

Distance constraints between microRNA target sites dictate efficacy and cooperativity

Pål Sætrom^{1,2,3}, Bret S.E. Heale¹, Ola Snøve Jr^{1,2,4}, Lars Aagaard¹, Jessica Alluin¹ and John J. Rossi^{1,*}

¹Division of Molecular Biology, Beckman Research Institute of City of Hope, Duarte, CA 91010, USA,

²Interagon AS, Laboratorienesenteret, NO-7489 Trondheim, Norway, ³Department of Computer and Information Science, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway and

⁴Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, NO-7489 Trondheim, Norway

Received November 21, 2006; Revised February 16, 2007; Accepted February 19, 2007

ABSTRACT

MicroRNAs (miRNAs) have the potential to regulate the expression of thousands of genes, but the mechanisms that determine whether a gene is targeted or not are poorly understood. We studied the genomic distribution of distances between pairs of identical miRNA seeds and found a propensity for moderate distances greater than about 13 nt between seed starts. Experimental data show that optimal down-regulation is obtained when two seed sites are separated by between 13 and 35 nt. By analyzing the distance between seed sites of endogenous miRNAs and transfected small interfering RNAs (siRNAs), we also find that cooperative targeting of sites with a separation in the optimal range can explain some of the siRNA off-target effects that have been reported in the literature.

INTRODUCTION

MicroRNAs (miRNAs) belong to a well-conserved class of non-protein-coding RNA (1–3) with important regulatory functions that control animal development and physiology [reviewed in (4)]. miRNAs are also implicated in disease, as they act both as tumor suppressors (5) and oncogenes (6) and their expression profiles can hold more diagnostic information than those of messenger RNAs (mRNAs) (7). The therapeutic potential of miRNA-like molecules, such as short interfering RNAs and short hairpin RNAs, also increases the importance of research that seeks to understand miRNA biology (8).

Endogenous miRNAs can direct sequence-specific down-regulation by cleavage (9–11), degradation (12–14)

or translational suppression (15–17) of mRNA. The 8.2 release of the miRNA registry holds 462 human miRNA genes (18) and current estimates suggest that the total number may be almost twice as high (19). Target motif conservation studies and extrapolation of data have shown that these miRNAs have the potential to target thousands of protein-coding genes (20–25). Indeed, the potential for such widespread effects have been confirmed with microarray analyses since ectopic expression of both endogenous miRNAs (26) and synthetic small interfering RNAs (siRNAs) (27,28) mediate significant off-target down-regulation of numerous genes.

Despite their importance, we do not yet have a clear understanding of the factors that determine whether a message will be targeted by miRNAs or by which mechanism silencing will be accomplished. Cleavage seems to be restricted to messages with near-perfect complementarity to the miRNA, whereas translational suppression occurs when partial complementarity to the message occurs within the 3' UTR [reviewed in (29)]. Several studies have demonstrated that sequence complementarity between the 3' UTR and 2–7 or 2–8 nt of the 5' end of the miRNA—often denoted as 'the seed'—is particularly important (20–23,25,30,31). Nevertheless, a seed site is neither necessary nor sufficient for miRNA down-regulation. miRNA target sites can tolerate G:U wobble base pairs within the seed region (32,33), and extensive base pairing between the 3' UTR and the remainder of the miRNA may offset missing complementarity of the 5' seed (23). Furthermore, even sites with extensive 5' complementarity can be inactive when tested in reporter constructs (32,34). Multiple target sites in the same 3' UTR can potentially increase the degree of translational suppression (35). Adding to the number of possible targets is the potential for several miRNAs to mediate cooperative effects by targeting the same transcripts (25). Consequently, a target site's activity will

*To whom correspondence should be addressed. Tel: +1 626 301 8360; Fax: +1 626 301 8271; Email: jrossi@coh.org
The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

depend on its surrounding context and a single ineffective site in a reporter may be effective in its original genomic context and *vice versa* (34). Furthermore, despite seed sites' inability to predict all miRNA target sites, seed sites may outnumber other sites by 10 to 1 (23) and as many as 85 and 50% of conserved and unconserved seed sites may be functional (36). Seed site analysis is therefore currently the preferred method to predict and analyze global trends in miRNA targeting (37).

In two cases, intervening sequences between repeated target sites have been necessary for strong miRNA regulation of targets (30,33), but the relationship between spacing of miRNA interaction sites and functional suppression of translation is poorly understood. Here, we show that the conservation pattern of seed sites is characterized by spacings of ~10–130 nt. We verify this experimentally by demonstrating that optimal potency requires an even tighter distance of 13–35 nt between neighboring sites. Furthermore, we show that off-target effects in siRNA experiments are related to cooperative interactions between endogenous miRNAs and the transfected siRNAs. Instances where seed sites for the transfected siRNAs and endogenous miRNAs are optimally spaced occur more frequently in off-target genes. Similarly, in genes that are not off-targets, despite having siRNA seed sites, siRNA and miRNA seed sites are more frequently too close for cooperative interactions. The distance dependence for seed site effectiveness is also supported by existing data.

Our results may explain why some miRNAs are more specific than that which is predicted from the properties of the individual sites and why miRNA targeting can depend on the specific 3' UTR context (34). These results could be used to optimize algorithms for miRNA target and siRNA off-target predictions.

MATERIALS AND METHODS

MicroRNA seeds

Seed sequences from families of highly conserved miRNAs (20), which contain 148 miRNAs with 62 unique hexamer seeds (2–7 nt) and 63 unique heptamer seeds (2–8 nt) were used for our analyses. Our control seed sequences were shuffled sequences obtained using the same procedure as that of Lewis *et al.* (20). In addition, to prevent any bias caused by differences in 3' UTR lengths, we required that the total length and the total number of 3' UTRs containing two or more occurrences of the shuffled seeds are comparable ($\pm 7.5\%$) to that of a miRNA seed. From this, we obtained 756-hexamer controls; see the Supplementary Data for details on heptamer controls. Thus, even though there will be some miRNA seeds in the control set, we expect that the relative number of false negatives will be low and therefore should not affect our results (Supplementary Figure 1). Furthermore, the miRNA and control seeds match a similar number of repeat regions in the 3' UTRs (663 ± 478 and 667 ± 347 ; average \pm SD), as defined by RepeatMasker (<http://www.repeatmasker.org>).

Distance between miRNA seed sites

The distance from the start of one seed to the start of another was used to identify the spacing of target sites. Others have referred to the spacing between targets as the number of nucleotides that separate sites (38), but this term is ambiguous because where targets begin and end cannot easily be determined. Furthermore, as miRNA-binding sites have very different structural properties (39), that particular annotation makes it difficult to compare results between studies. In addition, since we use the distance between seed starts, the length of the seed itself does not change the spacing parameter.

mRNA dataset

UCSC's table browser was used to download every human RefSeq mRNA with a 3' UTR of more than 6 nt (40). We then removed multiple 3' UTR transcript variants by mapping each sequence to a UniGene ID and keeping the longest 3' UTR. This resulted in a database of 17448 entries. Furthermore, we used aligned 3' UTRs from human, chimp, mouse and rat from UCSC's multiz multiple alignment files to deduce whether a given site was conserved between species.

Distance between siRNA and miRNA seed sites

When looking at the distance between the siRNA seed site and the closest miRNA seed site, we used the same highly conserved seeds as in our distance conservation studies. To define which miRNAs were expressed in HeLa, we required that the miRNAs' log-expression level, as reported previously (41), was above 2. This gave a list of 49 expressed miRNA seeds.

MicroRNA targeting assay using inhibition of endogenous miRNA

HeLa cells (seeded in 48-well plates on the day prior) were transfected in triplicate with 50 ng of reporter construct containing the various let-7 seed match combinations and either 15 pmol of a 2'-O-methyl RNA complementary to let-7a (anti-let-7a) or an irrelevant 2'-O-methyl RNA control (5'-CACAAUGCGCUCUCGAACGUUA-3') using lipofectamine2000 according to the manufacturer's recommendations (Invitrogen). Forty hours post transfection, the cells were lysed with Passive Lysis Buffer (Promega). Renilla and Firefly luciferase levels were then analyzed from 10 μ l lysates using the Dual luciferase reporter assay (50 μ l of each substrate reagent, Promega) on a Veritas Microplate Luminometer (Turner Biosystems). Changes in expression of Renilla luciferase (target) were calculated relative to firefly luciferase (internal control) and normalized to the irrelevant control.

MicroRNA targeting assay using miRNA mimics

HEK293 cells (seeded in 48-well plates on day prior) were transfected in duplicate or triplicate with 50 ng of reporter construct containing the various let-7 seed match combinations and either 25 pmol duplex RNA mimicking let-7f-2 (5'-UGAGGUAGUAGAUUGUAUAGUU-3' annealed to 5'-CUAUACAGUCUACUGUCUUUC-3'

at 100 μ M concentration) or 25 pmol of an irrelevant siRNA (5'-UAUACGAAGUUAUCGAAGCUU-3', 5'-GCUUCGAUAACUUCGUAUAUU-3'). Twenty-four hours post transfection, the cells were analyzed for Renilla and Firefly luciferase levels as described above and normalized to the irrelevant control.

MicroRNA targeting assay using miRNA expression constructs

HEK293 cells (seeded in 24-well plates on day prior) were transfected in duplicate with 25 ng of reporter construct and either 250 ng of mir-106b-25 wild-type construct, 250 ng of mir-106b-25 irrelevant construct or 250 ng of 'empty' control plasmid DNA (pcDNA3). The reporters contain either a 1.2-kb region of the BMPR2 3'UTR, a 17-nt-spaced miR106b/miR93/miR25 triple target based on BMPR2 or a 300 nt fragment of EGFP-coding region. Twenty-four hours post transfection, the cells were analyzed for Renilla and Firefly luciferase levels as described above and normalized to the 'empty' control.

Plasmid construction

Target sequences were cloned into the 3' untranslated region (UTR) of the Renilla luciferase gene in the psiCHECK2.2 vector (Promega). The let-7 targets and the super mir-106b-25 target sequence were cloned directly into the unique XhoI-Not I and the unique XhoI-SpeI restriction sites of the psiCHECK2.2 vector, respectively, using phosphorylated synthetic DNA oligos (IDT). All oligos are listed in Supplementary Table 1. A 1.2-kb fragment of the 3'UTR from the human bone morphogenetic protein receptor type II (BMPR2, NM_001204) gene was cloned into the XhoI-SpeI sites of psiCHECK2.2 from a PCR amplicon derived from human genomic DNA using the following XhoI- and SpeI-tagged primers: 5'-GTAACTCGAGGCTTTATCTTCCCATCTAACTCTT-3' (BMPR2-3UTR forward) and 5'-GTAACTAGTTGATATACAATTCTGTGTGCATGGC-3' (BMPR2-3UTR reverse). The psiEGFP plasmid was previously described (42). The polycistronic miRNA expression construct (mir-106b-25 wild type) was cloned directly from genomic DNA into pcDNA3 (Invitrogen), and a modified version expressing non-targeting small RNAs (mir-106b-25 irrelevant) was prepared from the mir-106b-25 construct (L. Aagaard, K. von Eije, J.J. Rossi and M. Amarzguoui; unpublished results). All constructs were verified by sequence analyses.

Statistical analyses

Randomization was used to determine whether the distance between two occurrences of the same miRNA seed site in 3' UTRs was different from what we could expect at random. More specifically, for each miRNA seed, we counted the number of times the distance from one occurrence to the next non-overlapping occurrence of the seed was less than or equal to a given distance threshold. We summed this count over all miRNAs and normalized this with the total number of two consecutive non-overlapping occurrences of the miRNA seeds. We then compared this normalized count to the same

count for an equal number of randomly selected control seeds. By repeating this process for several iterations and counting the number of times the miRNA count was higher than the random count, we estimated the *P*-value of whether the distance between miRNA seed sites is underrepresented for a particular distance threshold.

Similarly, we used randomization to determine whether the distance between conserved heptamers was different from what we could expect by random occurrence. To do this, we (i) computed the positions of all heptamers in the human 3' UTRs; (ii) randomly removed heptamer positions such that the expected number of heptamer occurrences was equal to the number of heptamers conserved between human, chimp, mouse and rat in the same UTR sequences; (iii) recorded the distribution of distances between two consecutive non-overlapping occurrences of the heptamers; and (iv) repeated this randomization process for several iterations to get an estimate of the average of distance distributions and standard deviations across all observed distances.

RESULTS

Distance-dependent conservation pattern for pairs of identical miRNA seeds

We investigated the spacing requirements for cooperative miRNA target site interactions by comparing the distances between pairs of identical miRNA seed sites to the corresponding spacing for conserved random controls (Figure 1a). The clearest pattern in the distance distribution is that miRNA seeds are underrepresented when the sites are close. This trend holds up to a distance of about 12 nt (Figure 1b).

MicroRNA seed sites separated by between 16 and 20 nt are also overrepresented. Although this pattern is less clear than the pattern for the close sites, the overrepresentation of miRNA sites is the largest for all distances except the large deviations for the two distances 208 and 1070. These deviations are, however, isolated and represent outliers in the distribution. Indeed, most counts for the two distances are from the same seed matching multiple genes in the proto-cadherin alpha and gamma gene clusters (43). No other gene families contribute such a disproportionate amount of counts to the distribution (data not shown). To better visualize trends within regions of the distribution, we smoothed the distance distribution by computing the average value within a sliding window (Figure 1c). This smoothing made the under- and overrepresentation of miRNA sites at distances less than 13 nt and at distances from 16 to 20 nt more clear. Further analyses with different control seeds confirmed these results (Supplementary Figure 2). miRNA heptamer and adenosine-anchored hexamer seeds had the same trends (Supplementary Figure 3). Finally, co-expressed miRNAs can also cooperatively regulate targets (25), and miRNAs located within genomic regions of about 50 kb tend to be co-expressed (41). We grouped the evolutionarily conserved miRNAs into putatively co-expressed clusters and measured the distance between occurrences of conserved seed sites for the miRNAs within each cluster.

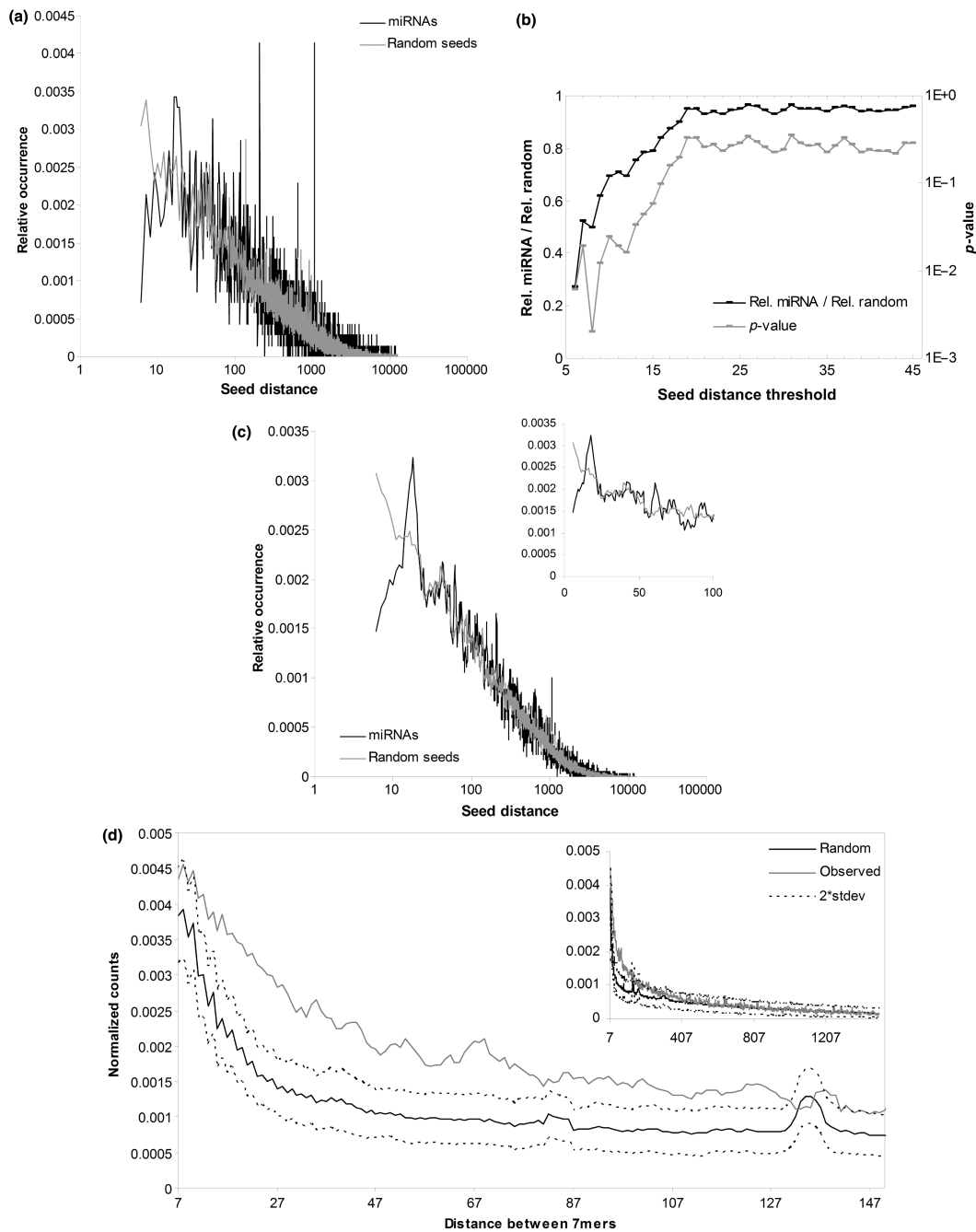


Figure 1. Pairs of identical miRNA seeds have a distance-dependent conservation pattern. **(a)** Pairs of miRNA hexamer seeds are underrepresented for distances of 13 nt or less. We counted the number of times the pairs of conserved miRNA hexamer seed sites were separated by a given number of nucleotides and compared the relative occurrences with the corresponding occurrences for random controls (see the Materials and methods section). Very close pairs of identical miRNA seeds occur much less frequently than the random controls do. **(b)** To determine the significance of the underrepresentation, we ran a randomization experiment that compared the relative occurrences of pairs of conserved miRNA seeds closer than a given distance threshold with the corresponding occurrences for random controls (see the Materials and methods section). The graph shows for increasing distance cutoffs, the average of the relative miRNA occurrences divided by the relative random occurrences (black; primary y-axis) and the estimated P -values (gray; secondary y-axis). The underrepresentation of miRNA seeds holds up to a distance of 12 nt after which the P -values increase rapidly. **(c)** The smoothed distance distribution indicates that miRNA seed sites are overrepresented for distances between 16 and 20 nt. We computed the moving averages of the miRNA and random distance distributions from **(a)** (moving average window size 5). In the resulting distribution, the largest deviations from random, except for the underrepresentation of miRNAs at distances less than 13 nt are the overrepresentation of miRNAs for distances between 16 and 20 nt. The graph in the upper right corner shows an excerpt of the distance distribution in a linear scale on the x-axis. **(d)** Pairs of heptamers are more likely to be conserved together when the distance is less than 130. The graph shows the distribution of distances between two consecutive non-overlapping occurrences of conserved miRNA heptamers (gray solid line), and the corresponding average distribution for simulated random conservation (black solid line). The real conservation distribution differs from the random distribution for distances between 10 and 130. Outside this range, the real conservation distribution approaches the random distribution (graph in upper right corner). The graph is smoothed by using a moving average with a window size of 5; see Supplementary Figure 5 for the original distribution.

A maximum distance of 50 kb gave 13 clusters with non-redundant seeds and the resulting distance distribution showed the same trend of under- and overrepresentation as the distribution for multiple occurrences of identical seeds (Supplementary Figure 4).

To further study the relationship between conservation and distance between seed sites, we carried out another randomization experiment in which we compared the actual evolutionary conservation patterns of all pairs of identical heptamers to that of random conservation (see the Materials and methods section). Figure 1d demonstrates that heptamer pairs separated by more than 130 nt are no more likely to be conserved together than would be expected if sequences were conserved randomly. Additionally, very close heptamer pairs have a random conservation pattern. The implication for pairs of miRNA seed sites is that they must be relatively close to have a high probability of being conserved together.

Distance between seed sites affects target down-regulation

Based on our observations of multiple seed site spacings, we hypothesized that miRNAs would have suboptimal efficacy for target sites that are very close to one another, whereas distantly spaced sites may not contribute to enhanced efficacy. If the overrepresentation of miRNA seeds at distances between 16 and 20 nt (Figure 1c) is biologically significant, certain distances between seeds should be optimal for strong, cooperative effects. We therefore set out to experimentally test this prediction using several constructs with varied spacing between seed sequence binding sites.

As shown in Figure 2a, we chose distances of 9, 13, 17, 21, 24, 35 and 70 bases between seed sites to examine the seed site spacing requirements within the region in Figure 1. Heptamer seeds are reportedly sufficient for efficient targeting even without 3' pairings beyond the seed (23). To ensure that seed complementarity would be enough to generate down-regulation, we therefore chose 7-nt target sites for our experiments. Each site was designed for targeting by let-7 miRNA. To avoid interactions that could interfere with our analyses, we minimized the binding potential between the 3' ends of let-7 with our target sites such that all target sites had similar binding energies (Supplementary Figure 6). Furthermore, we designed the constructs to have a low potential for forming stable self-interacting secondary structures. This design should prevent radical differences in target site accessibility influencing our analyses.

To evaluate the relative strengths of the variably spaced multiple seed target sites, we measured the ratio of increased target expression following transfection of a let-7 antagomir (44) into HeLa cells, as these express relatively high levels of endogenous let-7 (data not shown). Expression levels of the reporter were normalized to those obtained after transfection of an irrelevant control antagomir. A pattern of activity that depends upon the distance between target sites, similar to the genomic distribution of seed pairs, emerged from these experiments (Figure 2b; see Supplementary Figure 7 for

supporting results with a different miRNA). First, two sites that are very close, such as 9 bases, can inhibit efficacy in comparison to a single site (P -value 0.002, Student's t -tests comparing 'single' and '9' for let-7 antagomir data). Second, favorably spaced paired sites yield about twice the efficacy of a single site, but this additive effect falls off as sites become separated by an increasing number of nucleotides. To ensure that these correlations hold in a different cell line with another assay, we confirmed our results by transfecting a let-7 mimic siRNA into HEK293 cells (Figure 2c). We chose HEK293 cells because this cell line expresses less endogenous let-7 than HeLa cells (data not shown). Note that the trends are clearer for the overexpression than for the knockdown assay.

Our data suggest a model in which the extent of miRNA-mediated down-regulation depends upon the distance between sites, and may indicate that the miRNA-containing effector complexes interact cooperatively. To further test the distance requirements for multiple sites, we designed two additional target sites, which are shown schematically in Figure 2d. The results show that three individual sites are slightly more effective than two individual sites. Furthermore, two optimally spaced pairs of seed sites (17 bases between seeds) separated by 50 nt produced even greater inhibition than the triple seed site (Figure 2e and f). The dependence upon distance between seed pairs suggests cooperative interactions between miRNA complexes interacting at these sites, perhaps stabilizing the interactions of the complexes with the target sequence.

Note that differences between down-regulation levels are not as large in the let-7 mimic assay as in the let-7 inhibitory assay—especially between the most effective sites. This is probably due to saturation, as even a single site has about 50% down-regulation. In both assays, however, the increased potency of the triple site over the double site is relatively small compared to that of the two pairs of sites spaced by 50 bases. Previously reported experiments with four and six sites at distances of 24 nt gave consistently increased knockdown (35). We speculate that a distance of 17 nt between seeds gives suboptimal down-regulation for more than two seed sites. Pairs seem to be well tolerated, but three sites do not give the expected increase in potency, perhaps due to steric hindrance between complexes. Optimally spaced pairs may also stabilize the miRNA complexes and give cooperative interactions at longer distances than single sites do.

Distance between seed sites affects cooperative down-regulation by different miRNAs

Our experiments with artificially designed let-7 targets suggest that the distance between miRNA target sites is more important than previously recognized. To investigate whether this result could be generalized to other miRNAs and endogenous target sites, we searched for potential targets of miR-106b, miR-93, and miR-25—three miRNAs that are processed from a single intron of the MCM7 gene on chromosome seven (6). One possible

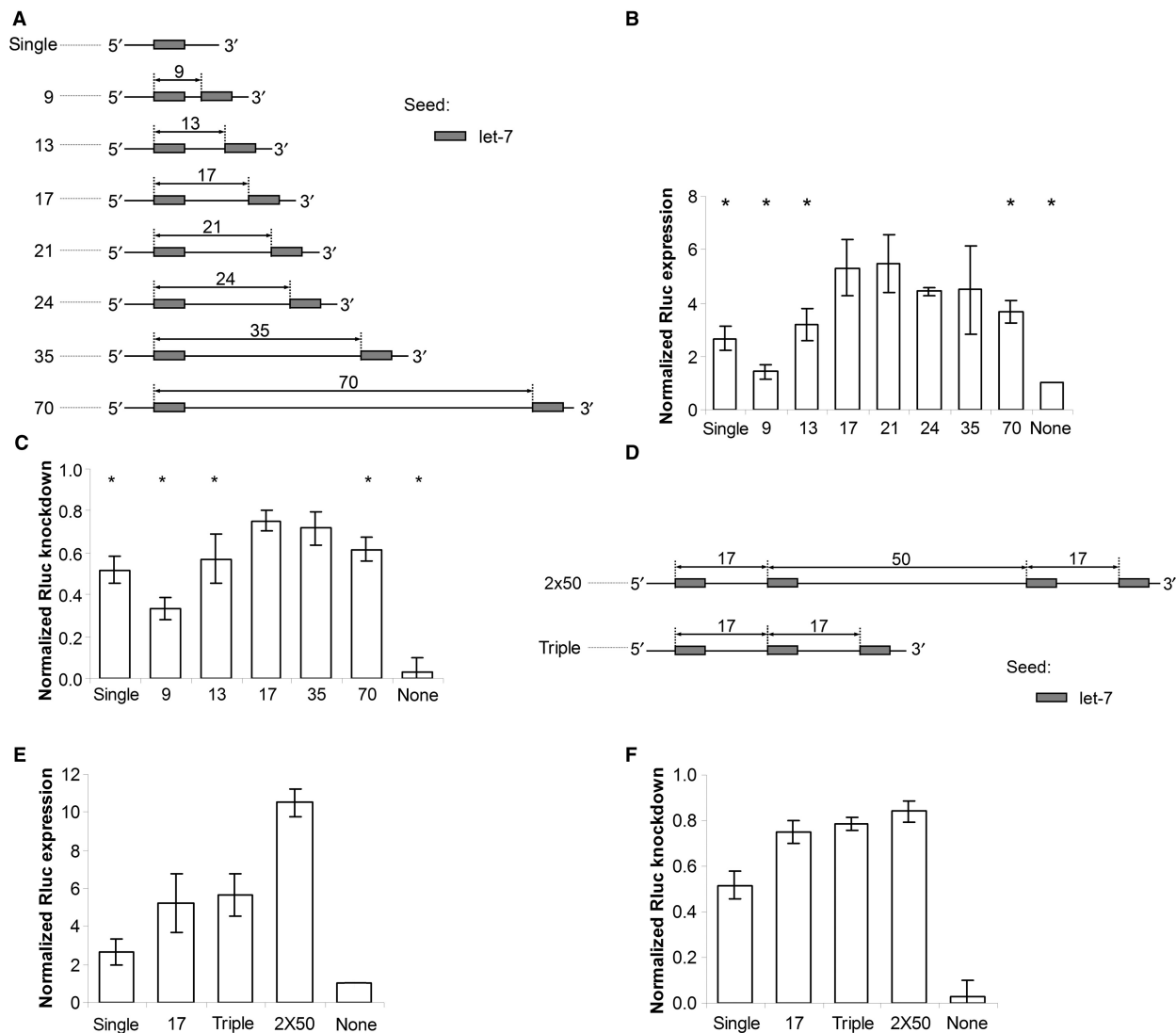


Figure 2. The distance between seed sites affects target down-regulation. We cloned different let-7 target site configurations into the 3' UTR of Renilla luciferase reporter constructs, transfected the constructs along with an anti-let-7a 2'-O-methyl RNA into HeLa cells [(b) and (e)] and a let-7 mimic in HEK293 cells [(c) and (f)], and measured the change in luciferase expression compared to irrelevant controls. (a) Schematic depiction of target sites with distances of 9, 13, 17, 21, 24, 35, and 70 between seed starts. (b) Ratio of increased expression in HeLa; and (c) percentage knockdown in HEK293 normalized to a control without a seed site for targets shown in (a). Asterisks (*) mark values that are significantly different from that of the seed sites with distance of 17 [Student's *t*-test, confidence level 0.05; (b) *P*-values for single, 9, 13, 21, 24, 35, 70, and none were 1E-5, 2E-4, 0.01, 0.8, 0.1, 0.3, 0.002, and 1E-8; (c) *P*-values for single, 9, 13, 35, 70, and none were 2E-4, 3E-5, 0.02, 0.5, 0.01, and 5E-8 (c)]. (d) Schematic depiction of one target site that has three optimally spaced seeds and another that has 50 nt between two optimally spaced pairs. (e) Ratio of increased expression in HeLa and (f) percentage knockdown in HEK293 normalized to control without a seed site for targets shown in (d). In (b), (c), (e), and (f), columns are the average of at least three independent experiments carried out in triplicate; error bars are standard deviations.

target for these miRNAs, referred to as the mir-106b-25 cluster, is *BMPR2*, as its 3' UTR contains possible seed target sites for each miRNA in the cluster (Supplementary Figure 8), which makes *BMPR2* a good candidate for cooperative targeting. The three miRNAs give no detectable knockdown of a reporter that contains part of the *BMPR2* 3' UTR sequence with the three predicted target sites (Figure 3). Importantly, when the same target sites are moved closer together in a configuration resembling the triplet in Figure 2c, we observed 30% down-regulation versus the non-specific controls. Thus, the spacing between binding sites also influences cooperativity between multiple miRNAs.

Distance between siRNA and endogenous miRNA seed sites affects siRNA off-targeting

Off-target down-regulation by siRNAs is related to the presence of siRNA seed sites in the off-targeted transcripts' 3' UTRs (27,28,45–47). Nevertheless, only a small percentage of transcripts that contain seed sites are significantly down-regulated by the siRNAs. To illustrate, Birmingham *et al.* report at most 73 significant off-targets for the 12 siRNAs used in their study, but these siRNAs have hexamer seed sites in between 1007 and 5627 of the 3' UTRs in our dataset (45). Jackson *et al.* note that siRNA off-target transcripts share many characteristics of

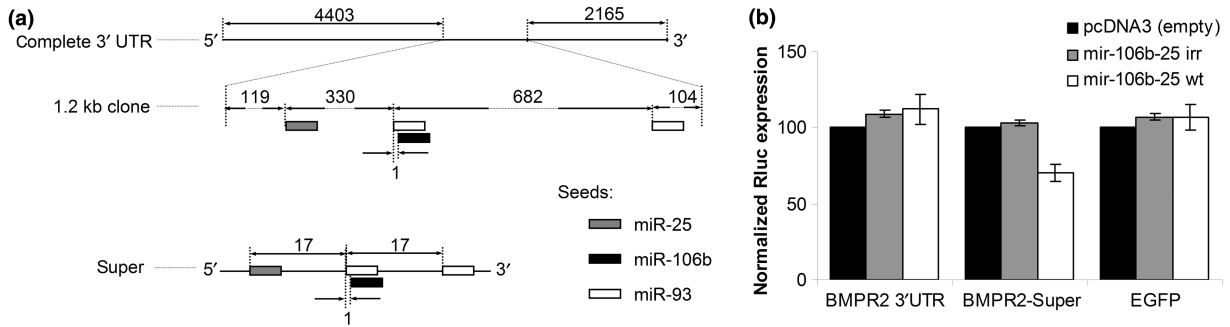


Figure 3. Polycistronic miRNAs from the MCM7 intron show no collaborative effect on the predicted endogenous target BMPR2, but produce 30% down-regulation when targets are moved closer. **(a)** Schematic representation of the 3' UTR of BMPR2 and the predicted targets of miR-106b, miR-93, and miR-25 from the MCM7 intron. **(b)** The percentage knockdown of the wild-type mir-106b-25 polycistron (mir-106b-25 wt) and a modified polycistron containing irrelevant controls (mir-106b-25 irr) on a Renilla luciferase reporter harboring 1.2 kb of the endogenous target (BMPR2 3'UTR) and the modified target (BMPR2-Super). Columns are the average of three independent experiments carried out in duplicate; error bars depict standard deviations.

miRNA targets and are enriched for miRNA target sites (28). We therefore hypothesized that siRNA off-targeting is partly caused by cooperative interactions with miRNAs expressed in the cells.

To test this hypothesis, we carried out an experiment in which we measured the distance from a siRNA hexamer seed site to the closest non-overlapping miRNA hexamer seed site in the 3' UTRs of the off-target genes reported by Birmingham *et al.* (45). Figure 4 shows that siRNA seed sites in off-targeted genes have fewer miRNA seed sites within a distance of 14 nt compared to the reference set of all 3' UTRs containing siRNA seed sites. This corresponds to our previous results, where distances of 13 nt or less between identical miRNA seed sites were underrepresented in conserved 3' UTRs and gave similar or reduced knockdown compared to single sites. Thus, given that a 3' UTR with a siRNA seed site represents a potential off-target, it seems that some of the potential off-targeting is prevented by the negative interactions of the siRNA seed site being close to a miRNA site.

In the previous experiment, we looked at several miRNAs, some of which may be expressed at low levels in the HeLa cell line used in the original off-target study. We therefore redid the analysis, but limited the dataset to the miRNAs previously reported to be expressed in HeLa (41). The trend that off-target genes have fewer miRNA seed sites within a distance of 14 nt compared to the reference set became even clearer in this analysis (Figure 5; Supplementary Figure 9), but we also saw that siRNA hexamer seeds are more often at a distance of 14–25 nt from miRNA seeds in off-target genes than in the reference set. Our experimental results show that this distance interval gives optimal cooperative down-regulation. Thus, it seems that some off-target effects are caused by the siRNAs cooperating with endogenous miRNAs to down-regulate mRNAs.

To confirm these findings, we analyzed the data from a different study in which three miRNA duplexes were over-expressed in HeLa cells and microarrays were used to follow target knockdown (26). This analysis revealed the same trends (Supplementary Figure 10). Thus, both off-targeting by siRNAs and targeting by over-expressed

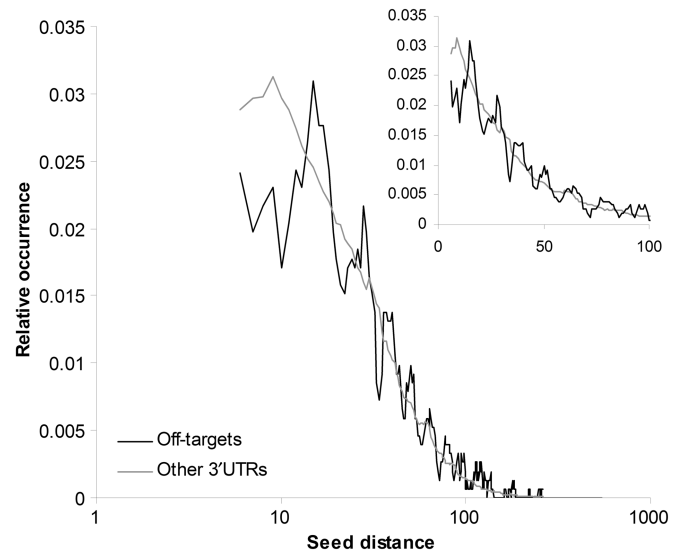


Figure 4. Short interfering RNA seed sites are located farther from miRNA seed sites in off-targeted genes than in other genes containing siRNA seed sites. The graphs show the smoothed (sliding window of size 5) distance distribution for the distance between siRNA hexamer seed sites and the closest non-overlapping miRNA hexamer seed site in off-targeted 3' UTRs (black) and other 3' UTRs that contain siRNA seed sites (gray). The graph in the upper right corner shows an excerpt of the distance distribution in a linear scale on the x-axis. The miRNA seeds are the seeds from the highly conserved miRNAs defined by Lewis *et al.* (20).

miRNAs are related to distance-dependent interactions with endogenously expressed miRNA seed sites.

Previous studies on cooperative down-regulation support our results

Several earlier studies have examined cooperative down-regulation via multiple miRNA target sites (Table 1). In most of these studies, the target sites were optimally spaced between 16 and 29 nt and showed cooperative down-regulation (23,30,31,35). Doench and Sharp (30) also looked at target sites 8 nt apart and found that in the context of two optimally spaced (24 nt) flanking target

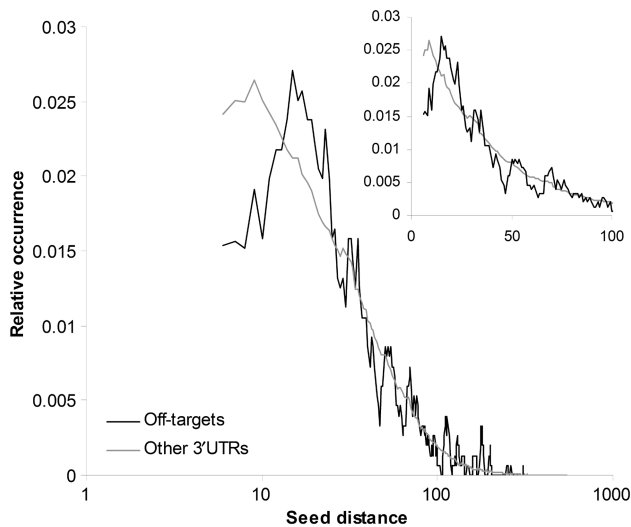


Figure 5. Short interfering RNA seed sites are located farther from seed sites of expressed miRNAs in off-targeted genes than in reference genes, and are more often located at an optimal distance to expressed miRNAs in off-targeted genes than in reference genes. The miRNA seeds are those previously reported to be expressed in HeLa cells (41). See the legend of Figure 4 for additional information.

Table 1. Previous studies of cooperative down-regulation by multiple target sites have primarily looked at target sites separated by an optimal distance

Study	Species	Distance	Sites	Effective
Ref. (35)	Human (HeLa)	24	2, 4, 6	Yes
Ref. (30)	Human (HeLa)	24	2 (4) ^a	Yes
		20	2 (4) ^a	Yes
		16	2 (4) ^a	Yes
		8	2 (4) ^a	No (1 site) ^b
Ref. (33)	<i>C. elegans</i>	47	2	Yes
		32	2	Slight
		24	2	Slight
		47 ^c	2	No
Ref. (23)	<i>D. melanogaster</i>	23	2	Yes
Ref. (31)	Zebrafish	29	2	Yes
		81	2	Yes
			1	No

[TFN] Distances are between the 3' ends of target sites. ^aThe two sites with varying distances were flanked by two target sites at a distance of 24 nucleotides. ^bThe two close sites gave the same knockdown as a single site in the context of the two flanking target sites. ^cMutated linker sequence.

sites, the two close sites gave the same knockdown as a single site. Each of these studies is consistent with our predictions and observations.

Two studies have also looked at cooperativity between more distant sites (31,33). Kloosterman *et al.* (31) used a GFP-reporter to look at let-7 regulation of lin-41 in zebrafish. In wild-type lin-41, the two sites are 81 nt apart and the authors found that let-7 down-regulated a GFP-reporter harboring the wild-type region, but not versions with one mutated site. Moving the sites closer to one another, to a distance of 29 nt, by deleting the region between the sites, also gave down-regulation of the reporter. It would have been interesting to see whether the closer sites gave stronger down-regulation than the

distant sites, but the authors did not, however, quantify the degree of down-regulation.

In the other study, Vella *et al.* (33) looked at cooperative down-regulation of lin-41 by let-7, but in *Caenorhabditis elegans* instead of zebrafish. The *C. elegans* wild-type lin-41 contains two let-7 target sites separated by a 27-nt spacer, which in our reference system means that the sites are 47 nt apart. This spacer sequence is important, as mutating the sequence abolished lin-41 down-regulation by let-7. Nevertheless, removing part of the spacer to bring the target sites closer together (32 and 24 nt distances) reestablished some of the down-regulation. The authors speculate that the linker sequence contains binding sites for proteins or RNA co-factors that are necessary for let-7 to down-regulate lin-41. Indeed, the linker does contain a potential binding site for cel-mir-265 (Supplementary Figure 11). This site is 27 nt from the 5' let-7 site and 20 nt from the 3' let-7 site and it is disrupted in the mutated spacer sequence (Supplementary Figure 11c).

Even though the cel-mir-265 site in the spacer sequence is a prediction and needs experimental verification, the presence of an miRNA-binding site in the spacer would explain the experimental results. Assuming the spacer contains a target site, mutating the spacer sequence disrupted the cel-mir265 target, left the two let-7 sites at a suboptimal distance for cooperative regulation, and abolished any detectable down-regulation. Removing the spacer however, reestablished down-regulation as it brought the let-7 sites closer, but the down-regulation would only be partial as the target now only contained two instead of three optimally spaced target sites. Thus, theoretically it is likely that lin-41 down-regulation in *C. elegans* requires cooperativity between three miRNA sites.

DISCUSSION

Since the first validated targets contained multiple sites, it has been proposed that more miRNA-binding sites automatically result in higher potency (35). As our experiments have demonstrated, this is not necessarily true and very close sites can even yield lower efficacy than a single site. Strong target sites, however, should potentiate the extent of target protein down-regulation. Optimally spaced sites are strong targets which are likely to result in the miRNA acting as translational inhibitors (16,17,48).

Optimal spacing between functional sequence elements is not uncommon. For example, the spliceosome depends on proximal exonic splicing enhancers to separate true splice sites from random occurrences of identical short motifs throughout introns (49). Furthermore, clusters of short sequence-specific transcription factor DNA-binding sites contribute to higher specificity and much stronger RNA polymerase II activity than do single sites (50). For transcription, multiple binding sites can be synergistic, which has also been proposed for RNAi (35).

One possible explanation for the distance dependency between seeds could be that the miRNA guides RISC to

the complementary target sites, but occupancy at a site is dependent upon the strength of the miRNA-complex interaction with the target site. Binding of one complex may serve as scaffold for attracting cofactors necessary for repression. If the sites are too close there may be steric hindrance resulting in reduced function as we observed with the 9 base spacing versus a single site (Figure 2b). Optimally spaced sites, however, facilitate complex or cofactor interactions with adjoining sites. When target sites are too distal, complexes may not be capable of physical interaction.

Our findings should be of use in developing improved miRNA target prediction algorithms, as we have now incorporated the concept of sub-optimal versus optimal spacings between sites as a predictor of efficacy. Very potent targets are likely to result in multiple miRNA-containing complexes binding within a narrowly defined region of the target to optimize functional interaction. To illustrate, there are 12 735 non-overlapping conserved pairs of hexamer seed sites throughout human 3' UTRs for the miRNAs in version 8.0 of miRBase (18), but only 2257 pairs which are separated by more than 13 and less than 100 nt. The corresponding numbers for heptamers and adenosine-anchored hexamers are 286 of 1666 and 196 of 1103 (see Supplementary Table 2 for a comprehensive list of conserved, human 3' UTR pairs of the various seed types).

Our results also indicate that multiple co-expressed miRNAs will cooperate to down-regulate targets that contain multiple consecutive optimally spaced seed sites. A recent study reports that human 3' UTRs contain mosaics of non-overlapping sequence elements that are related to miRNAs (51). The distance between the starts of such consecutive elements is most frequently between 18 and 31 nt, with 18 and 22 nt being the most frequent distances. In light of our results, these consecutive sequence elements have the potential to be clusters of cooperating miRNA target sites. Whether or not these clusters strongly down-regulate a candidate target will however, likely depend on how many and which miRNAs are expressed in the cell at a given time. Off-targeting by siRNAs can also be explained in this context, as off-targets may be the result of whether or not the siRNA can significantly affect the regulatory clusters already present in a gene or cooperate with endogenous miRNAs to establish new regulatory clusters.

Rigoutsos and colleagues reported that coding regions (CDR) and 5' UTRs contained mosaics of sequence elements as well (51), and miRNAs can target both 5' UTRs and CDRs (31). However, we could not recover the distance patterns from the 3' UTRs in these regions (Supplementary Figure 12).

In summary, our results indicate that the distance between pairs of seed sites is important for the strength of down-regulation for a particular target. Cooperation between multiple RISC's requires target sites to be close and is most effective when the distance is between 13 and 35 nt. Furthermore, our results indicate that siRNA off-targeting is related to cooperative down-regulation by endogenous miRNAs. We therefore expect that more effective algorithms for predicting both miRNA targets

and siRNA off-targets can be derived from the results and analyses presented here.

SUPPLEMENTARY DATA

Supplementary data is available at NAR Online.

ACKNOWLEDGEMENTS

O.S. and P.S. received support from the Norwegian Functional Genomics Program (FUGE) and the Leiv Eriksson program of the Norwegian Research Council; L.A.A. was supported by the Alfred Benzons Foundation; and J.J.R. received support from the NIH (AI29329; AI42552 and HLB 07470). The authors would like to thank H.S. Soifer and O.R. Birkeland for reviewing and providing helpful comments on the manuscript. Funding to pay the Open Access publication charge was provided by NIH funding.

Conflict of interest statement. None declared.

REFERENCES

- Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
- Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
- Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, **294**, 862–864.
- Ambros, V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
- O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. and Mendell, J.T. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, **435**, 839–843.
- He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W. *et al.* (2005) A microRNA polycistron as a potential human oncogene. *Nature*, **435**, 828–833.
- Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
- Hannon, G.J. and Rossi, J.J. (2004) Unlocking the potential of the human genome with RNA interference. *Nature*, **431**, 371–378.
- 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.
- Yekta, S., Shih, I.H. and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, **304**, 594–596.
- Zamore, P.D., Tuschl, T., Sharp, P.A. and Bartel, D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, **101**, 25–33.
- Bagga, S., Bracht, J., Hunter, S., Massirer, 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.
- Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P. and Schier, A.F. (2005) MicroRNAs regulate brain morphogenesis in zebrafish. *Science*, **308**, 833–838.
- Wu, L., Fan, J. and Belasco, J.G. (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proc. Natl. Acad. Sci. USA*, **103**, 4034–4039.
- Olsen, P.H. and Ambros, V. (1999) The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.*, **216**, 671–680.

16. Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, **403**, 901–906.
17. Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, **75**, 855–862.
18. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–144.
19. Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U. *et al.* (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.*, **37**, 766–770.
20. 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.
21. 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.
22. Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S. and Kellis, M. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, **434**, 338–345.
23. Brennecke, J., Stark, A., Russell, R.B. and Cohen, S.M. (2005) Principles of microRNA-target recognition. *PLoS Biol.*, **3**, e85.
24. 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.
25. Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
26. 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 downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.
27. 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.
28. Jackson, A.L., Burchard, J., Schelter, J., Chau, B.N., Cleary, M., Lim, L. and Linsley, P.S. (2006) Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity. *RNA*, **12**, 1179–1187.
29. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
30. Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.
31. Kloosterman, W.P., Wienholds, E., Ketting, R.F. and Plasterk, R.H. (2004) Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res.*, **32**, 6284–6291.
32. Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.S., Tam, W.L., Thomson, A.M., Lim, B. and Rigoutsos, I. (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**, 1203–1217.
33. 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.
34. Didiano, D. and Hobert, O. (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.*, **13**, 849–851.
35. Doench, J.G., Petersen, C.P. and Sharp, P.A. (2003) siRNAs can function as miRNAs. *Genes Dev.*, **17**, 438–442.
36. Chen, K. and Rajewsky, N. (2006) Natural selection on human microRNA binding sites inferred from SNP data. *Nat. Genet.*, **38**, 1452–1456.
37. Rajewsky, N. (2006) microRNA target predictions in animals. *Nat. Genet.*, **38 Suppl**, S8–13.
38. Vella, M.C., Reinert, K. and Slack, F.J. (2004) Architecture of a validated microRNA::target interaction. *Chem. Biol.*, **11**, 1619–1623.
39. Sethupathy, P., Corda, B. and Hatzigeorgiou, A.G. (2006) TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA*, **12**, 192–197.
40. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A. *et al.* (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.*, **34**, D590–598.
41. Baskerville, S. and Bartel, D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
42. Rose, S.D., Kim, D.H., Amarzguioui, M., Heide, J.D., Collingwood, M.A., Davis, M.E., Rossi, J.J. and Behlke, M.A. (2005) Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Res.*, **33**, 4140–4156.
43. Wu, Q., Zhang, T., Cheng, J.F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Noonan, J.P., Zhang, M.Q. *et al.* (2001) Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res.*, **11**, 389–404.
44. Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M. and Stoffel, M. (2005) Silencing of microRNAs in vivo with ‘antagomirs’. *Nature*, **438**, 685–689.
45. Birmingham, A., Anderson, E.M., Reynolds, A., Ilesley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K. *et al.* (2006) 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods*, **3**, 199–204.
46. Fedorov, Y., Anderson, E.M., Birmingham, A., Reynolds, A., Karpilov, J., Robinson, K., Leake, D., Marshall, W.S. and Khvorova, A. (2006) Off-target effects by siRNA can induce toxic phenotype. *RNA*, **12**, 1188–1196.
47. Jackson, A.L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J.M., Lim, L., Karpilov, J. *et al.* (2006) Position-specific chemical modification of siRNAs reduces “off-target” transcript silencing. *RNA*, **12**, 1197–1205.
48. Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D. *et al.* (2005) RAS is regulated by the let-7 microRNA family. *Cell*, **120**, 635–647.
49. Maniatis, T. and Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*, **418**, 236–243.
50. Kadonaga, J.T. (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, **116**, 247–257.
51. Rigoutsos, I., Huynh, T., Miranda, K., Tsirigos, A., McHardy, A. and Platt, D. (2006) Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes. *Proc. Natl. Acad. Sci. USA*, **103**, 6605–6610.