
Determinants of targeting by endogenous and exogenous microRNAs and siRNAs

CYDNEY B. NIELSEN,^{1,3} NOAM SHOMRON,^{1,3} RICKARD SANDBERG,¹ ERAN HORNSTEIN,^{2,4} JACOB KITZMAN,^{1,5} and CHRISTOPHER B. BURGE¹

¹Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

²Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

ABSTRACT

Vertebrate mRNAs are frequently targeted for post-transcriptional repression by microRNAs (miRNAs) through mechanisms involving pairing of 3' UTR seed matches to bases at the 5' end of miRNAs. Through analysis of expression array data following miRNA or siRNA overexpression or inhibition, we found that mRNA fold change increases multiplicatively (i.e., log-additively) with seed match count and that a single 8 mer seed match mediates down-regulation comparable to two 7 mer seed matches. We identified several targeting determinants that enhance seed match-associated mRNA repression, including the presence of adenosine opposite miRNA base 1 and of adenosine or uridine opposite miRNA base 9, independent of complementarity to the siRNA/miRNA. Increased sequence conservation in the ~50 bases 5' and 3' of the seed match and increased AU content 3' of the seed match were each independently associated with increased mRNA down-regulation. All of these determinants are enriched in the vicinity of conserved miRNA seed matches, supporting their activity in endogenous miRNA targeting. Together, our results enable improved siRNA off-target prediction, allow integrated ranking of conserved and nonconserved miRNA targets, and show that targeting by endogenous and exogenous miRNAs/siRNAs involves similar or identical determinants.

Keywords: target prediction; Dicer knockout; RNAi; miR; mouse embryonic fibroblast

INTRODUCTION

Precise control of mRNA and protein levels in different cell types requires regulation at multiple levels. In metazoans, a large proportion of mRNAs is targeted for post-transcriptional repression by ~22 nucleotide (nt) microRNAs (miRNAs). Identified as developmental regulators, miRNAs are now known to play roles in diverse biological processes including control of proliferation, apoptosis, stress resistance, and metabolism (Ambros 2004; Bartel 2004; Filipowicz et al. 2005; Zamore and Haley 2005).

miRNAs were initially described as exerting their effects primarily by inhibiting productive translation of mRNAs (Lee et al. 1993; Wightman et al. 1993). More recently, several studies have demonstrated that animal miRNAs can

direct accelerated decay of targeted mRNAs (Hutvagner and Zamore 2002; Bagga et al. 2005; Lim et al. 2005; Rehwinkel et al. 2005; Giraldez et al. 2006) and that siRNAs commonly direct decay of "off-target" mRNAs (Jackson et al. 2003). When they possess near-perfect complementarity to a targeted mRNA, miRNAs can direct endoribonucleolytic cleavage of mRNAs ("slicer" activity) by Argonaute2 (AGO2) (Hutvagner and Zamore 2002; Llave et al. 2002; Meister et al. 2004). This type of targeting is predominant in plants, but appears to occur only rarely for animal miRNAs (Yekta et al. 2004; Jones-Rhoades et al. 2006). For typical metazoan targets that possess complementarity only to a segment at the miRNA 5' end, miRNAs appear to direct mRNA degradation by mechanisms that may involve AGO2 but do not appear to involve its slicer activity (Bagga et al. 2005; Schmitter et al. 2006). Instead, decay may be promoted by relocalization of targeted mRNAs to specific cytoplasmic locations, which can be sites of mRNA decapping and degradation (for review, see Valencia-Sanchez et al. 2006) and/or by acceleration of mRNA deadenylation (Giraldez et al. 2006; Wu et al. 2006).

In many studies, miRNA regulation has been assessed only at the protein level, without distinguishing the relative

³These authors contributed equally to this work.

Present addresses: ⁴Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel; ⁵NimbleGen Systems Inc., Madison, Wisconsin 53711, USA.

Reprint requests to: Christopher B. Burge, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; e-mail: cburge@mit.edu; fax: (617) 452-2936.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.768207>.

contributions of effects on mRNA decay and on inhibition of translation. However, for some individual targets, both mRNA-level and protein-level effects have been measured. For the classical let-7 and lin-4 target genes *lin-41*, *lin-14*, and *lin-28* (Lee et al. 1993; Wightman et al. 1993), a recent study found a predominant effect on mRNA stability (Bagga et al. 2005). Studies of transfected miRNAs or siRNAs using transfected reporters with moderate degrees of complementarity have typically reported significant effects on protein levels, with modest or negligible effects on mRNA levels (Zeng et al. 2002, 2003; Doench et al. 2003; Doench and Sharp 2004). However, studies that have examined changes in the expression of endogenous mRNAs in response to manipulation of miRNAs have generally observed widespread miRNA-associated changes in mRNA levels. Following miRNA overexpression, Lim and colleagues (2005) observed down-regulation of sets of mRNAs that were enriched for predicted miRNA targets and for genes with low expression levels in the tissues where the miRNAs were naturally expressed, supporting the physiological relevance of this effect. Inhibiting the expression of the critical miRNA processing enzymes Dicer and Drosha also yields specific derepression of predicted miRNA targets at the mRNA level (Rehwinkel et al. 2005; Giraldez et al. 2006; Schmitter et al. 2006). Thus, perturbations of miRNA expression commonly affect the levels of endogenous mRNAs, and effects on mRNA stability appear to be an important component of the endogenous function of miRNAs.

The special importance of the miRNA 5' end was suggested by early studies (Lee et al. 1993; Wightman et al. 1993; Lai 2002). Since then, the critical importance of pairing to the miRNA seed, comprising bases 2–7 from the miRNA 5' end, has been established through extensive comparative genomic and experimental studies (Lewis et al. 2003, 2005; Doench and Sharp 2004; Brennecke et al. 2005; Stark et al. 2005). The degree of conservation above background in orthologous 3' UTRs of “seed match” segments having Watson–Crick (WC) complementarity (“matching”) to the seed regions of conserved miRNAs can be used to estimate the number of conserved targets. This approach indicated that at least one-third of mammalian mRNAs are conserved targets of one or more conserved miRNAs (Lewis et al. 2005), and related methods have indicated that a comparably large fraction of *Drosophila* mRNAs represent conserved miRNA targets (Brennecke et al. 2005; Grun et al. 2005). Recent analyses of mRNA sequence and expression patterns have detected pervasive effects of miRNAs on mRNA expression and evolution, suggesting that most mRNAs are subject either to direct miRNA regulation or to evolutionary pressure to avoid miRNA targeting (Farh et al. 2005; Stark et al. 2005).

Some targets identified genetically possess complementarity to bases at the 3' as well as 5' ends of miRNAs, which may confer specificity to individual members of a miRNA family. However, comparative genomic approaches have

determined that “seed only” type targets comprise the vast majority of all conserved miRNA targets (Brennecke et al. 2005; Lewis et al. 2005). This conclusion is also supported by miRNA overexpression experiments; e.g., in the study by Lim and colleagues (2005), 88% of mRNAs whose expression was significantly repressed following transfection of miR-1 contained seed matches in their 3' UTRs, and replacing the 3' end of the transfected miRNA by unrelated sequences yielded a largely overlapping set of down-regulated mRNAs. However, the presence of a minimal seed match is not generally sufficient to generate detectable mRNA down-regulation; e.g., only about one-tenth to one-twentieth of expressed genes containing a 6 nt seed match in the Lim study were significantly down-regulated (not shown), suggesting that miRNA regulation is strongly influenced by additional targeting determinants.

To assess a variety of potential targeting determinants, we analyzed the effects on global mRNA expression in miRNA and siRNA overexpression studies. In parallel, the effects on mRNA expression of endogenous mouse miRNAs were analyzed following knockout of the *Dicer1* gene, which is essential for miRNA maturation in vertebrates. Our results uncover additional rules and determinants for targeting that hold for both endogenously expressed miRNAs and exogenous miRNAs and siRNAs.

RESULTS AND DISCUSSION

A hierarchy of extended seed match types associated with different degrees of target down-regulation

To explore miRNA targeting determinants, we analyzed global mRNA expression data following transfection of the tissue-specific miRNAs miR-1 and miR-124 into HeLa cells reported by Lim and colleagues (2005). To assess the impact of a putative targeting determinant on down-regulation, we compared the distributions of log fold change (LFC), defined as the log base 2 of expression in miRNA-transfected cells over that in mock-transfected cells, for mRNAs containing and lacking the putative determinant. The cumulative distribution functions (CDFs) of LFCs for these two mRNA sets could then be compared and the significance of differences assessed using a Wilcoxon rank sum test (Materials and Methods). Using this approach, mRNA sets with and without specific hexanucleotides (6 mer) in their 3' UTRs were compared for all 4096 possible 6 mer in both miRNA transfection datasets. Following miR-1 transfection, the most significant down-regulation was observed for mRNAs containing the 6 mer CAUUCU, which has perfect WC complementarity to miR-1 bases 2–7 ($P < 10^{-34}$, Bonferroni corrected for the 4096 comparisons performed). For miR-124, the most significant down-regulation was associated with GUGCCU ($P < 10^{-58}$) and UGCCUU ($P < 10^{-26}$), which are complementary to miR-124 bases 3–8 and 2–7, respectively. These observations,

obtained without the need to define significantly up- and down-regulated mRNA sets, are entirely consistent with the motif-finding analyses of significantly down-regulated mRNAs by Lim and colleagues (2005), and suggest that pairing to miRNA seed matches was a primary effector of mRNA down-regulation in this experiment.

Stronger down-regulation was observed for mRNAs containing additional matching to the transfected miRNAs in their 3' UTRs beyond the 6-base seed match (Fig. 1B,D). As shown in Figure 1A, we use the notation m_1, m_2, \dots to refer to miRNA bases, starting at the 5'-most base, and t_1, t_2, \dots to refer to positions in target mRNAs opposite miRNA bases m_1, m_2, \dots , respectively, in presumptive seed:seed match duplexes (Lewis et al. 2005). Those mRNAs that contained a seed match 6 mer flanked by a WC match to miRNA base 8 (Fig. 1, M8 7 mer; red curves) exhibited enhanced down-regulation relative to those that contained a 6 mer alone ($P < 10^{-5}$ for both miR-124 and miR-1). The presence of an adenosine at position t_1 (Fig. 1, A1 7 mer; blue curves) was also associated with greater mRNA down-regulation than a seed match alone for both miRNAs ($P < 0.03$, $P < 10^{-5}$ for miR-124 and miR-1, respectively). Those mRNAs that contained seed matches flanked by both of these features (Fig. 1, M8-A1 8 mer; purple curves) exhibited greater mRNA down-regulation ($P < 0.002$ relative to A1 7 mer for both miRNAs). Modest but significant down-regulation was observed for mRNAs that contained only a seed match 6 mer not flanked by an M8 or A1 base (Fig. 1, 6 mer; green curves) for miR-1 ($P < 10^{-4}$), but not miR-124 (NS). Therefore, the highly significant down-regulation observed for the seed match 6 mer in the independent 6 mer analysis is attributable primarily to the effects of M8 and A1 7 mer and M8-A1 8 mer. We consider these 7 mer and 8 mer and the seed match 6 mer to represent distinct "seed match types" and refer to these 7 mer and 8 mer collectively as "extended seed matches." These observations suggest that the presence of these types of extended seed matches, not just of a seed match 6 mer, may be generally required for effective miRNA-directed down-regulation of mRNAs. A similar hierarchy of seed match types was observed when mRNAs containing conserved and nonconserved extended seed matches were analyzed separately (Supplemental Fig. S1). All mRNA sets in the above analyses were mutually exclusive, and no significant differences between the distributions of expression levels of mRNAs containing different seed match types were detected by rank sum test.

There are multiple ways to think about the magnitude of the mRNA down-regulation effect attributable to a given seed match type. One perspective is to consider the set of mRNAs containing the given seed match type that were significantly down-regulated (e.g., those with LFC $< 97.5\%$ of control mRNAs lacking seed matches). By this criterion, 45% of expressed mRNAs containing 8 mer seed matches were down-regulated following miR-124 transfection. Among

these genes, the average LFC was -0.97 , corresponding to a $100 \times (1 - 2^{-0.97}) = 49\%$ decrease in expression. For M8 7 mer seed matches, 25% of mRNAs were significantly down-regulated, and these had a mean LFC of -0.87 , a 45% decrease in expression. The fraction of mRNAs significantly down-regulated, two measures of the magnitude of down-regulation, and rank sum P -values for all of the analyses shown in Figure 1 and Supplemental Figure S1 are provided in Supplemental Tables S1 and S2, respectively.

Another perspective is to consider all of the data and to calculate the mean normalized log fold change (nLFC), defined as the mean LFC for expressed mRNAs lacking seed matches to the transfected miRNA minus the mean LFC for expressed mRNAs containing the given seed match type in their 3' UTRs. (As defined, the nLFC will be positive if a seed match type is associated with mRNA down-regulation.) For miR-124, the mean nLFC value of the M8-A1 8 mer seed match type was 0.56, roughly twice that for the M8 7 mer (0.25). Thus, if fold change is multiplicative (i.e., log-additive) in the number of seed matches (as will be shown below), then the fold change associated with one 8 mer seed match is roughly equivalent to that associated with two 7 mer seed matches. Because it uses the largest possible set of mRNAs, and is less sensitive to the shape of the tail of the no-seed-match distribution, the mean nLFC is a more robust statistic for analyzing seed-match-associated effects than the fraction of significantly down-regulated mRNAs. For this reason, mean nLFC is used extensively in this study. However, by considering all seed-match-containing mRNAs, not just those with significant changes, the mean nLFC likely underestimates the true magnitude of miRNA effects on target mRNA levels, and mRNA fold change values underestimate protein-level changes (see below) because miRNAs often inhibit translation as well as mRNA stability.

Seed match hierarchy supported by siRNA, comparative genomic, and luciferase data

Exogenously added siRNAs complementary to seed match segments in mRNA 3' UTRs have been observed to direct similar effects at both mRNA and protein levels as transfected miRNAs (Doench et al. 2003; Jackson et al. 2003, 2006). Using global mRNA expression data following transfection of siRNAs generated by Jackson and colleagues (Jackson et al. 2003, 2006) and Schwarz and colleagues (Schwarz et al. 2006), similar shifts in mRNA expression for seed-match-containing mRNAs were seen for siRNAs as were observed for transfected miRNAs. Those mRNAs that contained M8-A1 8 mer matches to the siRNAs were most strongly down-regulated, with the nLFC value for 8 mer roughly twice that seen for M8 or A1 7 mer, as was seen for transfected miRNAs. The effects of M8 7 mer were comparable to that for A1 7 mer, and both had nLFC values more than twice that of 6 mer, the same ordering of seed

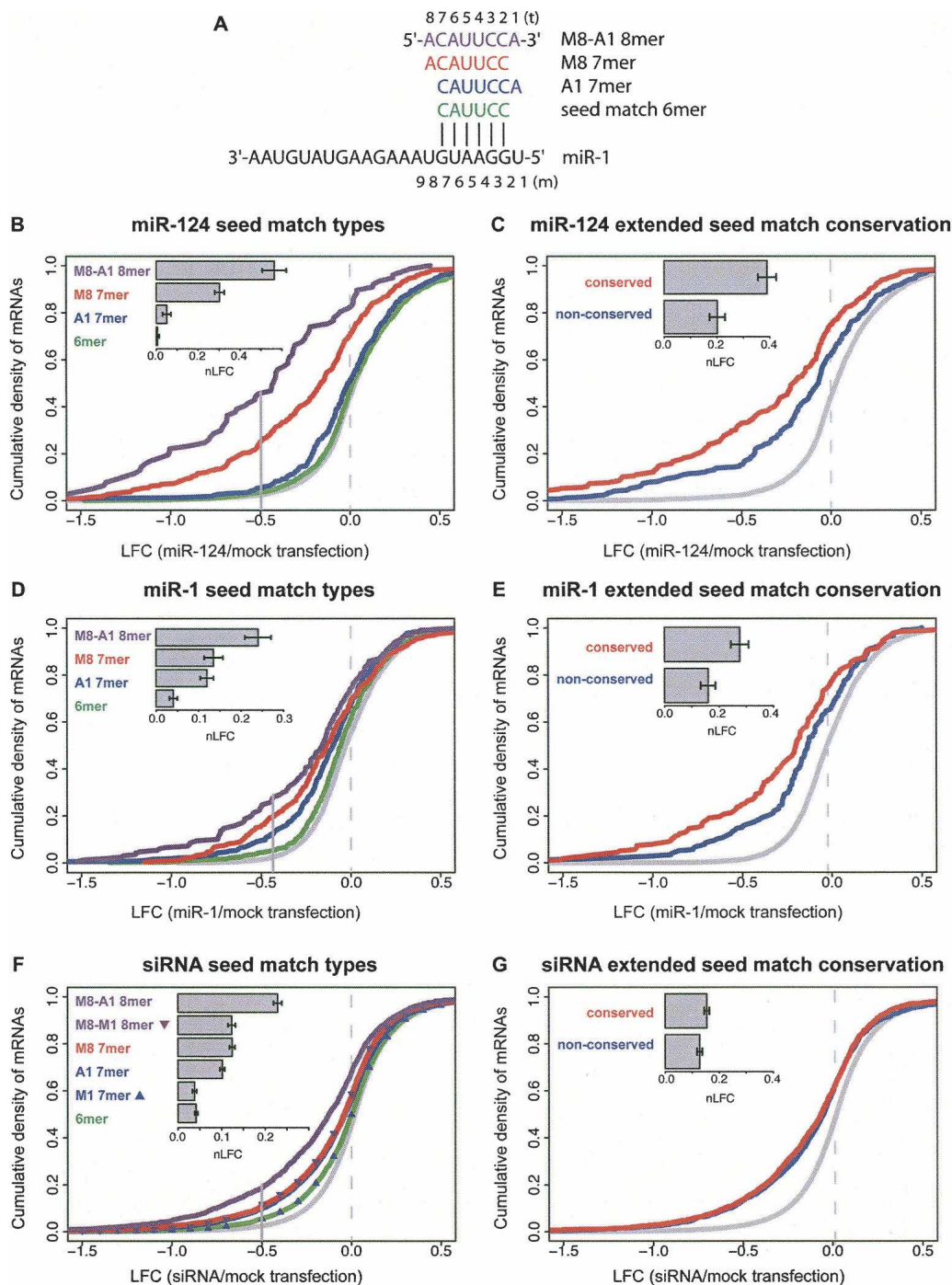


FIGURE 1. Effects of seed match type and conservation on mRNA repression for miRNAs and siRNAs. (A) Seed match types and numbering system, illustrated for miR-1. Positions in the miRNA are numbered 5'-3'. (Seed match 6 mer) WC inverse complement of miRNA bases 2-7; (A1) presence of adenosine opposite miRNA base 1; (M8) WC match to miRNA base 8. (B) CDFs (cumulative distribution functions) of LFCs (\log_2 fold change) for mRNAs containing indicated miR-124 seed match types (colored lines and labels) or no miR-124 seed matches (gray line) following transfection of miR-124. (Solid vertical gray line) The LFC above which 97.5% of the no-seed-match mRNA set falls. (Inset bar plot) nLFC (normalized \log_2 fold change) values for each seed match type with error bars indicating standard error. Data for panels B-E are from Lim et al. (2005). Plots include only mRNAs containing exactly one miR-124 seed match, and thus the seed match type sets are mutually exclusive. The distribution of mRNA expression values did not differ significantly between seed match type sets ($P > 0.05$ by rank sum test). All seed match types except the 6 mer have distributions significantly different from the no-seed-match class ($P < 0.005$ by rank sum test). (C) CDFs of LFCs for mutually exclusive mRNA sets containing conserved (red) or nonconserved (blue) extended seed matches to miR-124, or no seed matches (gray); the conserved and nonconserved sets are significantly different ($P < 0.001$ by rank sum test). The "nonconserved" mRNA set contains exclusively nonconserved seed matches; the "conserved" mRNAs may also contain nonconserved seed matches. The nonconserved set was sampled to match the conserved set in seed match type and count, overall UTR conservation, and initial mRNA expression level (Supplemental Fig. S2). (D) Same as B for miR-1. All seed match type classes are significantly different from the no-seed-match class ($P < 10^{-4}$ by rank sum test). (E) Same as C for miR-1. The CDFs of conserved and nonconserved mRNA sets are significantly different ($P < 0.01$ by rank sum test). (F) Same as B for pooled set of 33 "effective" siRNAs that begin with non-U bases (Supplemental Material). Additional seed match classes containing M1 are shown (triangles). All seed match types have distributions significantly different from the no-seed-match class ($P < 10^{-14}$ by rank sum test). (G) Same analysis and controls as C for pooled set of siRNAs. The CDFs of conserved and nonconserved mRNA sets are not significantly different ($P > 0.05$ by rank sum test). See Supplemental Table S1 for additional statistics.

match types as was observed for transfected miRNAs (Fig. 1F). Thus, the targeting rules observed for transfected miRNAs generally apply to transfected siRNAs, suggesting that transfected siRNAs and miRNAs enter similar or identical silencing complexes and mediate similar effects on their targets (Hutvagner and Zamore 2002).

Analyses of sequence conservation in mammalian 3' UTRs have previously found that a 6 mer seed match is the minimal unit of sequence that suffices to elicit a significant conservation "signal" above "noise" for conserved vertebrate miRNAs (Lewis et al. 2005), but that requiring conservation of M8 and/or A1 bases greatly increased the signal:noise ratio. In alignments of five vertebrate genomes, the signal:noise ratio increased from 2.4:1 for 6 mer to 3.8:1 each for M8 and A1 7 mer, to 5.6:1 for M8-A1 8 mer (Lewis et al. 2005). Thus, M8 and A1 bases adjacent to conserved seed matches in mammalian 3' UTRs are very often conserved, and the relative ordering of comparative genomic signal:noise ratios for different seed match types generally agreed with the relative magnitude of mRNA down-regulation effects observed above for transfected miRNAs (i.e., M8-A1 8 mer > M8 7 mer ≥ A1 7 mer > 6 mer). The agreement between these two orderings suggests that the miRNA effects on mRNA levels captured by microarrays are tightly correlated with the fold protein down-regulation—resulting from the product of mRNA decay and translational effects—which is presumably the effect that is under selection.

Comparing data from a panel of luciferase reporters following miRNA transfection (Farh et al. 2005) to fold change values measured by microarray (Lim et al. 2005) for the corresponding endogenous mRNAs (Supplemental Table S3), we observed a significant Spearman rank correlation of 0.63, despite the obvious differences in UTR context and whatever experimental noise was present in these assays. (As expected, average protein-level repression was somewhat larger than repression at the mRNA level.) This observation, though based on a small sample of genes, suggests that for typical targets, effects of miRNAs at the mRNA and protein levels may be reasonably well correlated.

Data from the panel of luciferase reporters (Farh et al. 2005) could also be used to address the effects of different seed match types. We observed that those reporters that contained at least one 8 mer seed match were more strongly repressed than those that contained exclusively 7 mer seed matches ($P < 0.05$ by rank sum test). Further, among those reporters containing exclusively 7 mer seed matches, those with at least one M8 7 mer were more strongly repressed than those containing exclusively A1 7 mer ($P < 0.01$ by rank sum test). Thus, the hierarchy of seed match types observed in the mRNA array data appears to hold also when miRNA effects were assessed at the protein level.

Effects of seed matches located in regions other than the 3' UTR were either very modest (coding regions) or not detected (5' UTRs), and so were not further explored here (not shown).

Evidence for direct recognition of t1 adenosines by the silencing complex

Preferential conservation of adenosine residues at the t1 position adjacent to miRNA seed matches was reported previously, even for the minority of miRNAs that do not begin with U (and have no known paralogs that begin with U). This observation led to the hypothesis that t1A residues in target mRNAs can be recognized directly by the silencing complex, in a manner that does not require pairing to the m1 base of the miRNA (Lewis et al. 2005). To directly test this hypothesis, we turned to data from three siRNA transfection studies by Jackson and colleagues (Jackson et al. 2003, 2006) and Schwarz and colleagues (Schwarz et al. 2006). To distinguish between direct recognition of t1A and possible base-pairing to miRNA base m1, Figure 1F includes data only for siRNAs whose first base was not U, representing 33 of the 44 "effective" siRNAs in these studies (see Supplemental Material). Strikingly, we observed stronger mRNA down-regulation associated with A1 7 mer (which lack complementarity to base m1) than for M1 7 mer (which have a WC match to base m1) for these siRNAs (Fig. 1F, cf. solid blue curve, blue triangles and inset nLFC plot, $P < 10^{-15}$), supporting direct recognition of t1 adenosines by the silencing complex. In fact, no stronger down-regulation was observed for M1 7 mer than for 6 mer flanked by nonmatching bases other than A (Fig. 1F, solid green curve), suggesting that base-pairing between the m1 and t1 bases, if it occurs, does not contribute to targeting. Similarly, stronger down-regulation was observed for M8-A1 8 mer than for M8-M1 8 mer (Fig. 1F, cf. solid purple curve and purple triangles, $P < 10^{-13}$). Again, no stronger down-regulation was observed for M8-M1 8 mer than for M8 7 mer with nonmatching, non-A bases at position t1. Together, these observations strongly support the hypothesis that t1A residues adjacent to 6 mer or to M8 7 mer are recognized directly by a protein component of the silencing machinery in human cells, and that pairing to the m1 base, if it occurs, is of little or no consequence for targeting. This conclusion is consistent with recent structural studies of an Argonaute protein homolog in complex with dsRNA or an siRNA-like duplex, showing that the 5' nucleotide of the guide RNA (corresponding to the m1 base in an miRNA:mRNA or siRNA:mRNA duplex) is not base paired (Ma et al. 2005; Parker et al. 2005). The predictions of widely used miRNA target prediction algorithms that reward WC matching at position 1 (e.g., John et al. [2004]) should therefore be improved by instead rewarding t1A independent of miRNA complementarity.

Stronger down-regulation of mRNAs with conserved seed matches

The widespread conservation of 3' UTR seed matches since the divergence of rodents, carnivores, and primates (>50

million years ago [mya]) raises the issue of whether miRNA targets conserved over this time span commonly possess other determinants of miRNA targeting. To address this question, the distributions of LFCs for mRNAs containing exclusively nonconserved 3' UTR extended seed matches to miR-1 or miR-124 were compared with those of mRNAs containing conserved 3' UTR extended seed matches to these miRNAs (which will be greatly enriched for authentic conserved targets of these miRNAs). Notably, the mean nLFC for conserved extended seed matches was twice that seen for nonconserved extended seed matches for both miRNAs (Fig. 1C,E). This difference was significant ($P < 0.001$ for miR-124, $P < 0.01$ for miR-1, by rank sum test), when controlling for overall UTR conservation, seed match type and count, and initial mRNA expression level (Fig. 1C,E; Supplemental Table S1, with controls performed as illustrated in Supplemental Fig. S2). This observation suggests that authentic conserved miRNA targets contain additional targeting determinants that make them substantially more repressible by miRNA-programmed silencing complexes. An additional control for generic effects of 7 mer conservation was performed using data from siRNA studies (Jackson et al. 2003, 2006; Schwarz et al. 2006). Because the siRNAs used in the Jackson/Schwarz studies are unrelated in sequence to known endogenous mammalian miRNAs, any conservation of seed matches to these siRNAs is purely coincidental and unrelated to regulation by endogenous miRNAs. No significant difference in the distribution of LFC values was observed between mRNAs containing conserved rather than nonconserved extended seed matches to the transfected siRNAs, when expression, seed match type and count, and overall UTR conservation were controlled for as above (Fig. 1G; Supplemental Table S1). These observations imply that the increased repression observed for mRNAs containing conserved miRNA seed matches results from selection to enhance miRNA-directed repression in conserved targets relative to other genes.

Inducible inhibition of endogenous miRNA expression in mouse embryonic fibroblasts

The analyses described above rely on systems in which miRNAs or siRNAs are transfected into cells in which these RNAs are not naturally expressed. Although supported by independent analyses of UTR sequence conservation, the results are therefore subject to any potential differences between the activities of exogenous and endogenously expressed miRNAs, e.g., resulting from differences in incorporation into silencing complexes, if such differences exist. Therefore, it was of interest to ascertain whether the targeting rules observed above, e.g., the differences between seed match types and between conserved and nonconserved seed matches, apply also to regulation by endogenous miRNAs.

To study the activities of endogenously expressed mammalian miRNAs, we developed a conditional *Dicer* knockout system. Following Drosha processing in the nucleus, ~70 nt hairpin pre-miRNAs are exported to the cytoplasm for secondary processing to the mature ~22 nt miRNA by the RNase III enzyme Dicer (Kim 2005). Vertebrates express only a single Dicer gene, *Dicer1*, which is essential for development in both the mouse and the zebrafish (Bernstein et al. 2003; Wienholds et al. 2003). All vertebrate miRNAs appear to require processing by the protein product of this gene. Mice homozygous for a conditional null allele of *Dicer1* were generated using a tamoxifen-inducible promoter driving Cre recombinase (Danielian et al. 1998; Hayashi and McMahon 2002) and a conditional LacZ reporter (Soriano 1999; Supplemental Fig. S3). Cells were harvested from embryos at gestational day 16 and propagated in culture according to standard protocols to generate mouse embryonic fibroblasts (MEFs), which we refer to as conditional *Dicer* knockout (CDKO) MEFs.

Exposure of these cells to tamoxifen (ortho hydroxy tamoxifen; OHT) induces expression of Cre recombinase, resulting in a deletion that heritably inactivates the *Dicer1* locus. By staining the cells for LacZ, we established the minimum concentration and time required to induce Cre expression and inactivate the *Dicer1* locus in essentially all MEFs (Fig. 2A; Supplemental Fig. S4). Within 24–48 h post-induction of Cre, proliferation slowed (Fig. 2B). Visual inspection of the cells suggested minimal levels of apoptosis, and a modest level of apoptosis not substantially greater than for control cells was confirmed by Annexin V staining (Supplemental Fig. S5). In these respects, the *Dicer*-deficient MEFs bore some similarities to *Dicer*-deficient T cells, which were reported to have reduced proliferation but only modestly increased levels of apoptosis (Muljo et al. 2005).

By day 4 post-induction, Western analysis with Dicer antibodies detected an approximately threefold decrease in protein levels (Fig. 2C). Day 4 post-induction represents an average of ~2 d post-inactivation of the *Dicer* locus (Supplemental Fig. S4). At this stage, total RNA was collected from the MEFs for microarray analysis. Untreated CDKO MEFs and MEFs derived from wild-type mice (untreated or subjected to OHT treatment) were used as controls, and each experiment was repeated twice.

Mature miRNAs were profiled in the MEFs using a spotted oligonucleotide miRNA microarray with standard miRNA probes present in quadruplicate. Using this array, expression of 99 miRNAs was detected at more than two standard deviations (SD) above background in seven of the eight miRNA arrays (Supplemental Table S4). Among the most highly expressed miRNAs were members of the let-7 family, miR-1, miR-124, miR-15a, miR-17–5p, and several other miRNAs previously detected in mouse embryos (Thomson et al. 2004). Expression of several of the array-detected miRNAs was also confirmed by Northern analysis

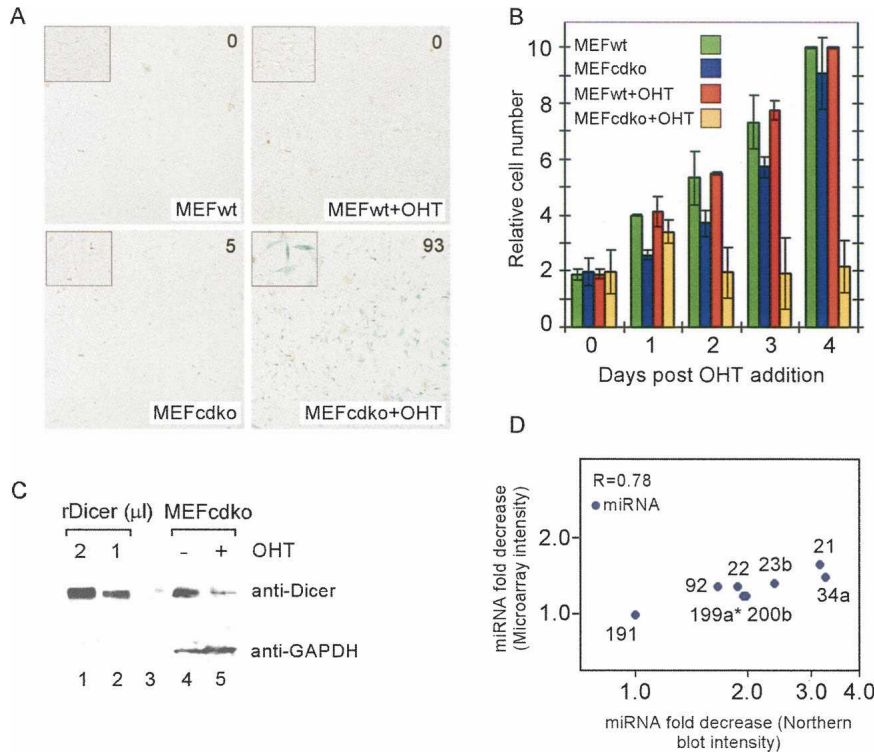


FIGURE 2. Characterization of conditional Dicer knockout (CDKO) mouse embryonic fibroblasts (MEFs). (A) Wild-type (wt) and CDKO MEFs are shown, untreated (*left panels*) or 4 d after ortho hydroxy tamoxifen (OHT) treatment (*right panels*). Cells were stained for LacZ, and the percentage of LacZ-positive cells is shown (*upper right*). (B) Proliferation of wild-type and CDKO MEFs, untreated or following addition of OHT. Error bars represent standard deviation of three independent counts. (C) Western analysis of MEFcdko cells, untreated (lane 4) or after OHT addition (lane 5), showing a three- to fourfold reduction in Dicer protein levels following OHT addition. GAPDH is a loading control. Westerns using different concentrations of recombinant Dicer protein are shown as a positive control (lanes 1,2). (D) Microarray hybridization intensity change and expression level change measured by Northern analysis (log scale, both axes) for eight miRNAs (miR-21, miR-22, miR-23b, miR-34a, miR-92, miR-191, miR-199a*, and miR-200b).

(Supplemental Fig. S6). Spiked control RNA and second-channel reference RNAs were used to enable comparison of miRNA array data between control and knockout cells. Levels of most miRNAs decreased following Cre induction/*Dicer* knockout (Supplemental Fig. S6). The fold change in microarray hybridization intensity and in expression measured by Northern analysis were correlated, with the array intensity change consistently lower than the fold change measured by Northern (Fig. 2D). The expression of most miRNAs tested was reduced by approximately twofold by Northern, consistent with the about threefold reduction in Dicer protein levels and the notion that miRNAs have fairly long, but not infinite, half-lives. Variability in the fold changes of different miRNAs was observed, which could reflect differences in miRNA stability, in pre-miRNA processing efficiency in the presence of limiting amounts of Dicer protein, or perhaps changes in miRNA transcription or processing in response to reduced Dicer protein or miRNA levels.

Targeting rules inferred from derepression of mRNAs following Dicer knockout

The expression of mRNAs was profiled in control and CDKO MEFs using Affymetrix Mouse Genome 430_2 arrays. This CDKO system has certain advantages over transfection-based systems for studies of miRNA function, including the potential to study the activities of endogenous miRNAs expressed at natural levels. The CDKO system generates a modest and gradual ebbing of miRNA levels, as opposed to miRNA/siRNA transfection, which effectively floods the cell with a specific miRNA/siRNA species. Although it requires administration of tamoxifen, use of targeted gene knockout to reduce Dicer levels, rather than RNAi, has the advantages of permanently inactivating the *Dicer1* gene and avoiding addition of exogenous siRNAs, which could exert “off-target” effects like those seen in Figure 1F, complicating analysis of mRNA expression changes.

Analysis of miRNA effects on mRNAs following *Dicer* knockout is necessarily more complex than for miRNA/siRNA transfection experiments because loss of *Dicer* results in decreases in the levels of dozens of miRNAs at once. One straightforward approach uses the median LFC (where LFC is defined for *Dicer* knockout

experiments as the base 2 log of hybridization intensity in treated CDKO cells over the intensity in control cells) over all mRNAs containing conserved extended seed matches to a particular miRNA, analyzing each miRNA independently. (In this analysis, untreated CDKO cells, and treated and untreated wild-type MEFs, served as controls; hybridization intensity averaged over these three types provided a control value for calculating LFC.) Applying this approach to the set of 99 miRNAs (representing 80 unique seed sequences) detected by miRNA array analysis yielded a distribution of median LFCs that was significantly shifted toward higher values than for control (random) sets of mRNAs or for mRNAs containing conserved extended seed matches to the miRNAs (representing 50 unique seeds) whose expression was not detected above background (Fig. 3A). (Median LFC was used in this analysis rather than mean because of its greater stability in the face of noise for the sometimes very small sets of conserved targets being analyzed.) This observation suggested that many of the changes in mRNA

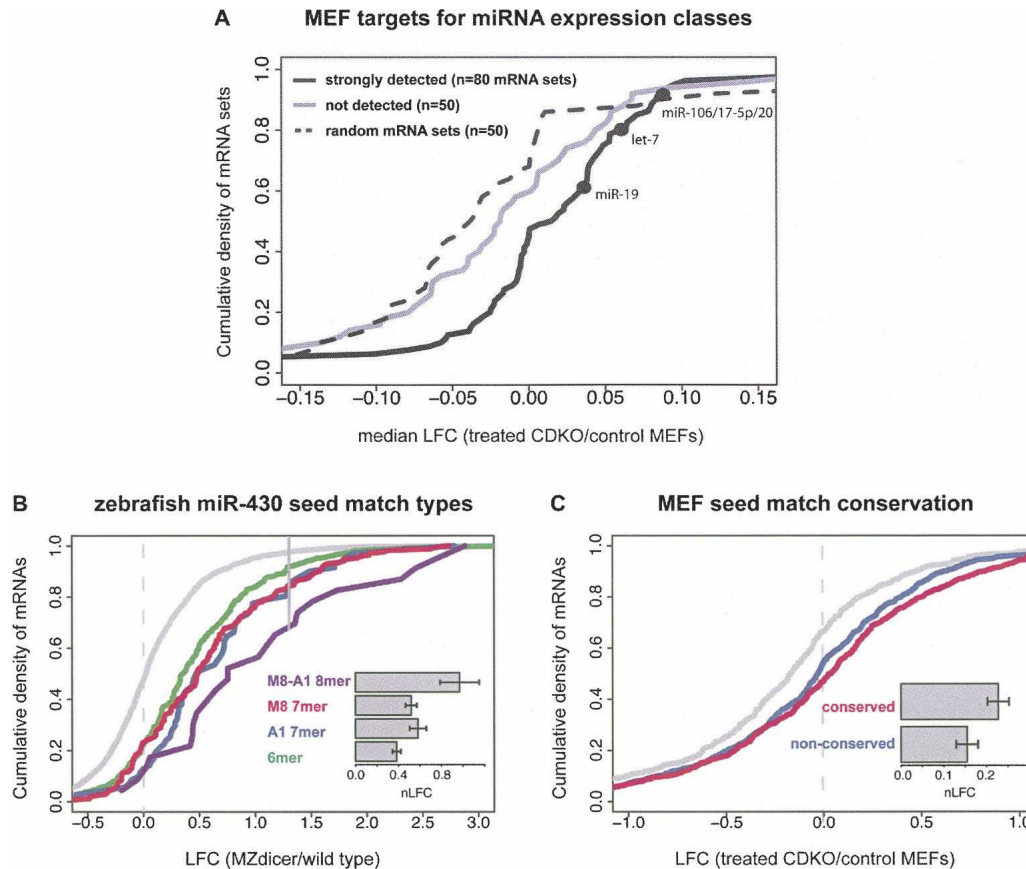


FIGURE 3. mRNA derepression following *Dicer* knockout varies with seed match type and conservation status. (A) CDFs of median LFC for three classes of mRNA sets. The expression classes were: (1) mRNA sets containing extended seed matches to the 80 miRNA families whose expression was detected above background by microarray (black curve, selected miRNA family names shown); (2) mRNA sets containing extended seed matches to the 50 miRNA families that were not detectably expressed (gray curve); (3) randomly selected mRNA sets (dotted line). Distributions of detected and nondetected sets are significantly different ($P < 0.01$ by rank sum test), but distributions of nondetected and random mRNA sets are not. (B) CDFs of LFCs for mRNAs containing the indicated miR-430 seed match types—or no miR-430 seed matches (gray curve)—for MZdicer zebrafish embryo data (Giraldez et al. 2006); only mRNAs with exactly one miR-430 seed match were included (seed match type mRNA sets are mutually exclusive). All seed match type LFC distributions differed significantly from the no-seed-match class ($P < 10^{-7}$ by rank sum test), and the pooled extended seed match types differed from the 6mer class ($P < 0.01$ by rank sum test). (Solid vertical gray line) The LFC below which 97.5% of the no-seed-match mRNA set falls. (Inset bar plot) LFC values for each seed match type, with error bars indicating standard error. (C) CDFs of LFCs for mRNAs containing conserved (red) or exclusively nonconserved (blue) extended seed matches, or no seed matches (gray) to the set of 31 “responsive” miRNAs in the CDKO MEF experiment (Supplemental Material). Seed match count, overall UTR conservation, and mRNA expression level were controlled between the sets. The distributions of mRNA with conserved and nonconserved seed matches are significantly different ($P < 0.05$ by rank sum test). See Supplemental Table S6 for additional statistics.

expression observed in this experiment resulted from derepression of genes whose mRNA levels were specifically repressed by miRNA-programmed silencing complexes prior to knockout of *Dicer*. The derepression of mRNAs containing conserved seed matches to many expressed miRNAs following about twofold reduction in miRNA expression suggested that, at least in this system, many miRNAs are not expressed at saturating levels relative to their targets. mRNAs with seed matches to miRNAs not detected by microarray were shifted to an insignificant degree toward higher values relative to random mRNA sets of the same size (Fig. 3A). A list of the mRNAs whose expression changed significantly following *Dicer* knockout is provided in Supplemental Table S5.

Three previous studies have analyzed the effects of inhibition of miRNA processing enzymes on global mRNA expression, two using RNAi knockdown and one using targeted gene knockout. Rehwinkel and colleagues (2005) found that the set of mRNAs derepressed following RNAi knockdown of Drosha in *Drosophila* cells were enriched for miRNA targets predicted using the algorithm of Stark and colleagues (Brennecke et al. 2005; Stark et al. 2005), which is based on rules for targeting that (like TargetScanS) emphasize WC pairing to miRNA bases 2–8. Thus, the Rehwinkel study supported the idea that endogenous miRNAs commonly regulate their targets at the mRNA level through mechanisms involving seed match pairing. Recently, Schmitter and colleagues (2006) studied global

changes in mRNA expression following RNAi knockdown of Dicer and Argonaute in cultured human cells. They observed up-regulation/derepression of overlapping sets of transcripts 2 and 6 d after knockdown of Dicer and 2 d after knockdown of Ago2, and again found enrichment for miRNA seed matches in the UTRs of derepressed mRNAs. Very modest effects were observed following knockdown of other Argonaute family genes. In the third study, Giraldez and colleagues (2006) used sophisticated gene knockout techniques to generate “MZdicer” zebrafish embryos deficient in both maternal and zygotic Dicer activity. The set of mRNAs whose expression was significantly increased in MZdicer embryos relative to wild type were enriched for seed matches to miRNAs of the miR-430 family, the most abundantly expressed miRNA family during early zebrafish development, representing ~50% of miRNAs cloned. These and related studies of MZdicer embryos convincingly demonstrated that miRNAs promote accelerated decay of targeted mRNAs in vivo.

The predominance of a single miRNA family in zebrafish embryos made this system suitable for assessing the effects of seed match type on mRNA regulation by endogenous miRNAs. Analyzing mRNAs containing different miR-430 seed match types, M8-A1 8 mer were associated with the strongest derepression, with a mean nLFC value almost twice that seen for M8 or A1 7 mer. The mean nLFC values for the two 7 mer types were similar to each other and higher than for 6 mer (Fig. 3B; Supplemental Table S6). Thus, the ordering of seed match types and the relative magnitudes of 8 mer versus 7 mer seed match effects paralleled those seen for transfected miRNAs/siRNAs (Fig. 1), indicating that the seed match hierarchy inferred from transfection data also holds for regulation by endogenous vertebrate miRNAs. For zebrafish miR-430, 6 mer had a higher nLFC value relative to 7 mer and 8 mer than in the mammalian miRNA/siRNA transfection experiments (Supplemental Tables S1, S2. The 6 mer nLFC value may be magnified by effects of other miR-430 superfamily miRNAs. Analogous seed match type comparisons were not attempted using the CDKO MEF data because most mRNAs contained a mixture of different seed match types, often to several different expressed miRNAs, so too few mRNAs containing only a single seed match type were available for effective analysis; the effect of the t1 position was not addressed in the CDKO MEF data for the same reason.

The repression of mRNAs containing conserved rather than nonconserved seed matches could be most effectively analyzed in the CDKO MEF data. In the zebrafish MZdicer data, the set of mRNAs containing conserved miR-430 seed matches was relatively small, and significant derepression relative to nonconserved seed matches was not observed (Supplemental Fig. S7). Seed match conservation is more difficult to assess in fish, as large differences in intergenic region sizes among the fishes yield less reliable genomic alignments, and classification based on seed match presence

is limited by the relatively sparse 3' UTR annotations available for orthologous fish genes. In the MEF CDKO data, far larger mRNA sets were available for this analysis. In these data, the mean nLFC for mRNAs containing extended seed matches conserved between human, mouse, rat, and dog (HMRD) to a set of 31 “responsive” miRNAs (see Supplemental Material) was ~50% higher than that for mRNAs containing exclusively nonconserved seed matches ($P < 0.05$), controlling for overall UTR conservation, mRNA expression, and seed match count (Fig. 3C). This analysis, indicating that mRNAs containing conserved extended seed matches are preferentially repressed by endogenous miRNAs, further supports the idea that conserved miRNA targets possess additional targeting determinants that contribute to their repression by the miRNAs that naturally target them.

Fold change increases multiplicatively with seed match count for both endogenous miRNAs and exogenous miRNAs/siRNAs

Using luciferase or other reporter assays, increases in the magnitude of miRNA-directed repression have typically been observed when the number of 3' UTR seed matches is increased (Doench and Sharp 2004; Vella et al. 2004; Pillai et al. 2005), but the quantitative relationship between seed match count and repression has not been established using large sets of targets. Grouping mRNAs based on the number of extended seed matches to transfected miRNAs in the Lim datasets analyzed in Figure 1, mean nLFC increased approximately linearly as extended seed match count increased from one to three for both miRNAs (Fig. 4A). The dose–response relationship between extended seed match count and mRNA nLFC further supports the idea that seed matches are the primary determinant of mRNA down-regulation by miRNAs. Although the Lim (Lim et al. 2005), Jackson (Jackson et al. 2003, 2006), and Schwarz (Schwarz et al. 2006) experiments used identical protocols and concentrations of transfected RNAs, the magnitude of the mean nLFC per seed match for the two miRNAs was roughly twice that seen on average for the siRNAs (not shown), suggesting some degree of optimization of target and perhaps miRNA sequences for efficient repression.

For endogenous zebrafish miR-430, a roughly linear relationship was also seen between the mean nLFC values of mRNAs containing one to three extended seed matches in the MZdicer experiment (Fig. 4B). Here, as for the miRNA transfection data, too few mRNAs were available to extend the analysis beyond three seed matches. Because of the greater diversity of miRNAs affected, the CDKO MEF experiment allowed analysis of the effects of a larger range of seed match counts on mRNA repression. For the set of 31 “responsive” miRNAs used above, an approximately linear relationship was again observed between mean nLFC and the count of conserved extended seed matches (Fig. 4C). This relationship

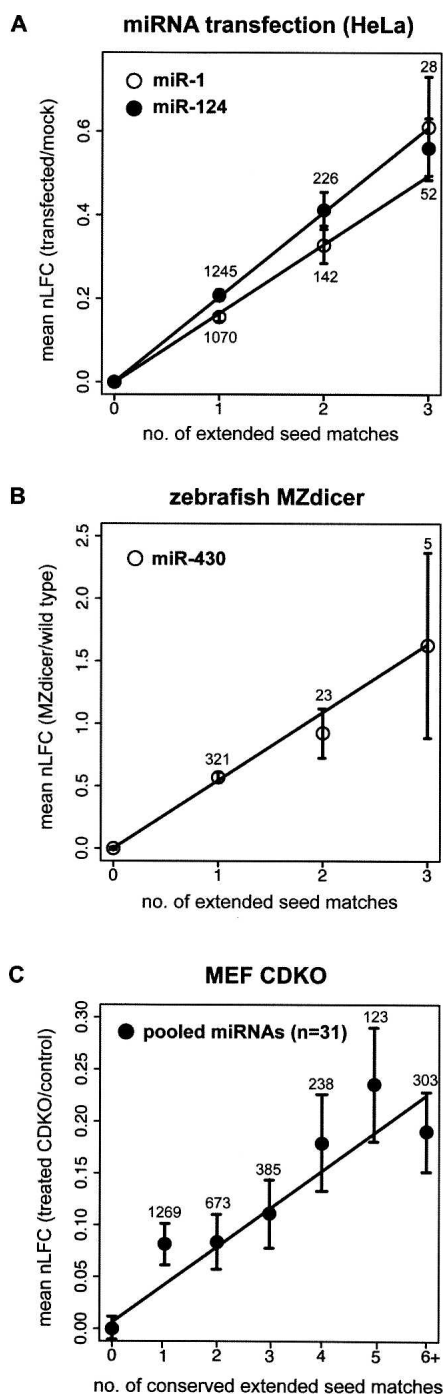


FIGURE 4. mRNA fold change increases multiplicatively with extended seed match count. (A) For miR-1 (open circles) and miR-124 (solid black circles), the total number of extended seed matches was enumerated for each mRNA, and the mean nLFCs in the Lim transfection experiments were determined for sets of mRNAs grouped by seed match count (set sizes indicated above or below points). (Solid lines) Least squares fit for the whole data set. Error bars correspond to standard error. For each of these plots, the proportions of different seed match types for different seed match counts remained fairly constant. (B) Same as A for miR-430 extended seed match counts following *Dicer* knockout in zebrafish embryos (Giraldez et al. 2006). (C) Same as A for conserved extended seed match counts to 31 “responsive” mouse miRNAs (see Supplemental Material) in CDKO MEFs.

held for conserved extended seed match counts from one up to at least five, suggesting that miRNA regulation is tunable over a very broad range. The essentially linear relationship between seed match count and the logarithm of the fold change observed in Figure 4 indicates that each seed match contributes multiplicatively to fold change in mRNA level. Multiplicative effects could be explained if RISCs act independently and each has a chance of interaction with a single effector site on the mRNA—such as the 5′ cap (Kiriakidou et al. 2007)—is required for RISC-mediated repression.

Evidence for A or U at position t9 as a targeting determinant

To search for additional targeting determinants, we analyzed the effects on mRNA down-regulation of nucleotides present at different target positions in the vicinity of seed matches. The Jackson/Schwarz siRNA transfection data were most suitable for this analysis because of the large number of independent siRNAs and array measurements. Although modest increases in down-regulation were associated with the presence of adenosine and/or uridine at a few other positions (not shown), the most pronounced effect was observed for the presence of A or U at position t9. Those mRNAs that had a t9W base (using the abbreviation W = A or U) were down-regulated to a greater degree following siRNA transfection than those with a t9S (S = C or G) residue (Fig. 5A). This effect was pronounced for M8 7 mer ($P < 10^{-6}$) and 8 mer ($P < 10^{-2}$) seed matches, with a marginal effect observed for A1 7 mer (not shown). The effect remained highly significant whether controlling for UTR CG content (as in Fig. 5) or not. The increased repression of extended seed matches containing t9W was observed independent of whether the base m9 of the siRNA was a match to t9 or not (Fig. 5B). No significant effect of t9 matching was observed, though the t9W match set for 8 mer in particular was quite small ($n = 141$), limiting statistical power to detect any effect that might exist. Taken together, these observations suggest that the presence of a t9W base adjacent to an extended seed match contributes to typical seed match targeting interactions, independent of pairing to m9.

The targeting role of t9 inferred from siRNA transfection data was corroborated by analyses of seed matches to miRNAs. Examining the composition of the t9 position for seed matches to a large set of conserved miRNAs in mammalian UTRs, we observed an increased frequency of t9W residues adjacent to conserved seed matches relative to control sets of nonconserved seed matches in UTRs matched for UTR CG content (Fig. 5C). Signal:noise values for this miRNA set, calculated with control oligonucleotides matched for both count and CG content, were significantly higher for t9W compared with t9S seed matches (Fig. 5D). Consistent with the siRNA analyses, the difference appeared independent of the base at position

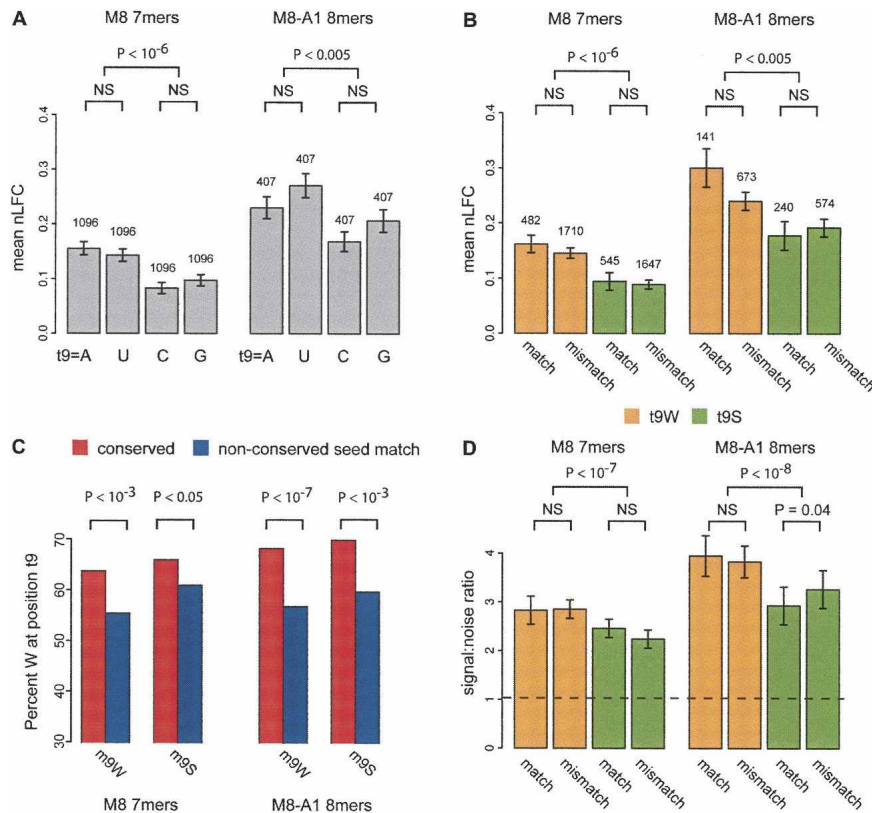


FIGURE 5. Increased down-regulation of mRNAs with adenosine or uridine at position t9. (A) Mean nLFC for mRNAs containing the indicated nucleotide at position t9 flanking siRNA M8 7 mer and M8-A1 8 mer (rank sum test P -values; NS = not significant at P -value cutoff 0.05). Error bars indicate standard error, and the numbers of mRNAs are indicated above the bars. Each mRNA contained exactly one seed match to any given siRNA (i.e., t9 sets are mutually exclusive), and mRNAs in each of the four t9 sets were controlled for 3' UTR GC content. Other variables, such as mRNA expression, 3' UTR conservation, or m9 composition, did not differ significantly between t9 sets. (B) Same as A, but reclassifying the controlled mRNA sets by whether the t9 base pairs with the siRNA m9 (match) or not (mismatch). (C) Enrichment of t9W nucleotides flanking conserved versus nonconserved miRNA M8 7 mer and M8-A1 8 mer in human 3' UTRs (χ^2 test P -values). The miRNA set consisted of conserved human miRNAs used for target prediction by Lewis et al. (2005) after removal of miRNAs with common m2–m8 seed regions but different m9 nucleotides, and pairs of miRNAs in the same superfamily. The nonconserved seed matches were sampled to match the seed match type, miRNA, and overall UTR CG content of the conserved set. (D) Mean signal:noise ratios for M8 7 mer and M8-A1 8 mer with t9W or t9S in match and mismatch configurations based on cohorts of control oligonucleotides (Lewis et al. 2005) matched for both count and exact CG content (error bars indicate standard deviation based on 14 control cohorts). (Dashed line) Baseline S:N value of 1. P -values based on Wilcoxon rank sum tests between indicated sets (NS = not significant at P -value cutoff 0.05).

m9. These observations extend previous observations of increased conservation of t9A residues, independent of complementarity to miRNA base m9 (Lewis et al. 2005), and support a role for t9W in miRNA targeting in vivo.

Increased conservation and AU content flanking siRNA seed matches associated with increased mRNA repression

Increased sequence conservation across mammals is observed in the vicinity of conserved miRNA seed matches

relative to those that are not conserved, even when overall UTR conservation is controlled for (Fig. 6A). The increase in conservation extends to 50 bases 3' and 5' of the seed match and beyond; similar patterns of increased local conservation are associated with other conserved UTR motifs (not shown; Lewis et al. 2005). One possible explanation is that the sequence context flanking authentic conserved target sites is enriched for features—e.g., protein binding sites or RNA structural properties—that, directly or indirectly, enhance the effectiveness of miRNA targeting. To explore these issues, we returned to the siRNA data and compared siRNA-directed mRNA repression between mRNAs having different levels of sequence conservation in the 50 bases 5' and 3' of the seed match, in sets matched for overall UTR conservation, expression level, seed match type, and local and global AU composition (Supplemental Fig. S9). Strikingly, substantially stronger mRNA repression was observed for siRNA seed matches with high conservation in the 50 bases upstream of the seed match relative to seed matches with low conservation in this region (mean nLFC = 0.16 and 0.11, respectively, $P < 10^{-4}$ by rank sum test; Fig. 6C). A similar effect of conservation in the downstream 50 bases was observed (mean nLFC = 0.16 versus 0.12, $P < 0.05$; Fig. 6C). These results were not affected by whether or not the siRNA seed match itself was conserved (not shown).

In vertebrate genomes, AU-rich sequences have higher levels of average sequence conservation than CG-rich sequences, at least in part because of the high mutation rate of CpG dinucleotides (Hwang and Green 2004), so it was important to determine whether the increased repression associated with conserved flanking regions resulted from an effect of base composition. The 30–50 bases just upstream of and downstream from conserved miRNA seed matches are indeed biased toward higher AU composition (Fig. 6B), suggesting that local AU composition itself might contribute to targeting. However, the effect of sequence conservation on mRNA repression was significant whether AU content was controlled for (as in Fig. 6C) or not. Conversely, significantly increased repression was observed for

siRNA seed matches flanked by high AU content in the 50 bases either 3' or 5' of the seed match relative to those with low AU content in these regions (not shown). When this analysis was controlled for the effects of sequence conservation, overall UTR AU content, expression level, and seed match type, seed matches with high AU content in the 3' 50 bases had significantly increased down-regulation relative to those with low AU content in this region (mean nLFC = 0.17 and 0.10, respectively, $P < 10^{-4}$ by rank sum test), but the effects of 5' AU content were no longer significant (Fig. 6D). Thus, conservation immediately 5' and 3' and AU content 3' of seed matches are independently associated with increased mRNA down-regulation by siRNAs, and 5' AU content may also enhance down-regulation. Choice of 50 base pairs (bp) as the size of the region to analyze was based on the distributions shown in Figure 6, A and B; however, the magnitude of the effects shown in Figure 6, C and D, were little changed when the size of the analyzed region was expanded or reduced by 20 bp.

In terms of the magnitude of nLFC change, each of these variables contributed to targeting to at least the same extent as the identity of the t9 base (e.g., Fig. 5B, compare mean nLFC value differences). The conservation of more distal 50-bp windows (e.g., bases 151–200 downstream from the seed match) was associated with increased repression when the controls on conservation in other regions were relaxed, but disappeared when either overall UTR conservation or conservation in the 50 bases immediately 3' of the seed match were controlled for (not shown). This observation suggests that effects for such distal windows observed in the uncontrolled analysis derive from the (fairly strong) positive correlation between conservation in nearby UTR regions and that the proximal 50 bp is of central importance. Similar results were obtained for AU content (not shown), again supporting the importance of the regions immediately adjacent to the seed match.

There are at least two plausible ways in which high AU content might increase effectiveness of adjacent seed matches. AU-rich sequences could be recognized directly by a component of the RISC or an auxiliary activating factor; a number of protein families are known that have affinity for A-rich, U-rich, or AU-rich RNA sequences (Barreau et al. 2005), and functional connections between AU-rich binding factors and miRNA regulation have been reported (Bhattacharyya et al. 2006; Vasudevan and Steitz 2007). Alternatively, AU-richness might enhance targeting by reducing the tendency for formation of stable RNA secondary structures that could interfere with RISC binding. Previously, it has been reported that predicted local folding of the mRNA in the vicinity of seed matches is a negative predictor of miRNA targeting (e.g., Robins et al. 2005). Consistently, we have observed that seed matches in regions of lower predicted thermodynamic stability using standard algorithms are associated with increased mRNA-level repression, but we have found that this effect disappears

when the AU content of the region is controlled for (not shown). Thus, the effect of AU content we observe may contribute to targeting by reducing the potential for inhibitory mRNA structures, but other effects of AU-richness are also consistent with the data.

Since siRNAs are not naturally expressed, siRNA seed matches tend to be distributed essentially randomly in UTRs and do not experience selection related to targeting.

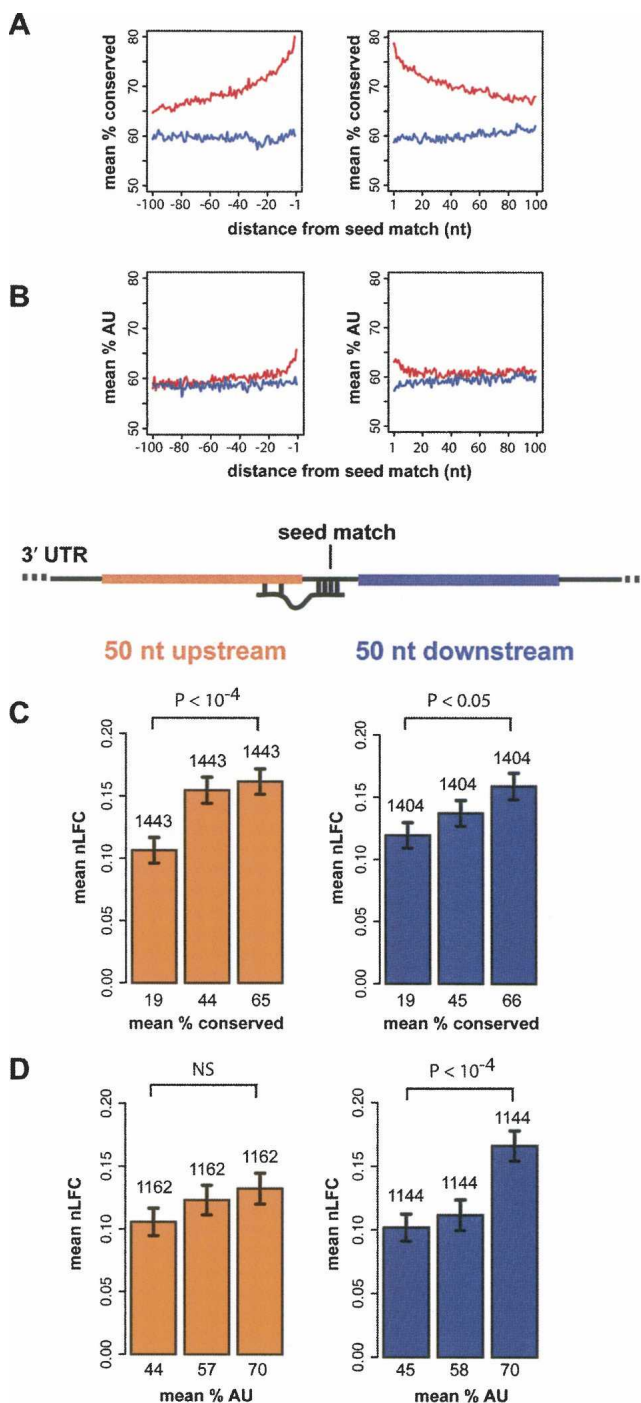


FIGURE 6. (Legend on next page)

The magnitude of the effect of local conservation on siRNA seed match efficacy (Fig. 6C) was similar in magnitude when siRNA seed matches falling within 100 bp of conserved miRNA seed matches were included (as in Fig. 6C) or excluded (not shown). Therefore, the enhanced repression observed for siRNA seed matches that occur in regions of high local conservation (Fig. 6C) is likely to represent a side effect of mRNA features that are conserved for reasons unrelated to targeting by the RISC, such as RNA binding sites for factors involved in other aspects of mRNA biology (mRNA processing, stability, localization, translation, etc.). For example, presence of proteins bound to sites nearby the seed match might increase accessibility to RISC by interfering with formation of local RNA secondary structures. Alternatively, the long coevolution of RISC and non-RISC factors binding nearby in mRNAs may have engendered more direct interactions, with common mRNA binding factors functioning to facilitate or stabilize RISC binding to nearby seed matches. Two very recent studies observed that, with the exception of the first ~ 20 bases of the mRNA, the density of conserved seed matches increases as proximity to the stop codon or poly-A tail increases, (Gaidatzis et al. 2007; Majoros and Ohler 2007); however, signal:noise in these regions was not assessed. Another recent study reported synergy between nearby seed matches located 13–35 bases apart, providing support for the idea that RISC activity is modulated by the presence of proteins or complexes in nearby flanking regions (Saetrom et al. 2007). Given the complexities of conservation effects on miRNA target analysis, including the phenomenon that mRNAs with lower overall UTR conservation have substantially

higher signal:noise values for conserved miRNAs (Lewis et al. 2005), analyses of the effects of local conservation and AU content on miRNA targeting were deferred pending availability of additional relevant data. Further investigation is clearly needed to understand the mechanisms underlying these phenomena.

Perspectives and applications to target/off-target prediction

Here, we have described specific rules for miRNA/siRNA targeting, including a hierarchy of seed match types, the multiplicative effects of multiple seed matches, and targeting determinants outside of the seed match, including t9W and local conservation and AU content effects. All of these rules and determinants were supported for exogenous siRNAs and/or miRNAs by direct effects on mRNA levels, and for endogenous miRNAs through direct effects and/or comparative genomic data. Thus, in the available data these rules and determinants appear to be applicable to targeting by exogenous siRNA/miRNAs and endogenous miRNAs.

Current miRNA target predictions have relied very heavily on seed match conservation, ignoring potential species-specific miRNA targeting. However, we observe that nonconserved 8 mer seed matches on average exhibit stronger repression than conserved 7 mer (Supplemental Fig. S8). For studies of miRNA function, it would be extremely useful to be able to predict which mRNAs will experience the strongest repression to facilitate choice of targets for in-depth characterization, and similar considerations apply to the design and interpretation of experiments involving siRNAs. The new rules and determinants identified here can be combined to produce an expected nLFC score for a seed match by summing the mean nLFC of the seed match type (Fig. 1F) plus the residual contribution to mean nLFC of the t9 base (Fig. 5A), and of flanking AU content and conservation (Fig. 6). Following the results of Figure 4, if multiple seed matches are present, their scores are added (scoring details are given in Supplemental Material). Because this score can be used to rank potential siRNA off-target effects, and to generate an integrated ranking of conserved and nonconserved miRNA targets, we call it the TargetRank score. Application of TargetRank scoring to sets of mRNAs, each with a single 7 mer seed match to transfected siRNAs (with parameters derived from a held-out set of siRNA transfections) yielded a dramatic separation between the LFC distributions of the bottom 20% and top 20% of TargetRanked mRNAs in the test set (Fig. 7A), with mean nLFC increasing from 0.07 to 0.26, and the fraction of significantly down-regulated mRNAs increasing from 5% to 20% (Supplemental Table S7). Because all of the mRNAs in this analysis contained single 7 mer seed matches, the separation of the two distributions results from the additional determinants identified in this study (t9W, flanking conservation, and

FIGURE 6. siRNA-directed mRNA repression is enhanced by local conservation and AU content. (A) Mean percent conservation at UTR positions within 100 bp 5' (left) and 3' (right) of conserved (red) or nonconserved (blue) extended seed matches to the set of conserved vertebrate miRNAs used in Lewis et al. (2005); overall UTR conservation was controlled for in the comparison of conserved and non-conserved seed matches. The average conservation differs significantly for the 100 bases 5' and 3' of the seed match ($P < 10^{-200}$ by rank sum test for both). (B) Mean AU content (sets controlled for UTR AU content); average AU content is significantly different both 5' and 3' of the seed match ($P < 10^{-18}$, $P < 10^{-30}$), respectively. (C) Mean nLFC for three equal-sized mRNA sets binned by percent conserved positions (in HMRD) in the 50-nt region immediately 5' (orange) or 3' (purple) of siRNA seed matches (5' region ends at position t10; 3' region begins one base 3' of position t1) for mRNAs containing single extended seed match to the relevant siRNA. Bars indicate standard error of the mean. Set size and mean percent conservation for each set are reported above and below each bar, respectively. *P*-values are for two-sided rank sum tests between the first and third bins. For both 5' and 3' conservation, the three bins have been sampled such that their distributions of overall UTR conservation, 5' (or 3') AU content, overall UTR AU content, seed match type, and initial expression level are not significantly different ($P \geq 0.05$) (Supplemental Fig. S9). (D) Same as C, but with UTRs binned by AU content in the same 50-nt regions. Bins are sampled to control for UTR AU content, 5' (or 3') conservation, overall UTR conservation, seed match type, and initial expression level (NS = not significant at *P*-value cutoff 0.05).

AU content), demonstrating their combined utility for siRNA off-target prediction.

Applying the same scoring system to expression data for Th1 thymocytes from mir-155 knockout mouse versus wild-type cells (Rodriguez et al. 2007) also yielded a strong degree of separation. When scoring mRNAs containing exclusively nonconserved single 7 mer seed matches, mean nLFC increased from 0.01 for the bottom 20% of Target-Ranked mRNAs (NS relative to no-seed-match mRNAs) to 0.09 for the top 20% of mRNAs ($P < 0.01$ relative to no-seed-match or bottom 20% mRNAs) (Fig. 7B; Supplemental Table S7), demonstrating the applicability of these additional determinants to regulation by endogenous miRNAs, and suggesting an approach for identification of important species-specific miRNA targets. In practice, scoring of 6 mer, 7 mer, and 8 mer seed matches yields a broader range of TargetRank scores and a correspondingly greater

separation between the distributions of higher and lower ranked genes. Grouping siRNA and nonconserved miR-155 targets into five bins of TargetRank scores demonstrated a strong and approximately log-linear relationship between mRNA down-regulation and TargetRank score (Fig. 7C,D). The relative ranking given by TargetRank is probably more useful than the score itself, since the overall magnitude of miRNA- or siRNA-associated repression will vary in different systems, as seen above.

Unlike purely conservation-based methods, TargetRank scoring of the expressed mRNAs in a cell type yields an integrated ranking of conserved and nonconserved targets, and should therefore be particularly helpful in identifying important species- or clade-specific miRNA targeting relationships. These results should also aid in interpretation of RNAi phenotypes and in prediction of the miRNA targeting effects of mutations and polymorphisms in human genes.

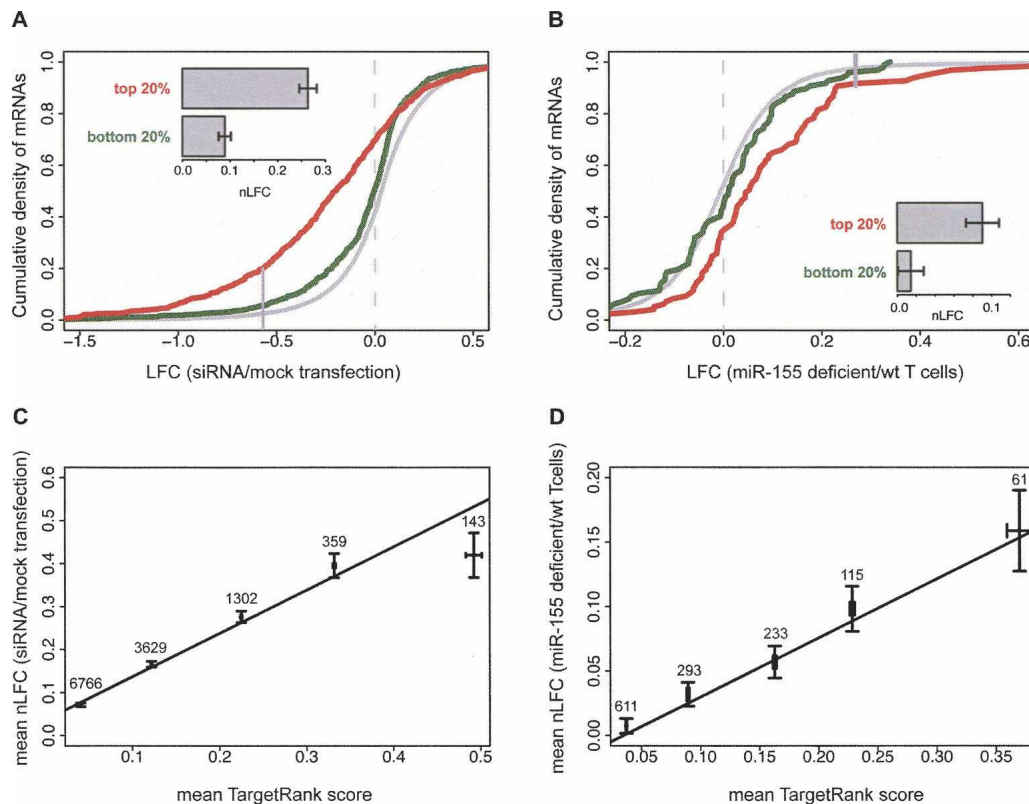


FIGURE 7. TargetRank scoring separates strongly and weakly down-regulated mRNAs. (A) CDFs of LFCs (as in Fig. 1) for the top 20% (red) and bottom 20% (green) of mRNAs to the relevant siRNA in the test set of eight randomly chosen siRNA transfection experiments ranked by TargetRank score. Only expressed mRNAs containing exactly one 7 mer seed match and no other seed matches of any type were used. For reference, the CDF for mRNAs lacking seed matches to the relevant siRNAs is shown (gray). (Solid vertical gray line) The LFC above which 97.5% of the no-seed-match mRNA set falls. (Inset bar plot) See Fig. 1. See Supplemental Table S7 for additional statistics. (B) Same as A, but for miR-155 knockout T cell data from Rodriguez et al. (2007). (C) All mRNAs containing seed matches (of any type or count) to the relevant siRNA in the eight test siRNA transfection experiments (same sets as in A) were scored using TargetRank. The mean TargetRank score and mean nLFC are plotted with standard error bars for mRNAs sets binned by TargetRank score (mRNA set sizes indicated above points). Line corresponds to least squares fit for entire data set (ANOVA $P < 10^{-100}$); $r = 0.23$ (Pearson correlation). (D) Same as C, but for miR-155 knockout T cell data from (Rodriguez et al. 2007). Line corresponds to least squares fit for entire data set (ANOVA $P < 10^{-18}$); $r = 0.24$ (Pearson correlation).

MATERIALS AND METHODS

Conditional Dicer knockout mice and MEFs

Male mice carrying one copy of the pCAGGCre-ER allele (Hayashi and McMahon 2002) and one *Dicer* floxed allele (Harfe et al. 2005) were crossed to *Dicer* floxed/floxed females harboring also a LacZ reporter (R26R) for detection of Cre activity (Soriano 1999). Timed-pregnant females were sacrificed at embryonic day 16 and embryos were dissected and dissociated to generate mouse embryonic fibroblast (MEF) primary culture (following Abbondanzo et al. [1993]). After 72 h of incubation, cells were frozen in aliquots. Mice were housed and handled in accordance with protocols approved by the Institutional Animal Care and Use Committee of Harvard Medical School.

Cell culture and treatments

MEFs were thawed prior to experiments, grown in DMEM supplemented with 10% FCS, penicillin/streptomycin and glutamine, split once, and induced for loss of functional *Dicer* by addition of 4-ortho-hydroxy Tamoxifen (0.5 μ M; OHT; Sigma). Following 4 d (and daily change of media and drug), total RNA and total protein were extracted. Control MEFs derived from wild-type mice were subjected to the same treatments.

RNA extraction

Total RNA was extracted using TRIzol reagent (Sigma), and RNA quality was measured using an Agilent Bioanalyzer.

MEF miRNA microarray analysis

MicroRNA microarrays were printed using a Cartesian PixSys 5500 Arrayer on epoxy slides (Corning) using Ambion's miRvana amine-modified DNA oligonucleotide probe set (version 1564V1) and scanned using an Axon Scanner GenPix 4000 (see Supplemental Material for further details).

Northern analysis

Thirty micrograms of total RNA was separated in 15% TBE-UREA gels (Bio-Rad), transferred to a GeneScreen Plus membrane (Perkin Elmer) using semidry electroblot apparatus (Owl) in 1 \times TBE (90 mM Tris-base, 2 mM EDTA, 90 mM Boric acid) at 350 mA for 35 min. The membrane was then UV cross-linked at 1000 μ J (Stratagene) and heated for 2 h at 80°C. Prehybridization and hybridization were carried out in PerfectHyb Plus Hybridization Buffer (Sigma) supplemented with Salmon Sperm DNA (20 μ g/mL) for 2 and 16 h, respectively, at 42°C, with a radiolabeled probe added to the latter. Washes were done in 2 \times SSC + 0.2% SDS (twice), then 1 \times SSC + 0.2% SDS (once) for 5 min at 50°C. Membranes were exposed to a PhosphorImager cassette for 3 d, then scanned (PhosphorImager, Molecular Dynamics, 445 SI) and quantitated (ImageQuant, Molecular Dynamics).

MEF mRNA microarrays

Affymetrix GeneChip Mouse Genome 430_2 Array labeling, hybridization, and scanning were performed according to the manufacturer's instructions. The data discussed in this publica-

tion have been deposited in NCBI's Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO Series accession number GSE6046. To map probes on the Affymetrix Mouse 430_2 array to Refseq transcripts, we used custom CDF file MM430_MM_REFSEQ_6, downloaded from <http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF>, the custom CDF project site. Refseq transcript expression levels were then calculated using GCRMA (GCRMA package, Bioconductor in R environment) using default settings. Genes with normalized log₂ intensity below 3 were excluded from analysis.

3' UTR datasets

Genome coordinates for 3' UTRs were obtained using Refseq annotations and alignments of hg17 with 16 other vertebrate genomes, available from UCSC (<http://hgdownload.cse.ucsc.edu>) for human (hg17, May 2004), mouse (mm5, May 2004), and zebrafish (danRer3, May 2005). Only Refseq transcripts mapping uniquely to the genome were considered. Annotated 3' UTRs shorter than 50 nt were excluded.

miRNA seed match counts and conservation

The 3' UTR sequences were searched for nonoverlapping seed matches to relevant miRNAs or siRNAs of the types shown in Figure 1A. For human and mouse analyses, multiple alignments were obtained for each 3' UTR by extracting the relevant region from genomic alignments available in multiple alignment format (MAF) from UCSC (<http://hgdownload.cse.ucsc.edu>, hg17 alignments of 17 vertebrate genomes). Seed matches with perfect conservation in aligned human, mouse, rat, and dog (HMRD) UTRs were labeled conserved.

miRNA and siRNA transfection datasets

Microarray expression data for miR-1 and miR-124 HeLa transfection experiments (Lim et al. 2005) were obtained from GEO accession GSE2075. Array platform information for these experiments was obtained from GEO accession GPL1749. Probes were mapped to the human genome using BLAST, and subsequently mapped to Refseq annotated 3' UTRs using Refseq genomic mapping files available from UCSC (<http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/>). Microarray expression data for siRNA HeLa transfection experiments (Jackson et al. 2003) were obtained from <http://www.rii.com/publications/2003/nbt831.html> and from GEO accession GSE5814 (Jackson et al. 2003, 2006; Schwarz et al. 2006) and GSE5291 (Jackson et al. 2003, 2006; Schwarz et al. 2006). Only values with Refseq IDs were used for this analysis. To remove poorly expressed genes, we excluded genes with log₂ intensity < -4.0 for both datasets. The analyses reported here are based on 24-h data where repression was typically stronger.

Zebrafish embryo Dicer knockout datasets

Microarray expression data for zebrafish wild-type and MZdicer mutant embryos (Giraldez et al. 2006) were obtained from GEO (accession GSE4201). Probe information for the Affymetrix GeneChip Zebrafish Genome Array was also obtained from GEO (accession GPL1319). Probes were mapped to Refseqs using genomic

mapping information for zebrafish Affymetrix Exemplar sequences from the UCSC annotation database. Only probes with a present (P) call were used for analysis.

bic/mir-155 knockout datasets

Microarray expression data for mouse wild-type and miR-155 deficient Th1 cells (Rodriguez et al. 2007) were obtained from ArrayExpress (accession E-TABM-232). Only probes mapping to mouse Refseqs were used for analysis.

Log fold change (LFC)

The mean normalized log fold change value for miRNA/siRNA transfection experiments was defined as the difference between the mean LFC for mRNAs lacking seed matches to the transfected miRNA/siRNA and the mean LFC for mRNAs containing seed matches of the given type (median nLFC was defined analogously). For *Dicer* miRNA knockout experiments, the nLFC was defined as the difference between the mean LFC for mRNAs containing seed matches to the relevant miRNAs (e.g., miR-430) and the mean LFC for mRNAs lacking seed matches to any relevant miRNA. The variability of the nLFC value due to experimental noise was estimated for 12 “effective” siRNAs where duplicate array data were available (MAPK14-M1, -M2as, -M4as, -M5as, -M6as, -M15, -M18, MPHOSPH1-2692, PRKCE-1295, SOS1-1582as, VHL-2651as, VHL-2652, where “as” indicates the strand antisense to the targeted mRNA). The nLFCs for expressed mRNAs containing one or more extended seed matches to the relevant siRNA were calculated for each array. The average Pearson correlation between nLFC values from duplicate array pairs was 0.83. The Pearson correlation among mean nLFC values across the duplicate array pairs was 0.91. These data indicate that while there is some variation among nLFCs for individual mRNAs, the mean nLFC is highly reproducible.

Statistical analyses

All test statistics were calculated using R (<http://www.r-project.org>). The Wilcoxon rank sum test was chosen over the t-test because it does not assume normality of the underlying distributions, and because it is more intuitive and familiar than non-parametric alternatives such as the Kolmogorov–Smirnov (KS) test. t-tests and KS tests using these data gave generally similar results. A *P*-value cutoff of 0.05 was used for all analyses.

SOFTWARE AVAILABILITY

A TargetRank Web server is available at <http://genes.mit.edu/targetrank/>.

SUPPLEMENTAL MATERIAL

Supplemental Material, including nine figures, seven tables, Experimental Procedures, and References can be found at http://genes.mit.edu/burgelab/Supplementary/nielsen_shomron07/.

ACKNOWLEDGMENTS

We thank R. Friedman and P.A. Sharp for helpful discussions and comments on the manuscript. Several of the analyses in this work

relied on data published by scientists at Rosetta Inpharmatics and at the Schier, Turner, and Bradley laboratories, for which we are grateful. This work was supported by fellowships from NSERC (C.B.N.), and from the Knut and Alice Wallenberg Foundation (R.S.), and by a Searle Scholar Award and a grant from the NIH (C.B.B.).

Received July 26, 2007; accepted August 8, 2007.

REFERENCES

- Abbondanzo, S.J., Gadi, I., and Stewart, C.L. 1993. Derivation of embryonic stem cell lines. *Methods Enzymol.* **225**: 803–823.
- Ambros, V. 2004. The functions of animal microRNAs. *Nature* **431**: 350–355.
- Bagga, S., Bracht, J., Hunter, S., Massierer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. 2005. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122**: 553–563.
- Barreau, C., Paillard, L., and Osborne, H.B. 2005. AU-rich elements and associated factors: Are there unifying principles? *Nucleic Acids Res.* **33**: 7138–7150. doi: 10.1093/nar/gki1012.
- Bartel, D.P. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.
- Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. 2003. Dicer is essential for mouse development. *Nat. Genet.* **35**: 215–217.
- Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.L., and Filipowicz, W. 2006. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125**: 1111–1124.
- Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. 2005. Principles of microRNA-target recognition. *PLoS Biol.* **3**: e85. doi: 10.1371/journal.pbio.0030085.
- Danielian, P.S., Muccino, D., Rowitch, D.H., Michael, S.K., and McMahon, A.P. 1998. Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Curr. Biol.* **8**: 1323–1326.
- Doench, J.G. and Sharp, P.A. 2004. Specificity of microRNA target selection in translational repression. *Genes & Dev.* **18**: 504–511.
- Doench, J.G., Petersen, C.P., and Sharp, P.A. 2003. siRNAs can function as miRNAs. *Genes & Dev.* **17**: 438–442.
- Farh, K.K., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. 2005. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* **310**: 1817–1821.
- Filipowicz, W., Jaskiewicz, L., Kolb, F.A., and Pillai, R.S. 2005. Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr. Opin. Struct. Biol.* **15**: 331–341.
- Gaidatzis, D., van Nimwegen, E., Hausser, J., and Zavolan, M. 2007. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* **8**: 69.
- Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**: 75–79.
- Grun, D., Wang, Y.L., Langenberger, D., Gunsalus, K.C., and Rajewsky, N. 2005. microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput. Biol.* **1**: e13. doi: 10.1371/journal.pcbi.0010013.
- Harfe, B.D., McManus, M.T., Mansfield, J.H., Hornstein, E., and Tabin, C.J. 2005. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. *Proc. Natl. Acad. Sci.* **102**: 10898–10903.
- Hayashi, S. and McMahon, A.P. 2002. Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: A tool for

- temporally regulated gene activation/inactivation in the mouse. *Dev. Biol.* **244**: 305–318.
- Hutvagner, G. and Zamore, P.D. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**: 2056–2060.
- Hwang, D.G. and Green, P. 2004. Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc. Natl. Acad. Sci.* **101**: 13994–14001.
- Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**: 635–637.
- Jackson, A.L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J.M., Lim, L., Karpilow, J., Nichols, K., et al. 2006. Position-specific chemical modification of siRNAs reduces “off-target” transcript silencing. *RNA* **12**: 1197–1205.
- John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. 2004. Human MicroRNA targets. *PLoS Biol.* **2**: e363. doi: 10.1371/journal.pbio.0020363.
- Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. 2006. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**: 19–53.
- Kim, V.N. 2005. MicroRNA biogenesis: Coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6**: 376–385.
- Kiriakidou, M., Tan, G.S., Lamprinak, S., De Planell-Saguer, M., Nelson, P.T., and Mourelatos, Z. 2007. An mRNA m(7)G cap binding-like motif within human Ago2 represses translation. *Cell* **129**: 1141–1151.
- Lai, E.C. 2002. MicroRNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* **30**: 363–364.
- Lee, R.C., Feinbaum, R.L., and Ambros, V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843–854.
- Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. 2003. Prediction of mammalian microRNA targets. *Cell* **115**: 787–798.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15–20.
- Lim, L.P., Lau, N.C., Garrett-Engle, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. 2005. Microarray analysis shows that some microRNAs down-regulate large numbers of target mRNAs. *Nature* **433**: 769–773.
- Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**: 2053–2056.
- Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. 2005. Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**: 666–670.
- Majoros, W.H. and Ohler, U. 2007. Spatial preferences of microRNA targets in 3′ untranslated regions. *BMC Genomics* **8**: 152. doi: 10.1186/1471-2164-8-152.
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**: 185–197.
- Muljo, S.A., Ansel, K.M., Kanellopoulou, C., Livingston, D.M., Rao, A., and Rajewsky, K. 2005. Aberrant T cell differentiation in the absence of Dicer. *J. Exp. Med.* **202**: 261–269.
- Parker, J.S., Roe, S.M., and Barford, D. 2005. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**: 663–666.
- Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. 2005. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**: 1573–1576.
- Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. 2005. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* **11**: 1640–1647.
- Robins, H., Li, Y., and Padgett, R.W. 2005. Incorporating structure to predict microRNA targets. *Proc. Natl. Acad. Sci.* **102**: 4006–4009.
- Rodriguez, A., Vigorito, E., Clare, S., Warren, M.V., Couttet, P., Soond, D.R., van Dongen, S., Grocock, R.J., Das, P.P., Miska, E.A., et al. 2007. Requirement of bic/microRNA-155 for normal immune function. *Science* **316**: 608–611.
- Saetrom, P., Heale, B.S., Snove Jr, O., Aagaard, L., Alluin, J., and Rossi, J.J. 2007. Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res.* **35**: 2333–2342. doi: 10.1093/nar/gkm133.
- Schmitter, D., Filkowski, J., Sewer, A., Pillai, R.S., Oakeley, E.J., Zavolan, M., Svoboda, P., and Filipowicz, W. 2006. Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Res.* **34**: 4801–4815. doi: 10.1093/nar/gkl646.
- Schwarz, D.S., Ding, H., Kennington, L., Moore, J.T., Schelter, J., Burchard, J., Linsley, P.S., Aronin, N., Xu, Z., and Zamore, P.D. 2006. Designing siRNA that distinguish between genes that differ by a single nucleotide. *PLoS Genet.* **2**: e140. doi: 10.1371/journal.pgen.0020140.
- Soriano, P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat. Genet.* **21**: 70–71.
- Stark, A., Brennecke, J., Bushati, N., Russell, R.B., and Cohen, S.M. 2005. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′ UTR evolution. *Cell* **123**: 1133–1146.
- Thomson, J.M., Parker, J., Perou, C.M., and Hammond, S.M. 2004. A custom microarray platform for analysis of microRNA gene expression. *Nat. Methods* **1**: 47–53.
- Valencia-Sanchez, M.A., Liu, J., Hannon, G.J., and Parker, R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Dev.* **20**: 515–524.
- Vasudevan, S. and Steitz, J.A. 2007. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* **128**: 1105–1118.
- Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K., and Slack, F.J. 2004. The *C. elegans* microRNA *let-7* binds to imperfect *let-7* complementary sites from the *lin-41* 3′UTR. *Genes & Dev.* **18**: 132–137.
- Wienholds, E., Koudijs, M.J., van Eeden, F.J., Cuppen, E., and Plasterk, R.H. 2003. The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat. Genet.* **35**: 217–218.
- Wightman, B., Ha, I., and Ruvkun, G. 1993. Post-transcriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855–862.
- Wu, L., Fan, J., and Belasco, J.G. 2006. MicroRNAs direct rapid deadenylation of mRNA. *Proc. Natl. Acad. Sci.* **103**: 4034–4039.
- Yekta, S., Shih, I.H., and Bartel, D.P. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**: 594–596.
- Zamore, P.D. and Haley, B. 2005. Ribo-gnome: The big world of small RNAs. *Science* **309**: 1519–1524.
- Zeng, Y., Wagner, E.J., and Cullen, B.R. 2002. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**: 1327–1333.
- Zeng, Y., Yi, R., and Cullen, B.R. 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl. Acad. Sci.* **100**: 9779–9784.