

# New Tricks for Animal MicroRNAs: Targeting of Amino Acid Coding Regions at Conserved and Nonconserved Sites

Isidore Rigoutsos

Bioinformatics and Pattern Discovery Group, IBM Thomas J. Watson Research Center, Yorktown Heights, New York

## Abstract

**MicroRNAs (miRNA) are short RNAs that affect the expression of a protein-coding gene either by directing the degradation of its “target” mRNA or by disrupting its translation into amino acids. Target selection depends on the underlying sequence as well as other, currently not understood, factors. In plants, miRNAs typically interact with the amino acid coding region of the target mRNA. However, in animals, research has been nearly exclusively confined to studying miRNA interactions with the 3′ untranslated region of their mRNA targets. This region-limited view of animal miRNA activity, together with the assumption that *bona fide* animal miRNA targets ought to be conserved across organisms, have been in effect for many years. Recent work has shown that miRNAs can target extensively the amino acid coding region of animal mRNAs and can do so at locations that are not necessarily conserved across organisms.** [Cancer Res 2009;69(8):3245–8]

## Introduction

MicroRNAs (miRNA) represent a class of short (~22 nucleotides in length), endogenous, noncoding RNAs that powerfully regulate eukaryotic gene expression (1). MiRNAs act on their targets in a sequence-dependent manner, and modulate the expression of the targets either by directing the degradation of the mRNA transcript or through inhibition of translation (2). What makes microRNAs particularly important is their involvement in most, if not all, fundamental biological processes. These include timing of developmental transitions (3, 4), induction of organ asymmetry (5), tumor suppression (6), oncogenic activity (7), invasion and metastasis (8), modulation of embryonic stem cell differentiation (9–11), neurodegeneration (12), etc. Moreover, extensive profiling work in different tissues has shown that, in various combinations, subsets of the currently known miRNAs are dysregulated in disease *versus* normal states (13, 14).

The founding members of the miRNA class, *lin-4* and *let-7*, were shown to repress their targets through interactions with the targets' 3′ untranslated region (UTR; refs. 3, 4, 15). It is probably partly due to this finding that subsequent research efforts focused, almost exclusively, on the study of 3′UTRs as targets of miRNA activity. Interestingly, in plants, most of the reported examples involved interactions with the amino acid coding region (CDS) of the target genes (1, 16), a result that for years has been fueling an ostensible distinction between plants and animals. Indeed, from a mechanistic standpoint, there is no compelling reason for singling out this particular region of the mRNA: as shown by artificial

insertion of known *let-7* binding sites in 5′UTRs and CDSs, these regions are perfectly capable of “hosting” miRNA targets (17, 18).

In the early days, 3′UTRs were mostly a *tabula rasa* of sorts with regard to their potential roles: they represented a largely unexplored segment of the typical mRNA and had average lengths that correlated conspicuously well with the apparent complexity of the organism (19). Such considerations made 3′UTRs a reasonable and appealing choice for packing gene-specific regulatory elements. By contrast, the nucleotides comprising the CDS already had an assigned role making this region a seemingly less-than-ideal choice for hosting regulatory signals. On the other hand, the very fact that the CDS nucleotides were constrained through their participation in codons also endows them with increased stability against mutations, which, in turn, makes them good candidates for forming persistent miRNA/mRNA heteroduplexes.

It can be argued that the initial focus on animal 3′UTRs was further bolstered by the many computational approaches that were devised for the prediction of miRNA targets. Characteristically, the majority of these methods relied on the available miRNA+3′UTR heteroduplexes for their training. This, in turn, enabled the discovery of more examples of miRNAs targeting animal 3′UTRs; these findings subsequently became part of new, augmented training sets (20). Not unexpectedly, as time went on, 3′UTR-centered miRNA interactions became almost synonymous with animal miRNA activity.

The first few validated heteroduplexes, as well as many of those reported later, made apparent a distinct preference for the reverse complement of the first few bases of the miRNA to be present at the miRNA's target site (4, 15). This core segment became known as the “seed” and typically corresponds to bases 2 through 7 inclusive counting from the 5′ end of the miRNA (1). Nonetheless, some of the early reported heteroduplexes involving miRNAs *lin-4* and *let-7* also provided exceptions: bases in the seed did not base pair with the target, forming bulges instead. As in the case of the 3′UTRs, this enforcement of the constraint led to the discovery of additional examples that supported it, whereas cases such as the bulged *lin-4/lin-14* heteroduplex were viewed as outliers for many years.

The importance of the seed was further corroborated by *in vitro* and *in vivo* experiments (and more recently by crystallographic data; ref. 21) that showed the adverse impact that single point mutations could have on gene expression (22, 23). These experiments typically focused on single point changes that would break a base pairing in the seed region, or replace a G:C bond by the weaker G:U one. Such studies affected two areas of great interest to practitioners, namely, gauging the magnitude of off-target effects, and obtaining better estimates for the number of genes that are under the control of a given short RNA. If an unpaired or weakly paired nucleotide in the seed region weakened or abolished miRNA activity, then potential off-target effects of siRNAs would present less of a concern (24). Analogously, if the miRNA seed acted as a binary specificity filter, then the eventual number of targets for a miRNA would be greatly reduced, especially if combined with a

**Requests for reprints:** Isidore Rigoutsos, IBM Thomas J. Watson Research Center, 1101 Kitchawan Road/Rte 134, Yorktown Heights, NY 10598. Phone: 914-945-1384; Fax: 914-945-4217; E-mail: rigoutso@us.ibm.com.

©2009 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-09-0352

focus on only the 3'UTR. This would, in turn, imply that a smaller fraction of the protein-coding genes in a given organism are under miRNA control: early analyses estimated this number to be ~25% of the known human genes (25), whereas later ones increased it to >90% (9).

One remaining consideration was how best to identify candidate miRNA target sites, and this is again where the appeal of the 3'UTRs received an additional boost. Generally speaking, the 3'UTRs of orthologous genes are not conserved. Thus, when presented with a multiple sequence alignment of orthologous 3'UTRs, conserved segments in the context of an otherwise nonconserved background look rather conspicuous (3). Focus on the sequences in such alignments of multiple species improves the ability to localize candidate miRNA target sites; this is achieved either by aligning sequences from distinct species of the same genus (3, 26) or sequences from different genera (27). With time, the increased power of cross-genome conservation to identify and localize regulatory sequences evolved into a seemingly *necessary* and *sufficient* constraint for identifying miRNA targets. Such an interpretation has an important complication attached to it: if it is eventually proven to be unwarranted, it will mean that the full complement of regulatory signals in a given organism is, in fact, higher than what sequence conservation could ever reveal. Interestingly, recent findings began lending support to this possibility (10, 11, 28, 29).

Unlike the miRNA target-finding schemes that preceded it, the scheme designed by the present author (known as *rna22*; ref. 9) has some unique characteristics that were meant to address the above considerations. *Rna22* obviates the need for validated heteroduplexes for the training phase, relying instead on the sequences of known *mature* miRNAs. Moreover, *rna22* does not use the seed sequence (whether implicitly or explicitly) of a given miRNA to subselect among candidate targets. Finally, it neither enforces nor relies on the conservation of a putative target sequence across organisms. These properties allow *rna22* to cast a wide net when searching for targets, and early on provided strong computational support for widespread regulation by miRNAs through the 5'UTR and CDS region of a mRNA, in addition to the conventional 3'UTR (9).

The above discussion makes the assumption implicitly that primary nucleotide sequence can serve as an effective proxy for gauging biological relevance in miRNA/mRNA interactions, i.e., if the quality of the base pairing between the miRNA and its putative target is high, then there is a good chance that one will observe a biological effect, and vice versa. Although intuitively appealing, mutation studies (30), and numerous luciferase assays involving three different miRNAs (9) provided counterexamples. Such observations led practitioners to consider the context in which such interactions take place and to incorporate constraints such as target accessibility, base-pairing topology, and nucleotide composition of adjacent regions (31–33). Although such filters ignore the effect of *chaperone*-like molecules, tissue-specific expression, temporal dependence, etc., they provide reasonable constraints that are expected to improve, at least in principle, the hit-miss ratio of computationally predicted miRNA targets.

## Beyond the 3'UTR: New Findings

During the last several months, various groups reported evidence in support of miRNA targeting of CDS. These efforts fall into two categories.

The first category comprises three publications. In the article by Duursma and colleagues (34), the authors present evidence for a

functioning target of miR-148 in the CDS of a DNA methyltransferase. The target is conserved in orthologues from several vertebrates and was identified because of its extensive complementarity (17 consecutive nucleotides) to miR-148. A second publication reports the presence of a miR-126 target, conserved in at least 8 vertebrates, in the homeodomain of the *HOXA9* gene (35). The discovery of this target was somewhat serendipitous: the on-line repository that the authors consulted had treated the CDS region of interest as the 3'UTR of a shorter, alternative transcript of *HOXA9* that lacks the homeodomain. The third contribution resulted from the study of a 17 genome multiple sequence alignment that revealed three conserved targets for let-7 in the CDS region of *Dicer* (27). All three let-7 targets were shown to be functional, suggesting cooperativity and the existence of a miRNA/*Dicer* negative feedback loop.

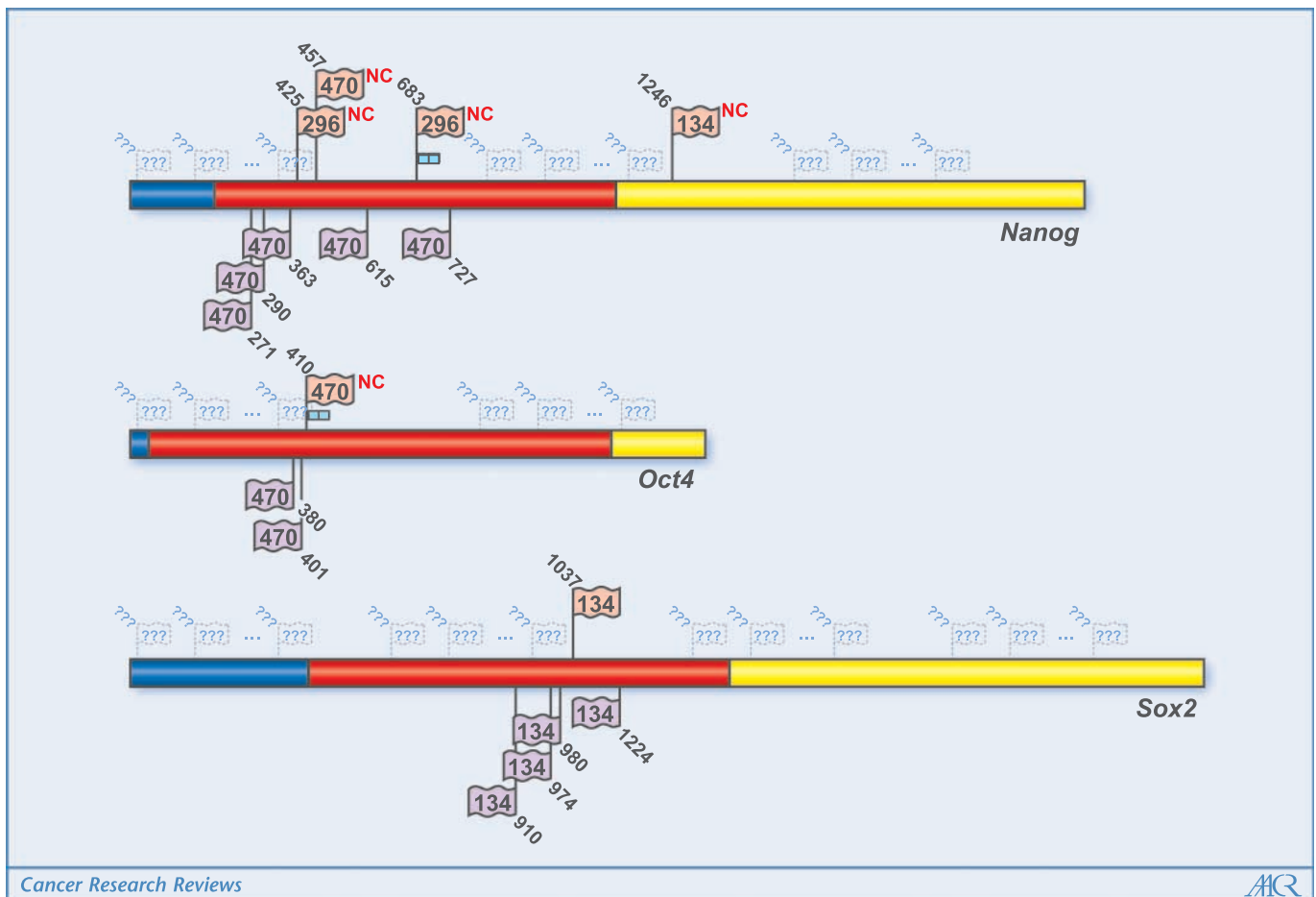
The two publications in the second category report CDS targets, identified using *rna22* (9), which are not conserved across genomes. Interestingly, both publications show evidence that the studied miRNAs act predominantly through translational inhibition. In the article by Lal and colleagues (28), the authors show that miR-24 suppresses p16 expression in human diploid fibroblasts and cervical carcinoma cells, primarily through a CDS target site (and to a lesser extent through a 3'UTR one). In the article by Tay and colleagues (10), a total of five CDS targets for miR-296, miR-470, and miR-134 are shown for *Nanog*, *Oct4*, and *Sox2* and validated using mutation studies and physiologic evidence. Several *additional* targets for miR296, miR-470, and miR-134 likely exist in the CDS regions of *Nanog*, *Oct4*, and *Sox2* as suggested by luciferase assays reported by the authors. The numerous targets discussed by Tay and colleagues (10, 11) exemplify the complexity of the interactions between miRNAs and mRNAs, something that we attempt to capture pictorially in Fig. 1.

## Implications

These recent findings reveal a novel facet of miRNA-driven regulation in animals, namely the targeting of CDSs. As evidenced by the context in which they are validated, CDS targets can be as physiologically relevant as their 3'UTR counterparts. Just as in 3'UTRs, one miRNA can have multiple targets in the CDS of the same mRNA, and multiple miRNAs can target the same CDS in a combinatorial and cooperative manner. Notably, the miR-296/*Nanog* interactions described by Tay and colleagues (11) show that a miRNA may target *only* the CDS of a mRNA. These results not only support a uniform paradigm governing miRNA interactions with the 3UTRs and CDSs, but also go toward bridging the conceptual gap between animals and plants with regard to miRNA targeting.

Of the five validated targets in the *Nanog*, *Oct4*, and *Sox2* CDSs, four have no complementarity to the seed sequence of the targeting miRNA (10). All four heteroduplexes contain a G:U pair in the seed region, whereas two of them also contain a bulge in the seed region. Thus, active heteroduplexes with either G:U pairs or bulges, or both, in the seed region may be more abundant than currently thought.

Additionally, the four targets in *Nanog* and *Oct4* (10) and the one in p16 (28) are not conserved in other organisms. Even restricting the examination to only the seed region of the target, we find that in three of the four *Nanog* and *Oct4* targets reported by Tay and colleagues (10), the seed region is not conserved between mouse and either the human or rhesus orthologues of the transcription factor at hand. Thus, enforcing cross-genome conservation constraints is likely to miss *bona fide* miRNA targets. The more



**Figure 1.** The ever-increasing complexity of miRNA-driven regulation as evidenced by the many-to-one and one-to-many interactions involving three miRNAs (miR-296, miR-470, miR-134) and three genes (*Nanog*, *Oct4*, *Sox2*). The three miRNAs are up-regulated and the three genes down-regulated during mouse embryonic stem cell differentiation. *Colored flags*, targets validated through mutation studies and physiologic experiments: each flag lists the targeting miRNA, whereas the number next to the flagpost is the distance from the start of transcription. *Cyan markers next to the flag-post*, targets spanning exon-exon junctions. *Dark gray flags*, predicted targets for which only luciferase assays were carried out. *Small light gray flags*, abundance of numerous predicted targets for other miRNAs that await validation. Each mRNA and the corresponding region are drawn to scale: *blue*, 5'UTR; *red*, CDS; *yellow*, 3'UTR. *NC*, corresponding miRNA target is not conserved in close organisms. All data shown here are based on earlier publications (9–11).

(and more divergent) species are included in the alignment, the higher the rate of these false negatives will be.

Another finding described by Tay and colleagues (10) was that two of the validated targets span exon-exon junctions. This result may have revealed another previously unrecognized, important property of miRNAs, namely an ability to recognize splicing alternatives of a given transcript. If true, it will mean that a miRNA can in fact distinguish among different isoforms, and selectively repress some but not others.

A final comment relates to a previously reported *Nanog* target. In an article by Tay and colleagues (11), miR-134 is shown to target the 3'UTR of *Nanog*. What makes this target notable is that it is internal to the sequence of a B2 SINE element embedded in the 3'UTR of *Nanog* (10). To the best of our knowledge, this is the first reported example of a heteroduplex involving a miRNA and a repeat element. It also suggests that sequences of repeat elements embedded in exons can be co-opted by the RNAi layer as miRNA target sites and made part of the normal cell process regulation. It is worth pointing out that the sequences of repeat elements are generally genus specific, which suggests another mode of contribution to miRNA/mRNA interactions that are not conserved

across organisms. Analogous findings were recently reported in the context of human and mouse intronic sequences that were shown to be involved in extensive conserved functional links in the absence of sequence conservation (29).

We have reviewed several recent results that depart from the model of animal miRNA targeting that has been in effect for almost a decade. However, additional studies will be necessary before we can answer beyond any doubt whether these findings represent experimental anecdotes or are of more general nature. Considering the ever-increasing complexity that miRNAs (and other short RNAs) contribute to gene expression, it will be prudent to keep an open mind and follow the path that is being illuminated by the accumulating data.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Acknowledgments

Received 1/29/09; accepted 2/19/09; published OnlineFirst 4/7/09.

I thank Peter T. Nelson and Zissimos Mourelatos for critical reading and comments on early versions of this manuscript.

## References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–97.
2. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 2008;9:102–14.
3. Ha I, Wightman B, Ruvkun G. A bulged lin-4/lin-14 RNA duplex is sufficient for *Caenorhabditis elegans* lin-14 temporal gradient formation. *Genes Dev* 1996;10:3041–50.
4. Reinhart BJ, Slack FJ, Basson M, et al. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 2000;403:901–6.
5. Didiano D, Hobert O. Molecular architecture of a miRNA-regulated 3' UTR. *RNA* 2008;14:1297–317.
6. Hammond SM. MicroRNAs as tumor suppressors. *Nat Genet* 2007;39:582–3.
7. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 2006;6:259–69.
8. Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA10b in breast cancer. *Nature* 2007;449:682–8.
9. Miranda KC, Huynh T, Tay Y, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 2006;126:1203–17.
10. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs at *Nanog*, *Oct4* and *Sox2* coding regions modulate embryonic stem cell differentiation. *Nature* 2008;455:1124–8.
11. Tay YM, Tam WL, Ang YS, et al. MicroRNA-134 modulates the differentiation of mouse embryonic stem cells, where it causes post-transcriptional attenuation of *Nanog* and *LRH1*. *Stem Cells* 2008;26:17–29.
12. Nelson PT, Wang WX, Rajeev BW. MicroRNAs (miRNAs) in neurodegenerative diseases. *Brain Pathol* 2008;18:130–8.
13. Deng S, Calin GA, Croce CM, Coukos G, Zhang L. Mechanisms of microRNA deregulation in human cancer. *Cell Cycle* 2008;7:2643–6.
14. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;6:857–66.
15. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 1993;75:855–62.
16. Mallory AC, Vaucheret H. Functions of microRNAs and related small RNAs in plants. *Nat Genet* 2006;38 Suppl:S31–6.
17. Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* 2007;104:9667–72.
18. Kloosterman WP, Wienholds E, Ketting RF, Plasterk RH. Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res* 2004;32:6284–91.
19. Pesole G, Grillo G, Larizza A, Liuni S. The untranslated regions of eukaryotic mRNAs: structure, function, evolution and bioinformatic tools for their analysis. *Brief Bioinform* 2000;1:236–49.
20. Bentwich I. Prediction and validation of microRNAs and their targets. *FEBS Lett* 2005;579:5904–10.
21. Wang Y, Juraneck S, Li H, Sheng G, Tuschl T, Patel DJ. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* 2008;456:921–6.
22. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. *PLoS Biol* 2005;3:e85.
23. Kiriakidou M, Nelson PT, Kouranov A, et al. A combined computational-experimental approach predicts human microRNA targets. *Genes Dev* 2004;18:1165–78.
24. Jackson AL, Burchard J, Schelter J, et al. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* 2006;12:1179–87.
25. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005;120:15–20.
26. Stark A, Lin MF, Kheradpour P, et al. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* 2007;450:219–32.
27. Forman JJ, Legesse-Miller A, Collier HA. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A* 2008;105:14879–84.
28. Lal A, Kim HH, Abdelmohsen K, et al. p16(INK4a) translation suppressed by miR-24. *PLoS ONE* 2008;3:e1864.
29. Tsigiris A, Rigoutsos I. Human and mouse introns are linked to the same processes and functions through each genome's most frequent non-conserved motifs. *Nucleic Acids Res* 2008;36:3484–93.
30. Didiano D, Hobert O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol* 2006;13:849–51.
31. Grimson A, Farh KK, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007;27:91–105.
32. Hammell M, Long D, Zhang L, et al. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat Methods* 2008;5:813–9.
33. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet* 2007;39:1278–84.
34. Duursma AM, Kedde M, Schrier M, le Sage C, Agami R. miR-148 targets human DNMT3b protein coding region. *RNA* 2008;14:872–7.
35. Shen WF, Hu YL, Uttarwar L, Passegue E, Largman C. MicroRNA-126 regulates HOXA9 by binding to the homeobox. *Mol Cell Biol* 2008;28:4609–19.