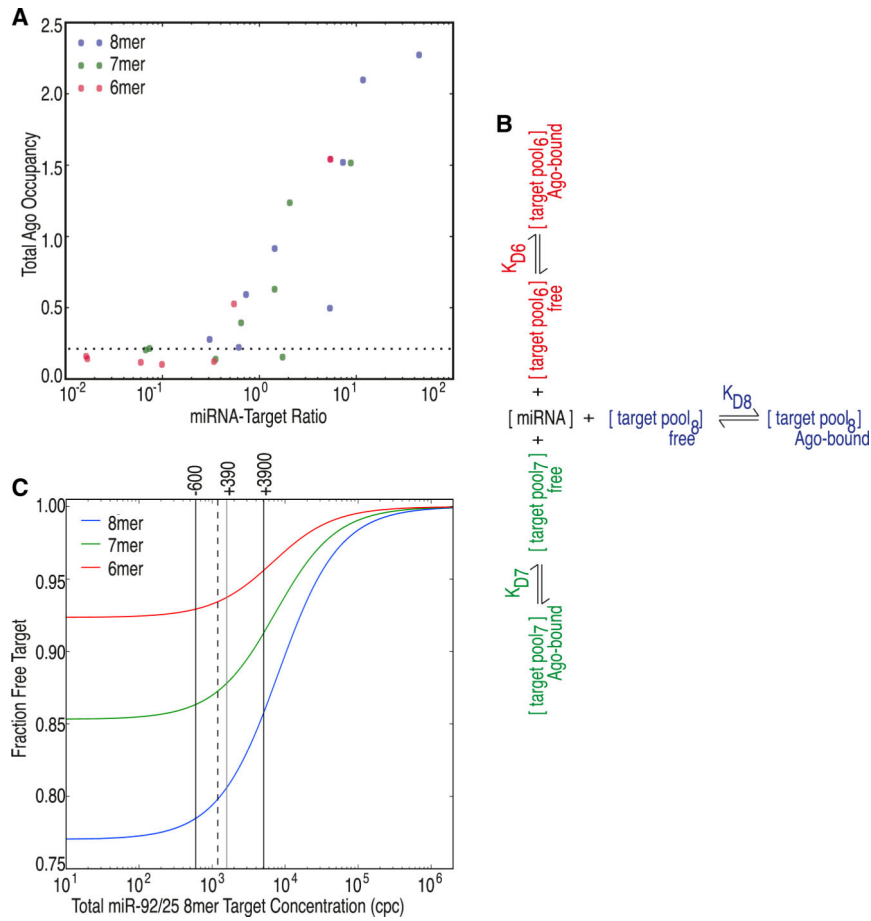


Figure 5. Mathematical Model of miRNA Binding Predicts ceRNA Effects Based on Intratarget Pool Competition

(A) miRNA and target stoichiometries per site type are plotted against total binding estimates from iCLIP, for each active ESC miRNA. miRNA:target ratio represents miRNA cpc/ site type target pool cpc. Total Ago occupancy represents iCLIP reads per million (RPM)/site type target pool cpc. The $p = 0.01$ is the empirical p value calculated from Total Ago Occupancy values of 1,000 random sets of control sites is indicated by dotted line. The x axis is log scale. Total Ago occupancy values for each miRNA family are listed in Table 1 and can be used to identify the corresponding point on the graph.

(B) Diagram of Ago-miRNA binding competition between different affinity target sites. The total target pool is partitioned by seed match type with different K_D and concentrations (indicated by color and subscript).

(C) Simulated target titration curves for a representative lowly expressed miRNA (miR-92/25) showing the relationship between proportion of free targets and total 8-mer target pool concentration in cpc. Dotted black vertical line indicates endogenous 8-mer target pool concentration in cpc. 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 3×8 -mer sites (600 site cpc) or 10- and 100-fold upregulation (right of dotted line) of an average target gene (13 cpc) containing 3×8 -mer sites (390, 3,900 site cpc). The x axis is log scale.

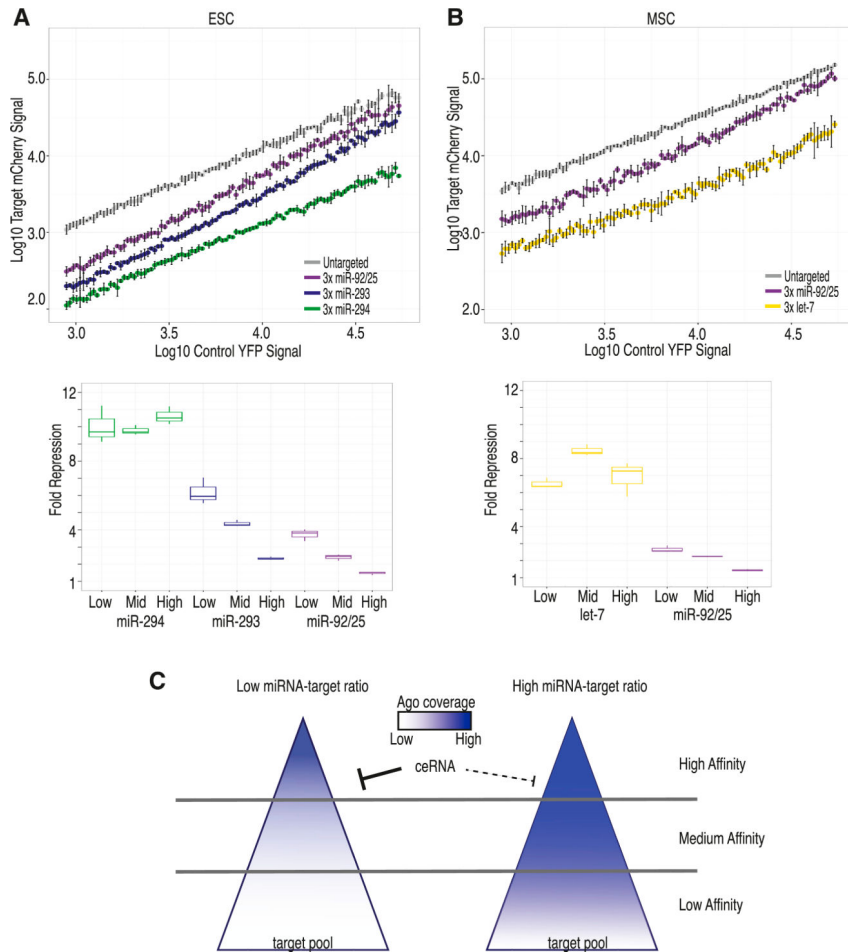


Figure 6. Single Cell Reporter Assays Confirm Differential Susceptibility to Target Competition Based on miRNA:Target Ratio

(A) Top: log-log plot of miRNA targeted or untargeted mCherry mean fluorescence across 100 bins of eYFP signal in ESCs for miR-92/25 (purple), miR-293 (blue), or miR-294 (green) reporters. Bottom: box plots of fold repression ($mCherry_{untargeted}/mCherry_{3 \times sites}$) in first ten bins (Low), middle ten bins centered around log₁₀ value of 3.75 (Mid), and last ten bins (High). Targeted mCherry constructs contained 3 × 8-mer sites. Untargeted mCherry contained 3 × 8-mer sites for the nonexpressed let-7 miRNA. Error bars represent SEM from three biological replicates. All fluorescent values are background normalized (Experimental Procedures).

(B) miRNA-mediated fold repression across a range of reporter induction as in (A) except measured in MSCs, using mCherry reporters containing 3 × 8-mer sites for miR-92/25 (purple) or let-7 (yellow). Untargeted mCherry contains a short 3' UTR with no miRNA binding sites. Error bars represent SEM from three biological replicates.

(C) Hierarchical affinity model of miRNA target competition. The endogenous miRNA target pool is indicated in a partitioned triangle illustrating the hierarchical affinity classes and their relative abundances. Ago-miRNA target pool coverage will depend on the relative miRNA:target ratio as indicated in red for exemplary low (left) or high (right) ratios and spread toward low-affinity targets. Individual miRNA target pools demonstrate differential

susceptibility to an identical level of high-affinity ceRNA overexpression (black bars) as described in the Discussion.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
miRNA and Target Pool Concentrations and iCLIP Coverage Values of Active miRNA Families

miRNA Family	Seed Match	[miRNA] cpc/nM	8-mer [target] cpc/nM	7-mer [target] cpc/nM	6-mer [target] cpc/nM	Total Ago Occupancy 8-mer Targets	Total Ago Occupancy 7-mer Targets	Total Ago Occupancy 6-mer Targets
ESC miR-294	AGCACTT	56,700/471	1,271/11	6,453/54	12,077/100	2.27	1.52	1.54
miR-292-5p	GTTTGAG	13,328/111	2,492/21	7,664/64	40,883/340	0.5	0.15	0.12
miR-292/467	GGCACTT	8,643/72	736/6	4,185/35	12,086/100	2.1	1.24	1.54
miR-293	GCGGCAC	2,644/22	360/3	1,847/15	4,807/40	1.53	0.63	0.53
miR-15/16	TGCTGCT	1,814/15	2,484/21	24,549/204	11,0918/921	0.59	0.21	0.16
miR-92/25	GTCCAAT	1,720/14	1,195/10	2,653/22	17,367/144	0.91	0.39	0.1
miR-26	TACTTGA	1,696/14	2,758/23	4,784/40	28,284/235	0.22	0.14	0.12
miR-291-5p	CTTTGAT	584/5	1,903/16	8,660/72	34,740/289	0.28	0.2	0.14
MSC let-7	CTACCTC	41,748/46.2	1,380/1.5	2,884/3.2	20,414/22.6	3.53	1.99	1.16
miR-125-5p	CTCAGGG	19,309/21.4	1,448/1.6	4,442/4.9	11,378/12.6	1.06	0.7	0.3
miR-26	TACTTGA	11,624/12.9	1,398/1.5	2,309/2.6	12,612/14.0	1.04	0.54	0.41
miR-15/16	TGCTGCT	6,610/7.3	1,450/1.6	12,833/14.2	56,220/62.3	2.72	1.08	0.71
miR-21	ATAAGCT	5,674/6.3	520/0.6	1,730/1.9	7,694/8.5	1.44	0.65	0.46
miR-99	TACGGGT	4,813/5.3	78/0.1	399/0.4	2,154/2.4	1.14	1.03	0.95
miR-31	TCTTGCC	2,228/2.5	1,426/1.6	3,603/4.0	13,285/14.7	1.12	0.79	0.41
miR-199	ACTACTG	1,226/1.4	912/1.0	2,827/3.1	10,837/12.0	0.91	0.59	0.5
miR-92/25	GTCCAAT	912/1.0	787/0.9	1,727/1.9	8,239/9.1	2.27	1.01	0.43
miR-27	ACTGTGA	849/0.9	2,266/2.5	6,011/6.7	21,091/23.4	1.29	0.68	0.52
miR-30	TGTTTAC	562/0.6	2,411/2.7	3,296/3.7	6,072/6.7	1.33	1.11	0.56
miR-17/20/93	GCACTTT	302/0.3	1,018/1.1	3,475/3.8	14,881/16.5	1.91	1.08	0.57

Measured copies per cell (cpc) and converted nM concentrations for all ESC and MSC miRNA families with statistically enriched average iCLIP coverage over random control sites ($p < 0.01$). Target concentration [target] is the endogenous iCLIP-estimated target abundance per miRNA and site type pair across all genomic regions (corresponding to values graphed in Figures 4C and S7E). Total Ago Occupancy, iCLIP reads per million (RPM) across corresponding sites/site type target pool cpc.